

**GNSO/SSAC
International Registration Data Working Group
TRANSCRIPTION
Monday 21 December at 14:00 UTC**

Note: The following is the output of transcribing from an audio recording of the GNSO/SSAC International Registration Data Working Group on 21 December 2009, 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://audio.icann.org/gnso/gnso-ird-20091221.mp3>

On page:

<http://gnso.icann.org/calendar/index.html#dec>

Present for the teleconference:

Edmon Chung -- GNSO Registry Stakeholder Group, .ASIA

Steve Crocker, Chair, SSAC, Shinkuro

Rafik Dammak -- GNSO Non-Commercial Users Stakeholder Group

Bob Hutchinson, GNSO Commercial Stakeholder Group

Yao Jiankang, GNSO Registry Stakeholder Group, CNNIC

Erick Iriarte Ahon - LACTLD

Steven Metalitz -- GNSO Intellectual Property Interests Constituency, Commercial Stakeholder Group

June Seo, GNSO Registry Stakeholder Group, VeriSign

ICANN Staff

Francisco Arias

Julie Hedlund

Dave Piscitello

Absent apologies:

Jeremy Hitchcock

Avri Doria

Ray Plzak

Julie Hedlund: Thank you everyone. Good morning. Good afternoon. Good evening to everyone. On this call today, the Internationalized Registration Data Working Group call, on the 21st of December, we have Rafik Dammak. We have Erick Iriarte Ahon. We have Jiankang Yao, June Seo, Edmond Chong, Steve Metalitz, Steve Crocker and from staff we have Dave Piscitello and Julie Hedlund. Have I missed anyone?

Rob Hutchinson: Rob Hutchinson.

Julie Hedlund: Oh, I'm sorry. Rob Hutchinson. I'm sorry about that. Yes, we also have (Rob Hutchinson). Thank you very much.

So Edmond would you like me to read off the agenda?

Edmond Chong: Yeah, I guess it's probably good that you would actually take the lead on the agenda and run the meeting.

Julie Hedlund: That's great. I'll do that. Thank you very much, Edmond. So for your agendas today, we will continue our discussion on the scope of the working group. And perhaps then proceed to our discussion of the work plan and timeline and deliverables. Does anyone have anything that they would like to add to the agenda?

Okay. Then we'll go ahead and proceed. Perhaps what I could do at this point is remind everybody what a little bit of the discussion was from our meeting two weeks ago on the 7th of December. We discussed, we continued the discussion of the scope of the working group. Dave Piscitello spoke a little bit about the survey that he had conducted and of the international postal union requirements and that, you know, based on that a possible requirement could be displaying data in also a local script.

There was some discussion of the degree of difficulty for displaying data in local script and also there was a question as to whether it might be useful to survey some ccTLDs with some specific questions on how registries currently receive and display data, although there was a note that on the goals of ccTLDs might be different from those of gTLDs.

There was also discussion about what was in scope for the working group and Dave Piscitello noted that the working group would not necessarily have to suggest changes to the way data is collected today. And Edmond suggested it might be useful to prepare a checklist of possible requirements

for receiving and displaying international registration data. And to use that list to decide what is or what is not in the scope of the working group.

So I think that summarizes much of the discussion to this point, so at this point we would continue the discussion of the scope of the working group and perhaps we could start with this suggestion of preparing a checklist of possible requirements for receiving and displaying internationalized registration data. I'm wondering if those on the call have some thoughts on this point?

Dave Piscitello: Julie, this is Dave Piscitello.

Julie Hedlund: Yes. Please.

Dave Piscitello: I started to scribble down some notes and I guess sort of one of the obvious requirements would be some discriminator, some way to identify the character set that is being stored and would be displayed so that one could know that UTF8 is being used or ASCII 7 is being used or UTF16 or whatever.

And so I would thinking that might be one way not only to be able to identify any character set, but if there were more than one present distinguish among those present.

I also think that given the amount of requests that we've had and the amount of concern that's been expressed by law enforcement and other parties, that it would be useful for the, you know, for the information to be machine readable. And so again that plays to having strong identification and so the machine can say, okay I'm looking at this particular character set, now I understand how to interpret the individual characters.

Julie Hedlund: Thank you, Dave.

Dave Piscitello: The next thing that goes form having the ability to recognize the character set for individual characters would be discriminator that would allow you to identify or type the data so that you would like to know that a particular field contains a number and it should only contain a number like a fax or a phone. And maybe another way of typing data that would help prevent, yeah, prevent people from submitting illegal characters that could be used to subvert a Web interface for example would be to have a strong syntax so that data would be typed, and for example you wouldn't allow special characters in a contact name like an ampersand or a hyphen or a star or something like that. I guess hyphen would actually be valuable.

But you know, it would be a defined set that would be appropriate. I think the biggest concern could be to eliminate the kinds of characters that could be used to do a cross-side script, scripting attack or to emulate a command line on a PC or a Unix machine.

So I know I dug way, way down into the depths of that data. I'm sure that there are probably higher level of requirements, but I just wanted to kick us off a little bit.

Steve Crocker: This is Steve Crocker. The one thing that I wanted to offer up is that it occurred to me a while ago when we first started that this is a topic for which the W3C might have some expertise to bring to bare since they have dealt with user interface issues on the Web forever and ever.

I sent a query off to (Thomas Wressler) and opening up the subject and he sent me back a reply and I'll pass it along on the list. The short of it was, yes I'd be interested, but it would probably take some money to for them to get involved and I'm not sure where we want to go with that, but nonetheless I think the idea is that they must have considerable depth in this area that we're just beginning to think about. And there might be a positive way forward by engaging with them.

Julie Hedlund: Thank you, Steve. That's very helpful. And thank you, Dave. Other comments on either of these suggestions?

Edmond Chong: This is Edmond. I think the (unintelligible).

Julie Hedlund: Edmond, you're breaking up. We're not. Edmond?

Edmond Chong: (Unintelligible).

Julie Hedlund: Edmond, we're having a very difficult time hearing you. We're just getting just fractions of what you're saying. I think Edmond is going to call back in.

Dave Piscitello: Yeah, Julie, the phone, the conference is having a little bit of trouble. I got knocked off in the middle of, I think it was, Steve's discussion.

Julie Hedlund: Okay. All right, we'll I'll send a message to the operator, let her know we're having trouble. We can just wait a moment.

Steve Metalitz: I have a question for Dave. You seem to be opening up in coding to any ISOs mechanism, is that correct or am I misinterpreting what you're saying Dave?

Dave Piscitello: I'm saying that if that is something that is going to be something that would be a recommendation as a group, then it would be necessary to discriminate what coding you would be using. And, you know, if you remember from prior calls, one of the things that I have mentioned several times is that it might be interesting to consider that there would be a mandatory sort of base encoding and that there would be optional encoding to satisfy local user needs.

So in the model that I've envisioned and this is by no means an ICANN position or anyone else's position, but in the model that seems to make sense from the many perspectives that I heard, having something that everyone can recognize which would be the de-babblizing language, plus a local language would be very attractive.

If you do that, you need to be able to say, the de-babblizing language is this one over here and yet this other language that I've chosen to also record registration data for the convenience for my most common users is this one over here. Does that make sense?

Steve Metalitz: Yes. It does make sense. Let me plumb a little bit further in terms of your assumptions. I was wondering whether or not you believe there is any corollary between the domain name registration script and the script that's being used to record the address of the registrar?

Dave Piscitello: You mean the encoding of the, for example, an IDN name and the encoding on the character encoding of the contact information?

Steve Metalitz: Yes. So in other words there's apparently there's a lot of effort that goes into saying what script characters you can use for a particular language, okay. And.

Dave Piscitello: I see. So there's two levels to that. Tow levels to answer that Steve. One is that, just to make absolutely certain, I'm not advocating that we use a label for encoding for information that would capture registration data other than the domain name. So I would not imagine that what we would have is something that would provide a extend dash, dash gibberish that translated back to Dave Piscitello if I were the registrar.

So that's the first clarification. The second clarification is that you're asking would it be appropriate for someone to submit their name where David contained a D that was from the US ASCII 7 character set and an A that came from a Cyrillic set. Is that what you're getting to?

Julie Hedlund: I see that Edmond has joined us as well. Edmond, I know you were trying to comment before when we weren't able to hear you. Did you want to comment?

Edmond Chong: Sure. Sorry. Am I (unintelligible) better now?

Julie Hedlund: We can hear you fine now. Thank you.

Edmond Chong: Oh, okay. Sorry about that. I was actually I think it was Steve getting some participation from W3CS. I think that should definitely be a good idea. I was going to respond also to the idea about character set and I dropped for a while and re-joined, so I wasn't sure what was asked in between, but I think we came back to the issue of sort of character set.

One thing, well two things about it, one is that besides having some sort of a tag, it's probably good to - probably important to consider whether we want blocks of data in different encoding. And how we encode it is probably something that might need to be talked about later as well, but maybe one of the questions is whether we would have blocks of data rather than sort of a tag for each field of data, whether that might be something we might want to consider.

The other thing is I sort of caught the discussion when I re-joined about the domain name, the language tag for a domain name versus a character set or a language tag for registration data. I think it's probably not useful to try to guess the language or the character set, but rather we should have specific steps, I mean, we shouldn't guess the language of the registration data with the language of the domain registered, but rather we should really, you know, think about it as potentially that they may differ even though that perhaps in many situations they could be the same, I don't think it's the best idea to restrict it to that way or make guesses based on the domain name itself.

I probably missed some of the discussion, so that might not be valid at all.

Dave Piscitello: This is Dave Piscitello and I dropped off again. I apologize. I'm not certain why I'm having so much trouble today. When I was talking about the domain

name encoding I was essentially advocating that we already have a standard for that field and there's no need for us to be venturing into that well-trodden space.

Because of the way A labels are encoded it's not even necessary to know what language the domain name happens to represent. And I agree with you that I should be able to as an English speaking person register a domain name in Chinese if I choose. So I don't think there should be any restriction on the character set that is used by a registrant to compose a domain name other than the existing restrictions that IDN guidelines provide. Does that make that clearer?

Edmond Chong: It probably does. I just, there's this one area that I think as we go down this discussion, I think it's probably important for us to distinguish the screen when we talk about language or scripts versus when we talk about character sets is they are probably very different things. Although there are some sort of overlap there. But it's probably better, you know, if we be very clear in terms of say, you know, when we talk about character sets versus when we talk about the language.

Dave Piscitello: And I agree with you. I am not the IDN expert here, and I've always differed to (Tina) here or one of the other people who have spent so much time in the IDN space to help me with the formal and correct language once I have, you know, once I have something that I'd like to say and write down. So I think if there are places where you can actually change or educate me I'm more than happy to learn what those differences are. They just simply escape me because I don't have that much experience in using characters other than extended Roman for Spanish.

Steve Crocker: Let me ask a question here Dave and Edmond, this is (Crocker). If for domain names typically I would think you would want them in a single script, that is you can look at it and say this is all Greek or this is all Mandarin or something. But for a WHOIS registration, I can easily imagine that you might

have a name that's in one script and then it's location, say the address information, to correspond to another script. So you may want to have, for example, a Chinese name that is then operating out of say, Korea and that you might want to use different scripts for those two parts of the whole entry.

I'm making this up, so Edmond you're probably a lot more, a lot closer to this so you can comment on whether or not this makes any sense at all.

Edmond Chong: Yes, I do think this makes sense. I mean, that's why I brought the issue that the language of the domain that's being registered may not be the language of what the registration data, the contact data. Or you know those kind of things may be - so maybe one of the things that we might want to consider is how we capture that and whether we do capture that or not or whether we.

The initial discussion was that started this discussion was the suggestion that at least there should be some sort of a tag for character set. I think that's probably useful, but before we go into that maybe a higher-level decision is, you know, whether we might have, flocks of data or rather that we simply tag certain amounts of data. That was really the jest of my point.

Dave Piscitello: There is actually a precedent for that in the data escrow requirements for ICANN gTLD registries. If you go and you look at like Appendix 5, I was just looking at it this morning, that's why I know this. In the dot arrow agreement, I think it's present in other agreements, there is a description of the XML document components or that are required for escrow, and in fact, there is like a registrar chunk or block and then there is a contact information block and there's a registrant name.

So as you say it's very possible to not have to necessarily strongly type every single individual element of the contact information, but if it were appropriate and if it was necessary to consider Steve Crocker's scenario then in fact you could do that in the data structure that you create.

Man: I guess the point that I was trying to make was a little subtler that you can point in that block, you can point to either some character set representation a tag of that nature or you can attempt to leverage the work that's gone into creating scripts in IDN. Not at the (unintelligible) level, but at the level above that that basically says these characters are allowed to be used in this class of languages.

And my question was, is that relevant to this, okay and I don't really know, I'm curious at this point what you know, Dave.

Dave Piscitello: Well, I mentioned at the outset that I thought that it would be valuable to actually consider excluding characters that are known to be used primarily for, you know, for malicious purposes. I don't know that, for example, if I went through the ASCII 7 character set, I don't know that I've ever seen a name composed with an asterisk.

However, I know that an asterisk can be used for wild-carding purposes for a malicious insertion of data in a Web script. So is that what you're trying to sort of exclude in the characters that you're talking about?

Steve Crocker: Well, you know, just a reference to a character set okay, is a pretty wild reference. There's a lot of character sets out there that have been used over time and they're, like you said, there's no constraint on it. You say, here I'm using this character set and I can use any character out of that character set with this tag.

That's one approach and it has its pitfalls, okay? And I basically was saying, I was curious as to whether or not, in order to discriminate characters within a character set, you would impose or utilize some sort of a coding that leveraging the already the work that's going on with IDN and that was just my question.

It seems like it might be appropriate, but maybe not. Maybe Steve Crocker's example is sort of useful although the other piece that I don't know is that I would think that all human names are representable in the character sets that IDNs are being done. I don't know of anybody who would say, oh in Arabic you couldn't represent this person's name therefore you have to use the whole character set.

So I'm assuming that whatever is adopted for IDNs, is a super set of whatever names that you would need to get described. Maybe that's not a (unintelligible) assumption.

Edmond Chong: This is Edmond. I think I finally got the question. Sorry I missed it earlier cause I dropped off. I think that's probably not the right assumption. Even if you, for example, take Roman based languages. The IDN characters, well I should say the repertoire of characters acceptable for IDN is really a small subset of characters that are being used for a particular language. Even for names such as in French or even English. There would be apostrophes or ampersands or different types of symbols, which would not be encoded into, well allowed into IDNs and thus the encoding methods there in probably is not the most appropriate to use.

And there are some other instances for other languages as well. So if specifically on with which we're just talking about an over-arching sort of encoding or everything and contemplating on whether the Unicode mechanism if you will is the best way to go about it, I would say probably it's something we might want to stay away from because the whole premise of IDN when it was designed was to really make it a subset rather than a superset.

Jiankang Yao: This is Jiankang Yao. Since IDN is based Unicode, Unicode has a defined all the languages characters. So Unicode almost all the languages (unintelligible). If we (unintelligible) it's necessary because IDN is based on Unicode. Unicode ready defined has a unified character set include every

character set. IDN has a (unintelligible). Another is UTF 8 based Unicode encoding. So our international (unintelligible) data will be have (unintelligible) first English based data for example A label.

Another is for U label. U label play some local languages. For registration or registrar data there may be source data in UTF 8 or local language, but if it is display pool the information for the customer or user they may first language is English. Another is UTF 8 character. Since UTF 8 is one kind of form of Unicode, so Unicode include every character. So I think this can solve the problem. Thank you.

Julie Hedlund: Thank you Jiankang Yao (Kong). This is Julie. Do we have other, some additional comment son some of these suggestions?

Edmond Chong: This is Edmond. One suggestion is I think all the discussion was really constructive and useful, but perhaps, you know. really to going forward, it's probably more useful to have some framework for discussion and I think that's sort of the next part of the agenda we talk about the timelines and approaches for really bringing the discussion to some sort of structure. So that sort of what I'm thinking.

Julie Hedlund: This is Julie. Thank you, Edmond. I think that's a good transition. In looking at our charter, there are some suggested, some suggestions for deliverables. And there has been some work that has already been completed. We have had a Webinar and also a presentation at Seoul. Our next suggested action was to per the IRD working groups to produce a study of the feasibility and suitability of producing the display specifications to deal with the internationalization of registration data.

And I guess, do we want to move to a discussion of the framework for this study, the, what we want to consider, what we don't want to consider as a working group. I've heard quite a bit of discussion on this call and other calls of where we might want to look for, you know, for information and then also

those areas that we consider out of scope or that are being dealt with elsewhere.

I'm not a technical expert in this area. I defer to others of you for some ideas of how we might proceed to the next step and that is perhaps putting together an outline for what we think the study might be.

Edmond Chong: This is Edmond. What is, I'm sorry, probably I should know this, but I don't, but what is the next deliverable and what are the next two deliverables actually that we need to work on.

Julie Hedlund: This is Julie. Thank you Edmond. That's a very good question. The charter is very broad. It was really just a sort of a suggested starting point for the working group and the working group can adjust it as it sees fit. The next step is to, you know, to develop, for the working group to develop a study. That really is right now in the charter of the Seoul next final deliverable for this working group, this study of the feasibility and suitability in producing display specifications for internationalized registration data.

What this study should be, what it should look at is something that I think we're discussing here, but that is at least in the charter, the draft charter, what is our only deliverable at this time and that is.

Edmond Chong: So if I can, that is a little bit vague for me. Who does the study? Does this group do the study or this group should form a, you know, a set of criterias for staff to do the study. That perhaps can help clarify a little bit more.

Julie Hedlund: This is Julie. Edmond that's an excellent question. Dave, do you have some thoughts on that as well?

Dave Piscitello: I think that it would be fair to say that there is expertise among the community that would be essential for staff to actually complete this work. And so I would think that frankly I think that a lot of the people that signed up for this

community would be exactly the people that could contribute in exactly the way that they're doing now.

I think if we have a better feel for what it is that we want to accomplish, staff can certainly take the lead and help write that report, as we do for other (unintelligible) activities. So my first inclination is that we may need to do a couple of, you know, a couple of other short pieces as I did earlier, I guess maybe two weeks ago, to describe some of the alternatives and means that other people in other areas of communication have adopted to deal with the multiple language, multiple character issue.

So perhaps one of the things that I can do in my copious spare time in January is try to put down some notes about the some of the requirements that have been mentioned here so we have a document to essentially add to and subtract from. I think one of the things we're struggling with Edmond is that we have such a broad remit and so many different questions to ask that we don't really have a focus point yet. And often a couple of little point documents may help that.

Edmond Chong: Right. That seems to be the case and that's sort of what I'm trying to drive at and here's a, and you were talking about it, here's a, some sort of a thought. And I'm just really thinking out loud and throwing out to the people. Perhaps what may be useful is if we can out of this whole topic of internationalized registration data, if we can identify maybe three or four areas of specific concern. And then here we would break into sub-groups to come up with sort of recommendations where and then staff could integrate the information for, you know, we would work on it completely as a group.

That might help the conversation cause it seems like every time we start on one task we going around in a lot of directions and try to cover a lot of things on this broad subject.

Julie Hedlund: Thank you, Edmond. This is Julie. I think that is very helpful suggestions. Do we, how do we want to approach coming up with some identifying.

Edmond Chong: In my mind, there's at least a few things, you know, one of which is obviously the domain name itself, the IDN. Is there additional fields that should be displayed or IDN that's definitely one area that should be considered? This area how it should be considered is whether blocks of data for each contact or blocks of data for each domain, the whole set of information in a separate language. That might be another area.

I'm just really thinking out loud here, but I think it may be really useful for people on this call to suggest some of these sort of higher level topics. And perhaps a way to go forward is maybe myself and (Jeremy) as chairs between now and next meeting to try to draft a set of these sort of broad areas where for the next meeting we could, you know, finalize on and then from thereon focus on discussion of these, you know, three, four or five areas which will eventually become the initial study. Does that make sense?

Dave Piscitello: I think so. And if you and (Jeremy) need some help or want to, you want to offload come of the writing chore, I'm happy to help out in that area, Edmond.

Edmond Chong: We're definitely going to need more help. If, you know, as we go about, but you know at this point I think the, it seems we've had a couple meetings, it seems what might be useful is just to have a list of topics and so we can, you know, fit what we're talking about into those topics and then we can have a more fruitful and constructive discussion. That's where my head is at and definitely we everyone's help and especially some staff, so.

Julie Hedlund: Edmond, this is Julie. Would it be helpful if Dave and not Dave, I don't want to volunteer you, but if Dave, you know, had the time to just start with a list of topics, send them to the list and you know just as a start, get the members of the working group, to, you know, then bounce back, you know, some ideas based on that initial list?

Edmond Chong: That would be perfect. I think that would be perfect and you know between now and next meeting, we should have a draft set of, you know, a small list, right now I'm talking about a small list, three at the most five areas that needs to be discussed. And then from there of course we will branch out into other things. That I think would be most useful for us.

It would be great if Dave can, you know, just throw that out to the list and we can talk about it between now and next meeting and try to make some decisions next meeting. And then either cut up the work or go about it together as a group, you know, depending on how the whole group sees it at this point based on the list.

Dave Piscitello: I think I can do that. I have about a page and a half of notes from comments that, you know, that Steve and you and a couple of people didn't actually offer their names, so I couldn't actually attribute the comments. Steve Crocker, so there were multiple Steves as well.

So I can put together a couple of, or a description of what we've covered so far and what those and try to lump them into bins or categories and I'll try to that, well, we have to decide when our next meeting is, but. I'll try to do that sometime certainly before the end of the year.

Edmond Chong: Cool. In terms of that, are we settling on a bi-weekly meeting, Julie? Was that the decision last week?

Julie Hedlund: Edmond, this is Julie. I don't think we formally made that decision, but generally most of the working group, unless, you know, there's a, you know, specific, you know, hard deadline that we're working towards been meeting on a bi-weekly basis. That seems so far to be a schedule that works well with this working group and of course now that we have the holidays coming up here I think at least for the next meeting we probably want to wait two weeks

which would be the 4th of January for the next meeting. DO others have thoughts?

So Edmond if that seems right to you, we could stay with the bi-weekly schedule.

Edmond Chong: Right. I think it's probably useful to make it regular and so far bi-weekly seems to be working well. And from between now and next meeting it's going to be the holidays, and but - the only thing that we're looking for to happen between this meeting and next is really a draft list of the items or bins or categories, however you call it. And then in the next meeting we can decide on them and actually break out the work. To me that seems like a discussion point.

Julie Hedlund: Dave, you had a comment?

Dave Piscitello: I think that that's manageable. I will try to get some time today or tomorrow to get a list to the list and if Julie can share what mode she's taken and I can use them along with mine and I think we'll have something out for people to at least start.

Julie Hedlund: Right. This is Julie. Dave, thank you. I will actually summarize the discussion, today's discussion. I may not catch all of the, you know, the technical issues quite accurately, but I think you'll be able to see what I thought. And I'll share those with you Dave and then perhaps that can be a basis between the two of us for an initial list. And then I'll send that out to the working group and ask for comments on the list and then I'll summarize those comments, Edmond and you know, see if I can prepare a summary prior to a meeting on the 4th of January.

Edmond Chong: I think that sounds good. I guess the question is anyone else want to add their thoughts on both this approach and on whether some of these topics

that you feel is pretty important that we air over the last couple of meetings we have never touched on so we don't miss anything unless that's overall.

Julie Hedlund: Edmond this is Julie. I guess we can take silence for agreement in this case.

Edmond Chong: Okay.

Julie Hedlund: So for everyone on this call then we have an action item between Dave and I that we're work on to collect the notes form this call and others to produce a list that we'll send out to the email list. We'll schedule a call for the same time the 4th of January and unless anybody has any other business they want to add today, is there any other business? Hearing non I want to thank everyone for being on this call and for Edmond for chairing and I wish everyone a very, very happy holiday and a happy New Year. And we'll see you or hear you all in the New Year.

END