GNSO/SSAC
International Registration Data Working Group
TRANSCRIPTION
Monday 10 May 2010 at 14:00 UTC

Note: The following is the output of transcribing from an audio recording of the GNSO/SSAC International Registration Data Workign Group on 10 May 2010 at 14:00 UTC. Although the transcription is largely accurate, in some cases it is   incomplete or inaccurate due to inaudible passages or transcription errors. It is   posted as an aid to understanding the proceedings at the meeting, but should not   be treated as an authoritative record. The audio is also available at:
http://gnso.icann.org/calendar/#may
http://audio.icann.org/gnso/gnso-ird-20100510.mp3

All recordings and transcriptions are posted on the GNSO calendar page:
http://gnso.icann.org/calendar/

Present for the teleconference:
Edmon Chung – Group Leader
Jeremy Hitchcock - DYN-DNS
Rafik Dammak -- GNSO Non-Commercial Users Stakeholder Group
Bob Hutchinson, GNSO Commercial Stakeholder Group
Ram Mohan - Afilias – SSAC Board liaison
James Galvin - SSAC, Afilias

ICANN Staff
Julie Hedlund
Steve Sheng
Dave Piscitello

Absent apologies:
Avri Doria – NCSG

Woman: The recording has started.

Julie Hedlund: Thank you very much. Good morning, good afternoon, good evening to
                   everyone. This is the Internationalized Registration Data Working Group
                   Meeting on the 10th of May.

                   Thank you all for joining. And this is Julie Hedlund. And I'd like to give a roll
                   call for those on the call at this moment.

We have our two co-chairs Edmon Chung and Jeremy Hitchcock. Also joining us are Jim Galvin...

Man: Ram.

Julie Hedlund: And we have Ram Mohan, thank you, Bob Hutchinson, and Rafik Dammak both who currently are off the call but I think will be rejoining shortly.

From staff we have Dave Piscitello, Steve Sheng, and myself. Have I missed anyone?

Thank you very much.

Bob Hutchinson: Hi Julie. I just dialed back in. This is Bob Hutchinson.

Julie Hedlund: Hi Bob. Great. I think (unintelligible) on mute (unintelligible) on the call but we're still getting an echo. Okay. Thank you very much Bob.

Everyone, welcome. I wanted to note that we've had some very helpful discussion on the list with respect to an issue that I think there is some further discussion. And that is with respect to transliteration.

And in particular there's been some discussion on how transliteration might be handled and whether or not there are tools that could be used.

And Steve I wonder if I could ask you just to repeat the questions and maybe if you could summarize a little bit of the discussion around this topic that's been on the list?

Steve Sheng: Yes, sure Julie. Thank you. I think I asked three questions on the mailing list. The first question is I asked the working group if they are aware of any

programming libraries that can transliterate most of the world scripts into a US-ASCII.

The basis for this - the reason that I ask this question is because in one of the proposed models that we are considering model three, we ask that registrars transliterate the registrant's input into US-ASCII.

And there were some concerns that this might be too costly. So it's not really feasible. So I want to ask since - I want to ask are there any programming libraries that can do this automatically so that, you know, it's cost feasible? So that's the first question.

And the second question is if the answer to the first question is yes then I want to ask can these programming libraries and APIs be integrated into the registrars back end system easily?

So the first two questions essentially together basically asking, you know, is it possible to do this, you know, relatively with a low cost?

And the third question I think is more of a policy question is would the transliterated results, you know, along with the local script. So what I envision is in the WHOIS you have the local script in there and then also the transliterated in US-ASCII for that local script.

So the question is would that be enough use for registrars and users of WHOIS? So those are the three questions that I ask and the rationales behind those questions.

Julie Hedlund: This is Julie. Thank you Steve. And I know we had several different responses. Was there - did you get any sense that we were getting a consensus in the responses? There seemed to be a variety of responses to the questions?

Steve Sheng: Yes. So in terms of the first question a couple of people had pointed me a couple of libraries. So that one library is the Google API for transliteration that used to transliterate Web pages. This is a Web API.

There is also - a Microsoft has a translation utility in several Java applets and Window application for transliterations.

So there are some Java APIs for transliterates. And in terms of whether - I'm not certain. You know, some working - a couple working group members says answer to the first question is no.

But so for example, (Andrea) said no but if ISO or accepted locally tables exist it can be done easily.

So I'm not a real technical expert on that area so I'm not sure whether that means a firm no or there's a possibility there. So I'm - I'd like to see (that) call...

Rafik Dammak: Sorry?

Steve Sheng: Go ahead.

Rafik Dammak: It's Rafik. So I think for the - such API (depends) to the language. I think I know one. It's the (unintelligible) ATAs for Arabic is more user (given) than the Google ATA.

And I will think (of that) with the guide and company developing such tools. And my understanding that is not stable (unintelligible) and such implementations (depends) to the language.

So I'm not sure that we can have a uniform solution for all language. And I'm not sure how we in this working group we can deal for each language if we can find just one solution.

Julie Hedlund: This is Julie. Was that Rafik?

Rafik Dammak: Yes, yes.

Julie Hedlund: And if I could remind all of you for the recording and the transcript to please announce yourself when you're going to speak. But thank you Rafik. That's very helpful.

I wonder if I could encourage other members on the working group to comment on this. In particular I think Rafik has raised an important point. And that is that while there may be some various APIs out there it's not perhaps clear that there is a uniform solution.

If there is no uniform solution I guess my question to the working group would be is it useful still to have a recommendation relating to transliteration even if there is no uniform solution?

Edmon Chung: This is Edmon. I I'd like to, you know, sort - I understand that there are the three questions. And I think, you know, all of it actually relates to the third question which is what we're going to use for it.

And, you know, that's sort of the (pop C) question I would say. And then we should look back at what "transliteration" really means.

And I think I brought this up a couple of times before. The question is whether - at least for in my mind, the question is what do we want to use it for, who would be using this information, and then decide what type of transliteration or any type of representation in a let's say ASCII form would suffice for the policy matter?

Because the reason, you know, if I give a more specific example is that let's say a Chinese address as an example, the question is, you know, if - do we

need an ASCII representation of which such that people who do not understand Chinese can at least try to do something with it?

But the question is whether we - I would need to be able to actually send an - send the mail to that address and reach the right person because the latter is much more difficult.

But if we only have some sort of ASCII representation for things like for example just comparing data, then that would be a, you know, entirely different proposition.

So, you know, if it's somewhat needs to be able to be actually used by humans to send a postal mail then we have a much bigger challenge.

And also then we would have to look into what types of standards or transliteration or Romanization is necessary.

But if it's only for some sort of comparison so that at least we can operate with the data then the requirements might be quite different and we might be able to tolerate different registrars using different and even different registrant's using different ways of transliterated - transliterating data.

Jim Galvin: So this is Jim Galvin. I agree with Edmon. And let me just characterize the question a little bit differently.

While I like the idea that we should provide some transliteration there should be - it'd be nice if there was a baseline and we can suggest and recommend that people to provide information at some baseline.

I struggle with how to make that a requirement and more importantly what the motivation is for making that a requirement.

I think we have identified a number of issues with insisting on transliteration and on insisting on things. You know, we obviously can't continue with things being required to just be an ASCII. You know, I'll just turn that around. I mean there are issues and then there are requirements that come up which by themselves create an issue.

So I think I'm just struggling with what's the motivate - what is the end goal? What is it that we're trying to achieve? There's - I mean if we're trying to achieve internationalization then it's not clear to me that a requirement about transliteration is the right thing to do.

And I think we can still provide some advice about it. We're just not going to be able to provide something uniform.

Julie Hedlund: Thank you Jim. This is Julie. Any comments on that question from the work group members?

Well this is Julie. So Jim I heard you say that, you know, really the fundamental question is is it necessary to make transliteration a requirement? I mean and that is what exactly are we trying to achieve?

And perhaps we could provide advice but not, you know, not sort of any kind of mandatory recommendation.

You know, and I know we've been talking around this topic I think for several calls. Does anybody have a sense that we need to have a uniform recommendation or a recommendation concerning some kind of uniformity around transliteration? And is this even a necessary requirement?

Ram Mohan: Well this is Ram. I guess my perspective is that there are - we should recommend some basic guidelines, you know, kinds of information that ought to be made accessible but not requirements.

So they'd be - I would suggest making it more recommended to do a good best practice rather than a must do.

Julie Hedlund: This is Julie. Thank you Ram, that's very helpful. Other comments?

Jeremy Hitchcock:This is Jeremy. I'd probably echo similar comments to Ram just because the - if you have a number of entities that are all doing transliteration regardless of how it's done or how strong or how detailed the recommendation is for the most part scripts getting transliterated into ASCII representation are going to have differences.

You know, they might be as minute kind of on the technical side as upper case lower case sensitivity. And in terms of trying to use them for string comparison might turn out to be difficult.

Or if it's used for postal or location or identification mechanisms then the actual inherent script of the local language is going to be the more important of them.

So I'm not sure that the transliteration is going to provide much value. It would seem like the actual underlying characters are more important to capture and be able to be as represented in some form or fashion.

Julie Hedlund: This is Julie. Thank you Jeremy.

Steve Sheng: This is Steve.

Julie Hedlund: Please go ahead Steve.

Steve Sheng: So does that mean that you think - Jeremy, that you think model two would be kind of a better fit in terms of the three models?

So just a quick recap of the three models. The first model is require US-ASCII from registrant and optionally they may put registration data in their local script.

The second is no requirement for US-ASCII. A registrant provide registration information in whatever script that registrars except.

So that's the second one. And the third one is provide transliteration service that the registrar provide a transliterated results into WHOIS.

So it seems that the working group is thinking that transliteration would be an option instead of a requirement. So would that make a model two the recommended one or no?

Jim Galvin: So this is Jim. I have a question to add to this. And I apologize because I've missed the last couple of meetings and I haven't gleaned the answer to this question out of looking at the meeting summaries.

But there was a time when we were talking about providing this transliteration service. And I thought it was about registrars making that available in some way, not necessarily automated.

We seem to have moved into a space of this needs to be an automated service in some way. And I guess I'm trying to understand where that requirement came from or why that was the way to deal with it.

You know, I mean if they registrar's going to be allowed to present things in their local language for the local users and that's essentially the primary use of the WHOIS information at least, you know, in that place where does the requirement for this being automated come from? And is that - can we make that distinction in your particular list of questions Steve?

Is embedded in your question the requirement that it be automated or are we allowing for an option that says it can still be a manual thing that if somebody needs a translation they could reach out to a registrar and get some help in getting that done?

Steve Sheng: Sure Jim. So in model two we have a registrar provider point of contact to deal with this translation issues.

So that's - what you're mentioning is that scenario is covered. And what we - the model three considered is instead of registrar providing the point of contact the registrar does the translation or transliteration on behalf of the registrant. So that was the original proposal.

And then after I put that in the working group deliberations basically the working - there's a couple people in the working group felt that translation, manual translation is way too expensive.

I mean currently the profit margin for registrars is already very low. And you want to - and if we want to impose a requirement on them to do the translation that might be not - that's not feasible for them.

And then there's another suggestion saying any system, IT system that developed will be much cheaper than having, you know, actual people maintaining or doing the translation.

So that's the - that's the direction we're thinking of maybe having some kind of an automated system would revive model three.

And I also recall in one of the working group discussions they said that translation is not really needed. It's only transliterations that what we really need. So I think that's currently how it comes to being.

Jim Galvin: But I guess that's your last comment there is exactly - well it's not exactly but that's kind of my point. And it goes back to what Edmon asked earlier and, you know, what I asked too as a follow-up.

                What is the motivation for this stuff? I mean in the early conversations about this issue, you know, law enforcement is a very (economical) example.

                You know, you're trying to research and find out who owns a domain name and learn something about it and where people outside of the region of where this registrar and the registration exists are trying to learn something about the domain. And that's the...

Steve Sheng: Right.

Jim Galvin: ...purpose of the transliteration service.

Steve Sheng: Right.

Jim Galvin: If that's the only purpose of it, you know, how often does it happen? How frequently does it happen? You know, in - when we were talking about it in that context there was not expected to be a lot of transliteration. And so cost should not be a significant factor.

                So I'm trying to understand where this - because having missed some of the discussions I obviously, you know, missed something in here.

                But what was the motivation for requiring this thing to be automated and make this investment in transliteration?

Steve Sheng: Right.

Julie Hedlund: Now this is actually Steve, if I could just jump in. This is Julie.

Steve Sheng: Go ahead.

Julie Hedlund: I think Jim there might be a slight misunderstanding. We weren't as staff suggesting that this would be a requirement for automation.

It was a question that was asked because if we do think that there would be a need for transliteration then that - if we wanted to keep cost down or if we don't want to require a manual process the question was raised simply for the purpose of discussion as to whether or not there were automated systems that could do could do transliteration.

And I should point out that as staff we're simply trying to see if we can help move the discussion along. We're not suggesting, you know, particular requirements. We're really trying to see if there's any consensus around a particular set of recommendations which we've gathered into these three models.

But I think getting back to your main question and then this is really where we're trying to go with this is is there a need for a requirement for transliteration if it is something that would only be used seldom or is this just something that we could leave off the table as far as requirements but have as Ram had suggested and you have suggested some guidelines relating to transliteration?

Steve Sheng: All right this is Steve again. So I think Julie summarized what I want to say, the first part very well that we're not trying - we're only raising the question for discussion purpose.

Are there any one from law enforcement on the line that perhaps - any law enforcement, IP?

Dave Piscitello: But this is Dave. Can I interject on - I don't want to necessarily say on behalf of either of those constituencies...

Steve Sheng: Okay.

Dave Piscitello: ...but as an observation.

Steve Sheng: Right.

Dave Piscitello: You know, Ram had suggested that perhaps transliteration would be something that we can encourage as best practice.

Unfortunately the people who are most likely to ignore best practice are probably the people who, you know, who could be, you know, could serve as a locus for malicious domain registration.

And so the situation that I believe that it's someone who's involved in anti-spam or anti-phishing is concerned about is one where it is, you know, it is impossible to act on a malicious domain because the information that is acquired from WHOIS is not useful meaning it's not something that is easily understood.

Now there are two things that you could imagine. One is that someone from an anti-phishing or anti-spam environment would like to be able to rely on the registrar who has actually created the registration to provide him with some assistance.

The second is that, you know, as long as he has, you know, has UTF8 he can go and find something like, you know, Google, you know, software or one of the other software that we've talked about on the list and do, you know, a best guess transliteration.

So as long as - so one way to look at this is that as long as is there is enough in the - in the WHOIS records to make it - to make the transliteration

somewhat precise or somewhat accurate then, you know, it really doesn't matter whether the registrar is involved in transliteration or not.

Rafik Dammak: So if...

Julie Hedlund: And who is that? Go ahead.

Rafik Dammak: It's Rafik here. Just I was wondering about something you talked about like Google (API). It's - but maybe it's (unintelligible) can raise a problem of privacy. We don't know how Google is managing this kind of data.

Because API we have the type also. I remember there was kind of (unintelligible) one day Google provided (unintelligible) how we can (see the) - how we can type in there for Japanese in Japan.

And some people was wondering if Google is - can collect the data when we use that (tool). So maybe we need and open solution (unintelligible).

Julie Hedlund: (Unintelligible) you're up.

Man: That's different circumstances. And in the case in a registrar I would not think that a registrar, especially one who has to deal with local privacy regulations would just simply type into Google, you know, public app and do a transliteration there.

And certainly I agree with you you'd probably want to find some open source or licensed software that you can rely on to set the parameters for privacy.

But I was referring more to somebody who's, you know, like Joe St. Sauver University of Oregon who does a WHOIS query gets back something unintelligible and, you know, the fastest most convenient tool is a Google app that allows him to enter the UTF8 and see something that looks like English

to him so he can then find the registrar or find the registrant and make a, you know, a best effort in contact.

Ram Mohan: This is Ram. I'd like to be in the queue.

Julie Hedlund: This is Julie. Ram go ahead. I think you're the only one right now in the queue. Do we have any others?

Steve Sheng: This is Steve.

Edmon Chung: Edmon as well.

Julie Hedlund: I heard Ram and then after Ram was before Edmon. I didn't hear who that was?

Steve Sheng: Steve.

Julie Hedlund: Steve, okay. Ram, Steve, Edmon please.

Jim Galvin: And Jim too on your list. Thanks.

Julie Hedlund: Great I've added you too Jim.

Ram Mohan: Thank you. This is Ram. Three specific things I want to comment on. First is I am wary of unfunded mandates. I am quite concerned that if you mandate that transliteration must be provided and we depend on what is currently a free tool, you know, Google or something like that. Tomorrow that could change.

And we are then adding a burden, adding a cost burden, you know, across the spectrum which eventually will get passed on to the end user.

And so if you're going to do it then we should then consider the economic impact and ensure that we provide some funding for the mandate as well. That's one thing.

The second is that as we all know all transliteration is going to be a best guess methodology anyway.

I mean it's not going to be really accurate in the real sense of the word. You have to know the language in order to actually be accurate.

So the reason why I'm suggesting that it remain a guideline and not a mandate, the second reason is because it is inherently a flawed process and inherently a - an accurate and potentially invalid process.

I worry about creating a rule that by its very definition is going to be broken from the very first day of implementation.

The third is that we're really talking about the world languages. And there is a ton of them out there. And I would suggest some caution in how expensive we want it to be.

If you want to make this happen two or three prerequisites it seems to me are required. First of all the existing WHOIS Port 43 mechanism, that protocol so to speak has to be replaced.

That's a - that's got to be a core prerequisite recommendation with some teeth to it that it actually - that conversion happens that - because that's the only way you can actually accommodate the UTF8 representation.

But let's just keep in mind that, you know, if you want to have (Ozo By Johnny) automatically transliterated into something and no API anywhere has converted (Ozo By Johnny) into some other language you're pretty much going to get stuck anyway. So those are my three thoughts. Thank you.

Julie Hedlund: This is Julie. Thank you Ram. Next in the queue is Steve.

Jeremy Hitchcock:Hi. I'd like to be added to the queue at some point. This is Jeremy.

Julie Hedlund: Thank you.

Steve Sheng: So we - I know we've been talking about law enforcement or IP community. You know, it seems to me they're the main users who benefit from a transliterated WHOIS.

I think, you know, the larger picture is also they're also, you know, they're users of WHOIS other than IP lawyers and law enforcement.

And these often reside in different countries and do not know each other's language. So in today's world, you know, since English is more of the common language if it's in US-ASCII there are more likely to reach out to the other registrar in a different country.

If we require, you know, any language goes into WHOIS it might make them harder to contact.

So this is one consideration where, you know, we kind of put them backward compatibility, usability of who is in the IDN world. So that's one thing.

But I hear Ram saying that the existing Port 43 WHOIS has to be replaced to accommodate a UTF8 representation.

Ram can you - I was wondering if you could elaborate on that a bit because that seems to be a very important issue?

Ram Mohan: Should I respond Julie?

Julie Hedlund: (Unintelligible) with the queue.

Ram Mohan: Okay this is Ram. Steve very quickly, Port 43 only supports ASCII.

Steve Sheng: What about in Canada that, you know, (Jay) has been raised like flagging character set in input?

Edmon Chung: Steve and Julie actually this is Edmon. And my comments and questions like bleed right into this.

Julie Hedlund: Okay perfect. Go ahead Edmon.

Edmon Chung: So I think a couple of things before I jump right into this is first of all Steve and I think Dave earlier mentioned about some - and some way of law enforcement and IP to somehow have some intelligible data.

And I think, you know, that argument is kind of interesting. But I don't see the difference between law enforcement or IP trying to dig through the data using Unicode code points versus ASCII.

Because really in terms from - at least from my perspective and from people I've worked with a lot of it is about pattern matching. So you would pretty much use similar tools to do pattern matching based on ASCII data as well as, you know, it could be used on Unicode data as well.

Of course there would need to be some adjustments to, you know, to the tools that are being used.

But I think ultimately if you say "the bad guys," you know, as much as they would also put garbage data for the - for any type of whether it's in ASCII form or English or IDN or, you know, multilingual form, that's going to be the case regardless.

So it's I don't think that is as much a argument if you look at it in terms of providing a particular language or transliteration for the data.

But instead I think this concept about backwards compatibility is may be something that we want to, you know, look into.

I think Ram made a good point in that the WHOIS doesn't - at least Port 43 doesn't allow captors beyond the - I think it's 7 bit ASCII in fact.

So even the flagged type of thing you would have to have a format, you know, that - I mean the question is whether we want to have a format that is backwards compatible to existing WHOIS tools?

And that might be something we want to look into rather than sort of transliteration. So perhaps if I may sort of suggest a fourth model whereby it would be - so the registrant or registrar would provide the data in local languages and then but then they would also need to provide some sort of backwards compatibility for Port 43.

Whether that's something that we want to make as a requirement rather than transliteration and so called backwards compatibility, that type of data may be really only used by machines rather than human.

Because I think and coming back to the law enforcement and IPI, they - and in fact the general users, there might be a lot of tools already out there that are - that is sort of parsing the WHOIS data that, you know, it might actually be useful for some sort of backwards compatible format.

And that format might not, you know, might be information that in ASCII but let's say a stream of code points in their code point numeric format for example.

That actually helps a lot of machines to be able to read that information without having a lot of changes that provide some sort of backwards compatibility. But it doesn't put much of a strain on in terms of resource for having to do transliteration or any form that's more human readable.

Julie Hedlund: Thank you Edmon. I have two others in the queue. I have Jim Galvin and then Jeremy Hitchcock. Jim?

Jim Galvin: Okay thank you Julie. I have three points. First I don't think there's any question the transliteration is a good thing. But I still keep coming back around to the question, you know, what is the goal? What is the problem that we're trying to solve?

I mean as I understood it, it was about use of the information out of the region for whatever definition of region that you want to offer.

I mean law enforcement is the obvious economical example but certainly there are others.

And the question always comes back to what is the size of the community? And to go to the point that Ram was making about where are we putting the burden here when we say we have to have this service?

I - you know, Dave made the observation that as long as it's possible to transliterate it, then maybe that's sufficient and that (shouldn't) be our focus here.

In fact actually so that brings me to my second point. I really think that we want to focus on the representation of the data, not on transliteration in particular.

So focusing on the use of UTF8 in the presentation of WHOIS material, that takes us back to Ram's suggestion about, you know, maybe we recommend the items that are of general utility.

And insofar as one wants to offer transliteration or you know, you want to ensure broad usage of your data, we make a list of that bit of information which is core information that everybody needs to have, they need to be able to offer it out in a UTF8 format if that's what we want to go so that people who need to use it can develop their own applications and find their own way to transliterate it if that's what they need.

Presumably if the WHOIS information is for primarily for the use of the region, transliteration is not an issue. It's for those outside of the region and they should have the burden of solving that problem for themselves.

Brings me to my third point which ties into with what Edmon was just saying about providing backwards compatibility, I make the following observation.

The transition of electronic mail from US-ASCII to 8 bit covered this in quite a lot of detail.

You know, one of the requirements for the introduction of MIME when it first came out, the Multipart Internet Mail Extensions complete backwards compatibility with 7-bit ASCII mail.

Whatever they did in order to introduce 7-bit things had to have 100% backwards compatibility. And I really think that that's possible even in this situation.

It may be that we want to change ports and it's not Port 43 anymore. But I think that from a technical point of view there is a means to say you can still use US-ASCII on that port. And oh by the way if your - you know, if you can

understand other things here's the mechanism by which you can have 8-bit sequences coming across there or use UTF8.

So I don't think that this is a hard problem. If it could be solve for email it can certainly be solved here. Those are my points.

Julie Hedlund: Thank you very much Jim. This is Julie. Jeremy you were in the queue?

Dave Piscitello: Can I ask - this is Dave. Can I ask for a clarification on what I think I heard Jim say?

Julie Hedlund: Go ahead Dave.

Dave Piscitello: Jim did you say that the onus is on parties outside the region to solve the problem of not being able to recognize the local language in which the registration was submitted?

Jim Galvin: Yes I put that out there as a model in all of this. You know, rather than putting the burden on the registrars to provide a service for those outside of the region those outside the region should bear the burden of that service.

And that was in part in response to your suggestion that if there are free services like Google and other kinds of things they have a means to at least start the transliteration process.

Dave Piscitello: So what I mean just as an observation, the end effect of that is that everybody has to solve it for every language except their local one, right.

Because ultimately almost any registrar is going to get, you know, have some occasion where there's going to be some language, you know, that, you know, of a - you know, that a domain submitted in that, you know, that is not recognizable, you know. And they're going to have to display it through their WHOIS.

Jim Galvin: Two comments to that. I mean I see where you're going with this and I agree with you. I'm not suggesting that I've got the 100% complete answer. But in the interest of the discussion about this, you know, there's always the question of choosing the base language that you want everything to go to.

So, you know, you want to put the burden back on the registrars to transliterate then now we have to have a discussion about what's the base language.

And then you get into the point that Ram was making about, you know, transliteration is not trivial. There's always issues. It's not a perfect process it takes a lot of experience and knowledge in order to make all this happen. So that's one point.

The other thing is I still question what is the size of this problem space? Who needs access to this transliterated information outside of the region and how big is that? And so I mean to me that backs into the economics of where do you put the burden?

Ram Mohan: This is Ram.

Julie Hedlund: Yes I...

Ram Mohan: Whichever (unintelligible) I'd like to get in the queue.

Julie Hedlund: Yes I hear Ram but I know - I - Jeremy did you want to - you were still in the queue and we haven't heard from you. Did you want to make a comment before we move ahead?

Jeremy Hitchcock: I would. I'll answer a question that was asked to me a while ago too. So just a couple of quick things and then we can get back in this discussion.

But I wanted to point out that the tools for anything has to be local, can't rely on an API that is, you know, available or unavailable - you know, is available today but may not be available at a later time.

And considering that we haven't really - I mean there hasn't been an exhaustive search but does not look like there are tools that are widely available for all the universal languages or all the languages that exist in the world. So that's kind of a concern as Ram had mentioned.

Now I think another thing that's important to point out is it has to be done kind of in a common way or in a comment fashion otherwise the value, you know, for whatever the value of transliteration is it's going to be pretty reduced.

To answer kind of Steve's question kind of the - in the sense of models, I have more faith that far flung registrars, you know, meaning non-American English or Western Europe English speaking entities, either registrars or registries can correctly represent their own script and their own encodings rather than relying on them to translate or transliterate to a particular standard.

So I think I'm more, you know, again personal opinion just based on the discussion that there's is more opportunity for registrars, registries, and registrant's to choose entities that know how to support their particular languages and requiring them or potentially boxing them out of participating with different service providers because they don't "support a language" I think would be something that we would may want to avoid.

So in that sense I guess it's the allowing registry and registrars to choose the support languages that are local themselves which I guess is the second model.

I do want to kind of go back to the WHOIS asking for the service. We've heard a couple of times that transliterating is useful. But I'm wondering about

the requirement to have a common technology platform to represent uniform scripts versus a common script to represent languages in a single format.

I mean I think that it - the distinction doesn't sound that subtle but it makes a huge difference of whether or not we provide a service on a particular port that can provide particularly any language and different types of encodings or we provide a single script that "represents" languages.

And just thinking of the sheer scale of the language that are out there that seems like a pretty tall order. And I don't think we have nearly the expertise to think about all the languages and all the representations and all the transliterations even if we're just thinking about translating to, you know, 8 bit ASCII.

I mean I don't even know if that's even possible and I don't think we have the expertise in this group to say whether that is or is not.

I'm sure that there's some language sets that just can't be represented in that format. And I don't know if we want to make such a huge recommendation or a huge requirement that some of the languages can't be represented. So, you know, just my bubbled up thoughts but I'll let the discussion go on and on the topic.

Julie Hedlund: This is Julie. Thank you Jeremy. Ram you're in the queue and Dave.

Ram Mohan: Thank you. Very briefly to Jim's question about how big is the demand? I think we should look at a different question which is when you need to get that data how urgent do you need it? And I think that is the real question here.

If you're - if you've got a problem with a domain name and you need to do something about it right away and you can't make out head or tail of what's in

the name or who owns it or who the register is then, you know, you're in real trouble.

So I think it's - that focus should be less about how many people are going to use it or how big the problem is and much more about how severe an emergency would it be.

You know, and I think that that should be another facet, not just the volume, the quantity, but also the urgency. Thank you.

Julie Hedlund: Thank you Ram. This is Julie. Dave?

Dave Piscitello: Well I'm actually sort of lost at this point. So I'll give up my spot in the queue. We've gone down too many tangents and I can't trace back to just what it is we're now trying to accomplish.

Julie Hedlund: Well this is Julie. Perhaps I can - I've been taking I think pretty close notes here and of course I'll summarize this conversation.

But I've heard I think some interesting and I think helpful suggestions. In particular I'd like to go back to a suggestion that Edmon has made. And perhaps we can discuss this and it was also something that Jim hit on as well. And that was perhaps what we have is a fourth model where the registrars provide the information in a local script and then also backwards compatibility to Port 43 or other WHOIS applications instead of a requirement or guidelines concerning transliteration.

And Edmon noted this could be ASCII and a code point numeric format that could be something that could be machine readable.

And I know that Jim mentioned also the backwards compatibility. And it might be possible but perhaps not on Port 43 but we might want to change ports.

I'm wondering if others have comments on this suggestion of perhaps a fourth approach?

Steve Sheng: Julie this is Steve. I wish to be added to the queue.

Julie Hedlund: Steve I think you are the first in the queue. Are there others?

Dave Piscitello: Actually now I'll come back in because now...

Julie Hedlund: All right...

Dave Piscitello: (Unintelligible).

((Crosstalk))

Julie Hedlund: ...for Steve and then (unintelligible). Go ahead Steve.

Steve Sheng: Edmon I was wondering if you could clarify more a little bit about what goes into the backward compatibility, like what data is part of the backward compatibility?

So to clarify, we talk about translation and transliteration. And those are only focused on the registrant's name and the contact address. All the other WHOIS data elements we talk about ways to internationalize them.

So for example, we talk about the registrars name has to be in US-ASCII. We talk about telephone numbers and we talk about email address. Those have its own standards.

So in terms of, you know, a backward compatibility for Port 43 I was wondering if you could clarify more like what data goes in there? Thanks.

Edmon Chung: Right. I guess very quickly it's really you understand there. It's we talk - we went down the path of transliteration. And I guess the idea is to have it in US-ASCII transliterated. Because you can also transliterate it into for example, Arabic characters or Russian characters right?

So that's sort of led me into the thinking about, you know, rather than saying transliteration that there should be a form where it's backwards compatible for - to Port 43.

And I think that it's overall rather - and some registrars could implement it using transliteration into US-ASCII. Some could be, you know, in different format. But that's really just it. I don't really have any idea on further details there.

Steve Sheng: Okay thank you.

Julie Hedlund: Dave?

Dave Piscitello: So I remember Jim had mentioned MIME. And the - so Jim the question I would have is is whether you would foresee the ability to (salvage) some convention where people would say, you know, I'm going to do a WHOIS query and I'm going to, you know, have some, you know, some indicator that says I expect MIME in my response as opposed to well I'm able to process MIME formatted data as opposed to purely ASCII (text).

Julie Hedlund: This is Julie. Jim did you want to respond?

Jim Galvin: Sure. You know, what I imagine could happen here in terms of completely supporting backwards compatibility to right, WHOIS protocol is I send in US-ASCII a token right, with a carriage turn new line at the end. And I pump that at Port 43 and then I get back a whole chunk of US-ASCII data, right?

So what we do is you simply say that you have a formatted string that's allowed to appear as the first few characters in that.

And that allows the querior to then say some things. And, you know, we could spend a lot of time discussing about what kinds of things they might say.

But you could even go so far as to say if I get this special, you know, front string of characters then I know I'm going to get a multi-link query and I have to read in all that query.

And that just means that on the response side I can then return a MIME object what the heck. You know, I'll start with content type multipart related and I'll dump at you all kinds of stuff. And then I get all of MIME to be able to come back to me.

And the only thing that's required on the server side is an 8-bit clean implementation and the ability to check queries for this magic token in the beginning that tells me I'm now doing extended stuff.

Dave Piscitello: Okay so I - that's what would - I expected you to answer. So what we would have at that point is another alternative to (IRIS) and (RESPUL) which would be MIME as a way to migrate from an existing, you know, highly limited WHOIS to something that would be - would provide a richer data environment.

Jim Galvin: Yes I mean the only benefit of this versus the other things -- and I guess I don't really want to create competition for other solutions -- but what you get here is you get to continue to use Port 43. Nothing that you're doing now changes.

You have complete backwards compatibility with the installed base and...

Dave Piscitello: Right.

Jim Galvin: ...those who know more get to ask for more and they get more.

Dave Piscitello: And that's what - I want to tease out those little details because those are the kinds of things that we probably should make the broader community aware that we've discussed.

                   So, you know, so this would be something that has a benefit and then comparing the use of MIME versus SML and the rest (unintelligible) would be something that again would be, you know, something that we could explore.

Jim Galvin: Hey MIME is just a framework. The data inside could still be XML.

Dave Piscitello: And in all those details are things that people are going to have to understand to know the, you know, the cost.

                   I mean Ram began, you know, part of this discussion on what was a - the beautiful euphemism he used, unfunded mandates.

                   So rather than throw an unfunded mandate out without any sense of, you know, the size of the task it's useful to say, you know, to say whether it's mandate or not here are the things you, you know, you have to consider in order to be able to support IRB and here are the choices.

Julie Hedlund: This is Julie. Thank you Dave. Thank you all. Are there more comments? We're now at two minutes past the hour. And I would obviously certainly encourage all of you to continue this discussion on the list. And I think as staff we can try to summarize this discussion and send some notes out.

                   But also perhaps we can - I think this last discussion in particular of a possible - another way of addressing backwards compatibility is one that we should definitely continue.

Any further comments before we finish up the call?

Steve Sheng: Hi Julie. This is Steve.

Julie Hedlund: Yes. Please go ahead Steve.

Steve Sheng: I think Ram raised a very important question. That is existing Port 43 has to be replaced to accommodate UTF8 representation.

I think this needs more discussion and this will have a far more implication than to WHOIS than any of the policy work that we have done so far.

And if this is the recommendation of this working group then, you know, this could have quite some policy implications because the GNSO is looking particularly to the expertise of the IRD group in one of their WHOIS studies about, you know, not able to display IDN characters.

Jim Galvin: So Steve so let me just respond very quickly to that issue. I mean I make the observation that SMTP when it was originally deployed right, the mail transfer protocol was 7 bit US-ASCII and that was it.

And yet we have MIME 8 bit now arbitrary content now being distributed and used over SMTP, okay.

So the - it replacing Port 43 is certainly an option if that's where you want to go. But I make the assertion that it's not essential in order to transition to an 8 bit environment.

Steve Sheng: Okay.

Jim Galvin: We have a working example.

Steve Sheng: Right. Jim it's not really that's where I want to go. It's where the community wants to go. I think that that's very - want to clarify that, yes.

Julie Hedlund: Right. This is Julie. I think that, you know, this discussion at the end of this - just this last discussion and what Jim has described and I'd like to see this perhaps laid out a little bit more on the list is this idea of, you know, having a solution that's backwards compatible to Port 43.

And if that is indeed a solution then to me that Port 43 does not have to be replaced. Am I characterizing that correctly Jim?

Jim Galvin: That's correct. Thank you.

Julie Hedlund: Any other comments before we close the call? Then I want to thank everybody for participating today. Thank you very much and to our two chairs in particular, Jeremy and Edmon.

I will summarize this call and get some notes out tomorrow. And we'll also try to see if we can continue the discussion on this idea of backwards compatibility and perhaps see if it might work into sort of a fourth model or a another approach.

And our next call will be in two weeks. And I would let everybody know too that we will be having a public forum at the Brussels ICANN meeting.

We don't have it scheduled yet but we will have if not preliminary recommendations certainly something to encourage participation across the community.

But our goal would be to try to have some preliminary recommendations to - because I think that would definitely encourage participation. So we will try to see if we can decide guide the discussion in that direction.

Thank you everyone. And the call, the next call will be in two weeks at the alternate time. And we'll send out notices. Thank you.

Man: Thanks Julie...

Man: Thanks Julie, looking forward to the conversations.

Julie Hedlund: Great.

END