

**GNSO/SSAC
International Registration Data Working Group
TRANSCRIPTION
Monday 29 March 2010 at 19:00 UTC**

Note: The following is the output of transcribing from an audio recording of the GNSO/SSAC International Registration Data Working Group on 29 March 2010 at 1900 UTC. Although the transcription is largely accurate, in some cases it is incomplete or inaccurate due to inaudible passages or transcription errors. It is posted as an aid to understanding the proceedings at the meeting, but should not be treated as an authoritative record.

Coordinator: Thank you for standing by. I would like to inform all participants that today's conference is being recorded. If you have any objections you may disconnect at this time. I would like to introduce your host for today's call. We have Ms. Julie Hedlund. Ma'am, you may begin.

Julie Hedlund: Thank you very much (Lori). Good morning, good afternoon, good evening to everyone here. This is the internationalized registration working group call on Monday, March 29. Thank you all for joining us. I'd like to briefly do a roll call please.

From the working group we have the Chair, Edmon Chung. We also have Andrei Kolesnikov, Ram Mohan, Rafik Dammak and Owen Smigelski. And we also have from staff, David Piscitello, Steve Sheng and Julie Hedlund.

Welcome to all of you. And Edmon if you'd like, I have a suggestion for our agenda, for your consideration.

Edmon Chung: Sure.

Julie Hedlund: For all of the working group members on this call, Steve Sheng sent around earlier today a matrix. This is a matrix that was discussed - oh and I do see we have another person who's joined. It's Bob Hutchinson. Hello Bob.

Bob Hutchinson: Good morning.

Julie Hedlund: Good morning. This is Julie Hedlund. Welcome. We're just getting started. We just did a roll call. And I was just suggesting that perhaps as our agenda item today Steve Sheng had sent around the draft matrix that was requested on our last working group call on March 15.

The matrix identifies three different models for registration data and the impact of each model on potential stakeholders. I'd like to suggest to our Chair, Edmon Chung that perhaps we could look at this draft matrix on this call and go through the different elements and discuss them.

Edmon, what do you think about that?

Edmon Chung: I'm looking at the matrix right now. It seems reasonable following from last meeting.

Julie Hedlund: Great. Wonderful. Then what I would like to do actually is to ask Steve Sheng who had put together the matrix, to perhaps lead us in a discussion of the matrix going through each of the elements. And perhaps we can discuss them one at a time and the different models and so forth.

Steve will that be okay for you?

Steve Sheng: Sure Julie. Thanks. So the matrix in the last meeting, I think - we think that a matrix that basically kind of systematically explores the various models on the table of their impact to the registrars registries - registrants and the use of those WHOIS will be helpful.

So we have a - I sent earlier today a draft matrix. Basically it's essentially a first cast - a first cut on the staff. We understand that there's considerable

expertise and representation from registrars registries and various constituencies in these working groups.

So we look forward for your feedback on this work. So roughly, to start with, if you have a moment, if you can pull out the Excel spreadsheet, we have three models currently under consideration. This is the initial three models that was summarized from past deliberations.

You know, it doesn't have to be these three models. We can add more models to explore as well. But to start, the first model is basically requiring must be present language of WHOIS in addition - in addition to that given an option to provide data in local languages.

So the second model is essentially a registrant can provide the registration data in whatever language or script, you know, that can be accepted by the registrar. But it's up to the registrar to provide a point of contact or for the translation of those records when there is a request has been made.

So the third model is very similar - no, it's in some ways somewhat similar to the first model, is that it's requiring both a must be present language and also additional language. But the difference here is registrants still can provide their registration data in whatever language the registrar can accept.

But it is up to the registrar to provide a translation or transliteration of the registration records. So those are the three initial models on the discussion. So to systematically kind of consider the pros and cons of each we consider its impact to the various stakeholders.

So on the first column, on Column A on the left, you know, we have impact to registrars. Here we separate out to the existing registrars and to the new IBM based registrars. Then we consider the impact to registries.

Here we separate out impact to thin registries like in the dot com model and to thick registries like dot org and others. Again we consider impact to registrant. And finally, we consider the impact to each of these models to the users of who is.

And here when we mean users of WHOIS we separate it into three subcategories. So the first category is the local user of who is, that knows the local language. For example, a Japanese user, you know, who also receives WHOIS in Japanese.

And the second is the local user, the type of user that receives WHOIS in any language but local language. So for example, a - like a Chinese registrant who does not know like Indian or Japanese but needs to look up domain names registered by registrants of those countries.

So that's a second category. And the last category is legitimate automation a user, basically someone that, you know, would use WHOIS legitimately to - for some automated purposes. For example, compliance, accuracy reports and such.

So that's a rough, that's a brief introduction of the matrix. So before going forward discussing each of the models, I would like to see - are there any comments from the participants of the working group, on the structure of the exercise and the different categories of stakeholders we've considered?

Bob Hutchinson: Yeah. This is Bob Hutchinson. Thanks for putting this together Steve. And I guess I would like to see added as an impact to the three different models, the impact on the WHOIS system itself in terms of its data storage and...

Steve Sheng: So just to clarify - okay, you mean the impact to WHOIS system in storage. So these impacts can occur, you know, both you know registrars could take some impact and registries could take some impact as well.

So are you suggesting a new category - or in the existing categories we consider the impact to the WHOIS system that each of those maintains?

Bob Hutchinson: Yeah. You have three different models. I'm not asking adding models. I'm asking to add another set of impacts which is impact of the - for that model on the existing WHOIS system itself.

Steve Sheng: Okay, sure.

Man: Like RFC standards?

Bob Hutchinson: Yes. In other words, there's a number of ways we can look at the problem that's in front of us. A lot of what we've been discussing has to do with what I consider to be policy issues. There's another way to look at this problem. I think Steve (unintelligible), I don't know if he's on the line, mentioned that.

Which is just look at it from this - from the pure data processing standpoint. Don't look at it from a policy standpoint. Just build a system that will do the worst case of these three models, for example, okay? And let somebody else deal with the policy.

Steve Sheng: Okay, thanks.

Ram Mohan: Steve, this is Ram.

Steve Sheng: Hi Ram.

Ram Mohan: Again, thank you for putting this together. I think the structure is fine. It provides a good way to look at the overall landscape. I wanted to speak just briefly on impact to registries.

Under impact to registries you said to think registries in Model 1 and Model 3, you've said thick registries can set a standard regarding what language and

encoding they want. And registrars have to comply. I don't believe this is actually the case.

Because at the end of the day, all the thick registry is doing is putting out data in regular standard ASCII right now. And registrars simply conform to that. So in some ways it's really not very different from the thin WHOIS case particularly in cases where it's just ASCII based.

Steve Sheng: Right. Thanks.

David Piscitello: Hey Ram this is Dave. That's actually my comment in the matrix. And my understanding was that certain registries would not accept UTF8 if it was submitted to them. That they would reject it and say no, it must be ASCII 7. And that's the kind of criteria we were thinking of, not anything beyond that.

((Crosstalk))

Ram Mohan: ...the text has to be modified a little bit because it's not - that clarity is not here. This seems to be more about output from registries rather than input into registries.

David Piscitello: Yeah, well we - this is just (staff) trying to go through and get some discussion points. And actually we were relying on people like yourself and the registrars to embellish on what the problems were because we...

Ram Mohan: Yeah. I guess my take on some of this is that, you know, if you're not accepting a UTF8 then why do we have this working group.

That, you know, kind of part we're trying to do here is to reflect the new reality that registrants - especially with IDNs becoming more prevalent, registrants are likely to give information in UTF8. And having a UTF7 and, you know, or a 7 bit solution, doesn't necessarily solve the problem.

David Piscitello: I agree. And we're just trying - I think the reason why we put the matrix together was that we were in cross conversations, sort of wandering in and out of these three different models.

And we thought that perhaps if we were able to build the matrix in this manner we could focus on the specific issues within each scenario.

And then if the conclusion is the world's got to be UTF8 and policy development ought to acknowledge that and technology development ought to, you know, acknowledge that and automation ought to acknowledge that then I think that's an important conclusion that this working group would present to the GNSO and to, you know, and to the community.

Ram Mohan: I understand. Thank you for that clarification. I guess from - the bias that I bring into this working group and this discussion is twofold. One is that, you know, folks do need to find some way to have some potentially uniform representational guidelines or representational characteristics.

But at the same time I don't think we can forget folks who need - or who are only able to display data in their local scripts. So I mean that's kind of the balancing act. So that's my own bias, is to insure that the local script piece is not forgotten.

The other comment I had on the impact to registrars which I thought might be something that, you know, we should parse a little bit more is that between the three models there is - there are words like significant increase in the cost, some increase in the cost, etc.

And if you're going to make variations in that then we probably ought to be a bit more precise because otherwise, you know, it can get lost in the semantics. What is the difference with a significant increase in Model 3 versus an increase in Models 2 and 1?

So those are my two comments or suggestions Steve.

Steve Sheng: Thanks. Thanks Ram.

((Crosstalk))

Julie Hedlund: Actually - this is Julie. I just want to remind everybody please do announce yourself each time you speak because it does help with the transcribing.
Thanks so much. Sorry for the interruption.

Steve Sheng: Hi. This is Steve. These are great comments. I think some of them are relating to, you know, each different sections. And I just want to reiterate that the staff is only, you know, kind of doing a first catch - doing a first cut.

And we're really relying on, for example, you know, representatives from registrars or registries, to help us fine tune on that and make sure that the language we put in is correct and suggest the most correct language.

So with that, are there any further comments on the structure of this exercise? If not, we're going maybe, you know, model by model and, you know, stakeholders by stakeholders. Hearing no - so can we go to Model 1, impact to registrars.

So in this model we require must be present as well as local language. The only - are there any one from running registrar who can respond to...

Andrei Kolesnikov: Yeah.

((Crosstalk))

Andrei Kolesnikov: ...I can speak. Regarding the price - regarding the cost increase well it's a general comment because it's required to any of the models in general. Any

technical development or a break of the system is the way cheaper than any organizational additional cost.

Like a translation of manual procedures or, you know, finding the person who understands the language, etc. So I believe that the comment in the Model 1 to existing registrars - if they do end up to use a local script, then present additional costs for them to sort the data.

I don't think that sorting the data is taking the additional costs. That's my comment.

Steve Sheng: Thanks. Is that Andrei?

Andrei Kolesnikov: Yes. It's Andrei.

Steve Sheng: Okay, thanks. Thanks Andrei. I'm taking down carefully. Any other thoughts? Okay. What about Model 1's impact to registries? I'm already has taking down your comments. So let me see if I can paraphrase it.

What you're saying is there's no difference to thin and think registries with the respect to the impact of Model 1. And merely is registries let's say requiring for ASCII or UTF8 can then - and then the registrar will provide that information in that encoding. Is that correct?

Ram Mohan: This is Ram. Yes, that's correct.

Steve Sheng: Okay. So another question we have on this is would the thick registry - would a registry plan to include, you know, in Model 1, languages other than the must be present language?

So for example, if a Chinese registrant or a Japanese registrant register a dot org domain, you know, from the registrar in China or Japan and provides the

local language, would the registry be able to show Chinese and Japanese in the registry choice?

Ram Mohan: This is Ram. The answer is actually it depends. In general, if you look at the org registry or any of the - in most registries they certainly have the capability to display language in both - display data in both the must be present language as well as local languages.

Storing that in an EPP based system is quite trivial. Displaying it however, is where some of the challenges come up. And that would be - today it is defined by registry policy.

So if you look at the dot Asia registry or the dot info registry or dot org, etc. if there is local internationalized registration data in a local character set on a Web page that can be displayed using HTML escape characters.

And it's, you know, it's ASCII-ish because HTML escape characters allow you the flexibility. In the registry itself, registries have to make certain that the databases that they're using can support UTF8 because if you have the data coming in EPP supports UTF8.

So if data comes into the registry in UTF8 there is no guarantee that it'll get stored UTF8. So if we had to insure that the database itself is UTF8 compliant otherwise we'll have data lost or worse, data corruption.

Steve Sheng: That's a great point. Andrei you're from the - a CCTLD. I was wondering what kind of format are you used to store the...

Andrei Kolesnikov: Yeah. Okay. We use - currently we use ASCII for (unintelligible). And also we will keep the ASCII for the dot (RF) for the IDN registration.

Steve Sheng: Okay.

Andrei Kolesnikov: And it's basically related mostly to the demands of the, you know, different legal and, you know, enforcement agencies who really don't want to lose, you know, track and they need to understand the data.

And the resolution of this problem which we're talking is, you know, we took a simple approach. We basically proved the Russian language in the license script. And so far for 15 years, somehow it works.

And I should say about the technical capabilities here, the database which we're using which is basically Oracle and the UTF8...

Steve Sheng: Okay.

Andrei Kolesnikov: ...where we can store any (chart) set.

Steve Sheng: Okay. That's good to know. Sure. UTF8 - all right. That's good. Any other thoughts and comments on the Model 1's impact to registries?

Ram Mohan: This is Ram. Sorry to monopolize the registries piece of the data. But I think the - there is really - in addition to there not being a big difference between thin and thick registries in terms of how the data is stored, I think one of the things that we should potentially consider in this working group is what are we going to do about statuses?

And I don't know if that's already been covered elsewhere. But, you know, we have several statuses for domain name. It can be in transfer pending. It can be in redemption grace period. It can be active, etc. And is there a need to have those be made accessible in local languages as well?

Now that will actually pose at least from - I think from a gTLD perspective that'll pose a little bit more of an implementation challenge as compared to simply adding more objects, more registration objects in multiple languages to just hold in.

((Crosstalk))

David Piscitello: We actually had a separate discussion that is I think a derivative of this discussion, which is whether it's necessary to provide the entirety of the domain record in local language or whether the - it would be appropriate or sufficient to only have - or continue the contractors of having the registrar provide, you know, what the registrar's information is in ASCII 7 only.

So I think that's like - some people, you know, have suggested that we could separate out contact information and establish conventions for them for catering to local languages because that's something registrants supply.

And the information that registrars supply to continue to be provided in the manner that it current is for gTLDs which is in ASCII 7.

Steve Sheng: Thanks Dave. Comments? Thoughts? Let's move onto - let me take a note of that issue for discussion. So moving onto the impact to registrants for Model 1 are there any advocates of registrants rights on the call that can tell us whether we've thought of the impact carefully?

Avri Doria: Hey, hi. This is Avri. I'm not sure I qualify as an advocate of registrants rights but I'll play one....

Steve Sheng: Go ahead.

Avri Doria: ...for the purposes of conversation. So in terms of Model 1 yeah, I think you've basically got - actually yeah. I think you do have it correct in terms of views of receiving it in local as opposed to - and also to the one receiving it any language because both of them are there.

So I think that this for the registrant is - yeah, is indicated correctly.

Steve Sheng: Thanks.

Avri Doria: I don't know what to say about automation issues but I assume that it's correct. That there's little impact.

Steve Sheng: Right. But that's for registrants. So...

Ram Mohan: Steve this is Ram.

Steve Sheng: Go ahead.

Avri Doria: Oh yeah. I was looking at both users and registrants.

Steve Sheng: Oh. Okay.

Avri Doria: Sorry. I was looking at the wrong place.

Steve Sheng: Okay, sure. Ram, go ahead.

Ram Mohan: I had to look at both registrants and users in a kind of commingled way. And the only thing to consider is that we cannot look at WHOIS just as a system in and of itself. Quite a lot of the user experience is going to be dictated by the application level usability.

And the applications don't support UTF8. It doesn't matter if the data that is being sent through is in UTF8. So I think somewhere you've got to just put a disclaimer that says that it's only going to be as good as, you know, at the user level.

But certainly in places like (India), there are plenty of spots where IE5 is still the prevailing browser. And, you know, good luck trying to show IDN stuff in a proper user-friendly format.

Steve Sheng: That's an excellent point Ram. So would you suggest we put another category saying applications to the...

Ram Mohan: Yeah. I totally agree with that - with your suggestion. I think that really is the elephant in the room. That's where the biggest amount of change has to happen or advocacy has to happen if data really has to be, you know, made universally accessible in local languages.

Steve Sheng: Okay. Okay. Other thoughts? Comments?

Avri Doria: Yeah, this is Avri again. So if I understand correctly, it's not - you're adding a category of impact to the application that people use because it'll work for the registrants but only if somebody has (unintelligible).

Ram Mohan: That's - Avri this is Ram. That's correct. That's exactly what I'm saying.

Steve Sheng: Yes.

David Piscitello: This is Dave. That's true for a lot of WHOIS applications as well and operating systems. I mean what (play) will have to do with, you know, what (spots) and character sets of work exist on the client, right?

Ram Mohan: Right. And Dave this is Ram. You know, one thing that since we're talking about at looking at this in a somewhat comprehensive way, one thing that is possible to do is as a working group, to make a recommendation that says - add a field that says if you cannot see this go here to download appropriate tools or look at minimum requirements.

Or something like that to provide users some clue potentially.

Avri Doria: This is Avri again. Yeah, I would think that something like that would be a good idea because I would think that there's no barrier at this point for most

of the languages we're talking in IDN in terms of getting more recent tools but that can handle the IDN, etc.

Steve Sheng: Thanks Avri. I've taken a note of that. So for impact to applications what are the - are there any subcategories that we want to consider?

I'm trying to think if staff were to ask to take a first pass at it what are the areas that we should consider - Web application, browser, that Portfolio 3? Any thoughts?

Ram Mohan: Steve this is Ram. I don't think there is - I don't think the first (unintelligible) necessarily. It's certainly the browser but it depends on what the users are utilizing, right?

Steve Sheng: Right. Right. Right.

Andrei Kolesnikov: Yeah. This is Andrei. Maybe we should call it the use case. Like who actually will be using the system because it can be robots or it can be real people. Ninety-nine percent of the real people use Web interface I think. I like the dog.

Julie Hedlund: Sorry.

Steve Sheng: Okay. So any other thoughts on Model 1 on its various impacts to registrars, registries, registrants and users of WHOIS?

Ram Mohan: This is Ram. Without trying to - I guess I - when I look at options one, two and three, my perspective is one seems to have the least amount of impact potentially to registrars and registries. But in so doing it may also have the least amount of use because local language display is purely optional.

And that may therefore not have the kind of universal utility value that, you know, we maybe looking to advocate. But having said that, I mean to me Model 1 is - it feels like it's the least common denominator option.

And, you know, down the road when we do comparisons that's something to keep in mind. It's the least amount of incremental work. It feels like it's the least amount of incremental work to add at least some basic support for IRB.

Steve Sheng: Thanks.

David Piscitello: This is Dave. It's also consistent with, you know, with the way that we do domain names for IDN. It's, you know, we've got the, you know, A and U labels. And, you know, it's, you know, in most cases it will be most valuable if both the A label and the U label are present.

And I happen to be going through like 2,000 (user) records at the moment for different projects. And I'm already noticing several registrars who are putting the domain name and underneath they have (puny) code or they have IDN.

And so they're already getting used to and prepared for having to support both representations.

Steve Sheng: Thanks Dave. So I'll take a note of that. Option 1 is the least common denominator. So should we go ahead and consider Model 2 and 3 or do we have other thoughts? Okay. So for - okay, let's move onto Model 2.

By the way, actually Ram would you - so would you suggest that we add an option saying that, you know, to require both the must be present language as well as the local language? Would that add more utility or does it...

Ram Mohan: Yeah. I believe so. I do believe that that is - in many ways, I mean that's kind of one of the founding reasons for our group. And it would be a pity if we - if as a group we don't arrive at that kind of a recommendation to have both.

Steve Sheng: Sure.

David Piscitello: This is Dave. In fairness, since Jay Daley is not on the call, I think we need to convince him or at least get him to weigh in on, you know, on how strongly he would object to that.

Because it was his perspective that pushing the problem to the edge and allowing every registrar to make a decision on what language would be most appropriate to store and have that information stored in the registries was in his mind, the only solution.

So we - I sympathize and actually, you know, have my own, you know, opinion that you really do need a common denominator. But, you know, Jay's argument is probably compelling to an awful lot of people.

Steve Sheng: Okay. This is Steve again. I...

Ram Mohan: So then Steve - this is Ram. That then makes the case that maybe you should add Model 4 which is what Jay is advocating which I will unkindly state as status quo.

Steve Sheng: Well I'm not sure Jay is referring to a status quo.

Ram Mohan: I know. I'm actually smiling as I say that, hoping that Jay will actually listen to it and get all riled up. He's a good friend. But we should probably, you know, list that option as well.

Steve Sheng: Okay. But let me - yeah, let me take that down and make sure that, you know, Jay's - what Jay's proposal is, actually is.

Ram Mohan: All right. That's fine.

Steve Sheng: Okay. So for Model 2 - so under the model, registrars will provide a point of contact to deal with translation issues. And Andrei was saying that could pose higher costs in terms of personnel, in terms of support costs. Right Andrei? Higher costs than in terms of just simply investing in technology.

Andrei Kolesnikov: That's correct.

Steve Sheng: Okay. Are there any others from an operating at registrar would like to comment on? No? Okay. I think there are a few more members in the working group that are operating in registrar. And I will also get their feedback on this offline.

So looking at Model 2's impact to registries Ram or Edmon actually, I was wondering if you have thoughts on the Model 2's impact to registries. Or if, you know, whether we have characterized it, the impact, correctly.

Edmon Chung: This is Edmon. Sure. I think - I guess generally yes, that I - you mentioned that after administering the domain the registry needs to know the - all the different languages.

I don't - I'm not sure, you know, if - even if there are other - even from Model 1 or Model 3, you know, how is that - how would that be different if they need to deal with the part where it is the UTF8 data.

So right now most of the cases registries would pass it onto registrars to deal with at least on the GTLD perspective. I'm sure these GTLDs maybe a little bit different. But I just - the GTLD perspective once the - most of the cases we need to work with the registrar together as well.

Steve Sheng: Okay. So are you suggesting that - hello?

Edmon Chung: I'm still here.

Julie Hedlund: Go ahead.

Steve Sheng: Oh okay. I don't know who - so are you suggesting that Model 2's impact to registry is the same as, you know, Model 1's and Model 3's impact to registry? Is that right?

Edmon Chung: In a way, yeah. In a way.

Steve Sheng: In a way. Okay. So when we - so just correct us - correct me if I'm wrong. So when we put together this grid we were thinking okay, so for registry for example, maybe a dot Asia, you have registrations for all over Asia, for example in China, in Japan, you know, for India, even from Pakistan.

Where, you know, registrants from these countries, register a dot Asia domain. And then - and for that domain they can input any of these local languages as a registration record.

This is different in my mind from Model 1 because in Model 1 the registry always has an option saying, you know, we want to require US ASCII. So that standardized - that greatly reduced the set of languages to one.

But that's my understanding. But what you're saying is, you know, that's not correct.

Edmon Chung: What I'm saying is that sure, you can operate with - I mean it seems to me that your assumption is that we could offer it based on the US ASCII data is what you're saying.

But in - I guess what I'm saying is that in reality if you look at the - what I was saying is that if you - if the registry needs to work with the multilingual data then it really doesn't matter.

Because even in Model 1 or Model 3 there maybe a situation where you would be - you would have to operate with - operate the multilingual data. Because the task at hand may not be, you know, trying to use that information to send an email, you know, or contact the registrant.

Steve Sheng: Okay.

Edmon Chung: You know, if you only assume that the task at hand is to use that information to contact with registrant then - or the contact. Then maybe that's true. But there are - the other reasons for using the data and - for all those situations. And it's not different.

Steve Sheng: Okay. I see. Okay. So we will - I will modify that. That's an interesting thought. Ram, do you have more thoughts on Model 2's impact to registries?

Ram Mohan: Yeah, I - this is Ram. Yeah, I have a somewhat - actually someone was talking about the different comments than what Edmon's comments are. I think the - one of the fundamental issues with Model 2 is that there is no default language. There is no default script.

It is - and there is implicit assumption that the registries are going to get registration (automation) in many, many languages.

But in practice what I expect will happen in a potential model to implementation is that registries will just wash their hands of any validation or verification of the validity of even the scripts of the languages coming in. They will accept whatever is given by the registrar and they'll just stick it in.

And if it makes no sense, if it is Cyrillic combined with Chinese combined with Hindi they'll take it, stick it in the database and then just show it. And I think we have a pretty high risk of larger amounts of garbage being stored than already is.

Steve Sheng: Okay. That's a great comment. Yeah. So what you're saying, in Model 2 it will actually increase data inaccuracy.

Ram Mohan: Well that's my intuition because it is not feasible for any registry regardless of its size, to really attempt to understand all the different languages that are coming its way. It might be more feasible for some ccTLD registries.

But, you know, even with - with very few exceptions, most ccTLD registries also deal with more than one language. But if you're a gTLD you're dealing with lots of languages. And the best option is to be - is to play safe.

And the way to play safe is not to make a guess at what language is attempting to be inserted into the database, to just take in whatever comes. And then that obviously leads to immense potential for gaming from those who would like to be hidden.

Steve Sheng: Yeah.

Avri Doria: This is Avri. Can I ask a question about - I guess I (unintelligible) see why there's a possibility of confusing information being put in. But I don't understand why it makes it any higher for people to put in garbage than if there were only one language.

I, you know, it seems like two indefinitely large sets that are pretty much identical. You know, wherever there's a registration someone can put in garbage. And the fact that you've got more scripts to do it with doesn't change the percentage of people who are going to put in garbage.

Ram Mohan: This is Avri. I think semantically what you're saying is right. But linguistically in a 7-bit ASCII world you cannot put in multilingual garbage into the system at all. You can try to put it in on the registry but simply reject it. In an 8-bit world you can just put in more or less anything.

You can put dingbats in and, you know, they'll just get accepted because it's just UTF8. You cannot really do that - so yes, semantically however you could put in Mickey Mouse in ASCII and you could put in, you know, Disney World as the address and that is garbage also.

Or you could put in, you know, ABCDEFG as the registrant name. That is also garbage. So functionally it's no different.

But in - just in terms of the variety of ways to confound somebody I think increases quite a bit with UTF8 because you could simply throw in all kinds of special characters, you know, including potentially, you know, hidden characters that do math and, you know, that do funky things that you cannot really do in 7-bit ASCII.

Avri Doria: Okay. So this is Avri here. But what you're saying is the number liars doesn't change. There's just more creativity for them to lie with.

Ram Mohan: That's right. More ways to lie.

Avri Doria: Okay. I'm not sure how much of a difference it makes but okay.

Owen Smigelski: This is Owen. Just to kind of point out as an example on this, I wish - I recently took over my company's Chinese domain names and noticed that over a period of years the number of individuals have registered my company as three different ways - in Chinese characters, in straight English characters and then also transliterated from Chinese characters.

And this is all legitimate but anybody trying to look or determine what was going on would certainly need a wild variety of ways to look at and translate that. So I guess without some sort of common way to enter data then, you know, anything goes.

Andrei Kolesnikov: It sounds like just the site (unintelligible) putting the garbage on the wrong information into the registry is a little bit different topic.

And it's not related actually to the, you know, to the mainstream of our workgroup because, you know, putting the best, you know, appropriate information into the database, you know, should be controlled by the different needs. Thank you.

Julie Hedlund: Thank you. This is Julie. I just wanted to let you all know that Steve Sheng unfortunately had to leave the call for another engagement.

He did want me to suggest to this group and we'll also suggest this in our follow up with our notes from this meeting, but if you could all take a good look at the matrix draft as it is now.

Steve has noted your various suggestions and comments. But if you have further comments as far as any inaccuracies you see in the language in the matrix if you could please send them to the list we would certainly appreciate that.

And of course we have about five minutes now so we can continue with this discussion as well. Any other comments on the ways that Model 2 can perhaps increase data inaccuracy. I know Ram and Owen and Andrei and Avri have all provided comments. Do others have comments?

Okay. Any comments in general on - any additional comments on Model 2? Probably - go ahead. Someone ready to speak?

Ram Mohan: This is Ram. In my - in looking at these three models to me, it feels like Model 2 epitomizes the highest cost with the lowest impact and the most amount of extra confusion. So that's how I would look at Model 2.

Julie Hedlund: Other comments on that? Do others agree or disagree? That's a helpful comment Ram. I appreciate it. And I've made note of it. Do you have any general comments you'd like to note on Model 3 Ram.

Not to put you on the spot but just in glancing through Model 3 any general comments from yourself and others?

Andrei Kolesnikov: Yeah. This is Andrei. I have a question regarding the transliteration service. If anybody knows that there are like - there are acceptance of transliteration between different characters exist in multiple languages.

We definitely have one for Cyrillic which really gets through how to transliterate between the Russian Cyrillic and the ASCII. So if members of this workgroup have this knowledge. Please share. Thank you.

Julie Hedlund: Thank you Andrei. This is Julie. So your question if I understood it right, is are there standards that we should be aware of as far as translating from other scripts into English. As you've mentioned, there are standards for transliteration from Cyrillic.

Ram, do you know if there are - do you know of any standards relating to translations say for inscription media and (unintelligible)?

Ram Mohan: This is Ram. I know of some but not any protocol defined standards, you know, that would work, say in a W3C standards track or IETF standards track manner. I mean there are local language institutes that provide, you know, specific guides.

But on the computer science side or the combination of computer science and linguistics I am not aware of something that is commonly available. I had a separate point that I wanted to make Julie when you have a moment.

Julie Hedlund: Please go ahead.

Ram Mohan: Okay. The - on Model 3 I think the - it makes sense to separate and be explicit about translation versus transliteration.

Now in the point it seems rather obvious but actually it's very meaningful when you get into linguistics because in some languages a translation can actually result in a different word or a different meaning in the must be present language than if you do a transliteration.

And so I worry a little bit about translation and/or transliteration being played - juxtaposed one to the other because I think the unintended consequences can be pretty big if you go for translation versus transliteration.

Andrei Kolesnikov: Yes. It is exactly different meanings. Thank you.

Julie Hedlund: This is Julie. Thank you Ram. That's a very important point. And thank you Andrei also for asserting that as well. It's I think one that we need to make sure we make clear in this matrix and also into any results we communicate out of this working group.

Any other comments with respect to Model 3? We can of course delve into it more deeply on the next call. But I was just wondering if there were anymore generalized comments.

Avri Doria: This is Avri.

((Crosstalk))

Julie Hedlund: Please. I heard Avri. If you might go first and then...

Avri Doria: Yes. Just quickly on your question - also with Hebrew while there is practice that probably comes out of linguistic language institutes in the country. I don't

know if there is being anything standardized for translation, automatic translation or anything or automatic transliteration.

Julie Hedlund: This is Julie. Thank you Avri. That's helpful. Who else wanted to comment?

Edmon Chung: Yeah, this is Edmon.

Julie Hedlund: Yes, please.

Edmon Chung: I have a couple of general comments actually. One is that there seems to be an assumption in terms of when registrar needs to deal with providing the information or translation into another form that the registrar is only dealing with a - I guess a local where they operate.

And - but the - I guess because of the nature of the business then they maybe getting registrations in many different places.

So knowing how to transliterate or translate between let's say Russian or Chinese, doesn't mean that you would be able to do the same when presented with a registrant coming from India, for example.

So having some sort of a model where - I guess that brings me to the second question or second comment - is what do you mean by transliteration? In a way, what are we trying to prove?

If we have a form whereby the registrant needs to provide to the registry and then for users for everywhere to use is that form - does that form have to make a lot of human sense or does it just have to be operational so that, you know, I can easily cut and paste and compare that data?

What I'm trying to say is that let's say you take an address and you say, you provide a "transliteration" of that and essentially its been some encoded form

in ASCII. Does that qualify as a form that is usable for the user who only can deal with US ASCII characters?

((Crosstalk))

Edmon Chung: Okay.

Andrei Kolesnikov: Transliteration - it's - how to say - it's a verbal translation. It's like you type the sound of the name like Andrei. You can type it in English - sorry, in Latin script or you can type it in Russian Cyrillic. It's not a translation of the meaning. It's how the word is sounded.

((Crosstalk))

Owen Smigelski: This is Owen. You say it's the phonetic pronunciation as opposed to what the word would mean in the local - in a different language.

Andrei Kolesnikov: That's right. It's a phonetic - I forgot the word. It's a phonetic thing. And it can be automated if in the country or the language there is an accepted standard on how to do it.

Edmon Chung: This is Edmon. Yes. I understand that. And, you know, I understand transliteration and, you know, there are, you know, transliteration has been standard for I think a lot of the - a lot of languages - Chinese, Japanese, Korean all have pretty standard transliteration or Romanization forms.

What I'm sort of asking is whether there is contemplation on any model or any possible model whereby a - the form of the - the form of a set of data is - because one of the things that would seem - we seem to be trying to solve is that yes, we want local users to be able to provide contact in their own language.

And then - but then we have a challenge for an international user - we don't know where they are right now - to use that data. And there seems to be an interest to provide that in a form where the international user can potentially use.

And that international user being in some form back what's compatible to what they expect today.

((Crosstalk))

Edmon Chung: Please.

Andrei Kolesnikov: Okay. I can give you one very simple example. My home address can be written in the Russian Cyrillic of course. We all use the Russian Cyrillic. But if you transliterate it, if you type it in Latin and send a letter to me of course, the post office will find my address and deliver your letter to my house.

Edmon Chung: I see. That might not happen for some languages then.

((Crosstalk))

Edmon Chung: And for address any type of automated transliteration or translation would definitely pose a challenge for Chinese I would know for sure. But I can't be sure for Japanese or Korean. So I guess that's a good point.

My question is whether we want the information to be so operational that you can put it on an envelope and then send it to the person. If that is the case then any type of (machine) translation won't reach that.

Andrei Kolesnikov: No. I was just giving you an example. I'm not saying that this will work for every language on the planet.

Edmon Chung: Right. So my main question is whether we need - whether that is a requirement. Is - if - for Model 3 or any type of model where we say, you know, we get two sets of data.

Do we need to be able to - must we be able to use that data and actually post a post address to the person, a snail mail to the particular person or is it just a set of data for, you know, supporting some sort of comparison usage or, you know, other type of operational usage is sufficient?

I'm not sure whether I'm making sense. Basically the question is when we - if there are - if we assess multiple formats of a particular address and we have a model whereby you must essentially provide it in a particular format, then is - how operational needs - does that data need to be?

Could I provide a transliteration in a sense that it's back and compatible to your existing systems but you might not be able to actually use that, put it in an envelope and send it to the person. Is that accessible?

Ram Mohan: Edmon, this is Ram. I think this is a really deep question that we should probably, you know, table for our next call. But I think it is a deep question because it points to what is the intended use and how we represent that data is dependent upon the intended use.

As you were explaining the problem I was imagining that my response would be vastly different based on whether I was a paralegal at an intellectual property firm researching something versus if I was an attorney who was going to sue somebody versus if I were somebody in law enforcement.

My answer to that same question that you posed would be quite different. So I think the usage, the expected usage model will have a significant impact.

The - but I think it's an important question that perhaps as a working group, we should plan to address and either put aside explicitly or take on explicitly

but not just leave it unaddressed. My point earlier was really to advocate against translation as a model.

Because I think much danger lies in that - going down that path.

Avri Doria: Ram can I ask, when you say translation you mean automatic translation? You mean translation by people? You mean translation by the person putting it on or do you mean any translation?

Ram Mohan: Yeah. I mean translation as it is represented in the matrix document there which uses translation interchangeably with transliteration.

I mean it in that form. In general I have trouble with translation unless there is some certification of some sort accompanying it which increases the cost, etc. Avri. Because automatic translations, you know, can result in more of a Tower of Babel rather than less.

Avri Doria: Okay, so - okay this is Avri. So you did mean mostly automatic and specifically in the notion of they then being done by some automatic tool, not in the version of Model 1 where someone would put in something but they would be translating it themselves.

Ram Mohan: Okay. I see where you - this is Ram. I see where you're going. Let me be - let me parse my language a bit more carefully. Automatic translation I would advocate very strongly against. Self presented translation I think can be taken on the honor system.

And I don't see too much of a problem with that because mostly people giving information, you know, one can assume that they have the right intent behind them.

But if you really - if we as a group are really looking for something that is reasonably bulletproof then we have to go for a verified translation model if

you are going to go for translation at all. So that's the precise parameters of what I mean about translation.

Julie Hedlund: Ram, this is Julie, and working group members, I want to point out we're now 12 minutes after the hour and I know people have places they need to be. I have noted Edmon's question and I think it's an excellent one for us to look at more carefully on our next call in two weeks.

And also to gather up any additional changes to the matrix and to discuss more fully Model 3 in detail. Is there anyone - anything anybody would like to add before we adjourn the meeting? Then I want to thank everyone for being on this call and particularly those that are joining us at a difficult hour.

We will rotate the time for the next call and we'll send (unintelligible) ahead of the call. And we will be sending around notes from this meeting in the next couple of days. Thank you everyone and have a good morning, afternoon, evening. Okay.

Man: Thank you. Bye-bye.

Man: Thank you. Bye-bye.

Avri Doria: Bye.

END