
STEVE SHENG: Thank you. Good morning, good afternoon and good evening. Welcome to today's call on the WHOIS Review Team Internationalized Registration Data Working Group on May 21st at 9:00 am Eastern Time. On the line I have Jim, [Naoki 00:26], I have Takao, I have Sarmad and I have Nishit. We received apologies from Dennis. We are waiting for others to join. It looks like we have quorum so Jim, over to you.

JIM GALVIN: Okay, thanks Steve. We have a comment from the ALAC, and I believe that's actually our only comment here, isn't it, Steve? There hasn't been any others, right?

STEVE SHENG: That's correct.

JIM GALVIN: Yes, so I'm hopeful that folks have had a chance to take a look at that. The one thing that jumps out at me in that comment was their emphasis that we need to consider Hong Kong characters, and I guess that's not a meaningful comment from me. I don't know how to interpret that comment. Can anyone help me out here, to understand what they're trying to tell us?

STEVE SHENG: It seems... That seems to... I'm puzzled as well. It seems they may have confused our work with the IDN Variant Generation Panel, Integration

Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.

Panel work. The way I'm envisioning it... Are we specifying what characters are allowed? I haven't seen if our deliberation touches on that, so I'm a little surprised by that comment as well.

JIM GALVIN: Okay, good. Then I don't feel too bad. I was a little concerned that I was just not understanding something. Would it be reasonable, Steve, to ask you to reach out to someone on the ALAC side, and ask them for more specific references to where that comment applies?

STEVE SHENG: Sure, I'm happy to do that. Let me reach out to them.

JIM GALVIN: Yes. Other than that, I did not interpret or understand anything in the ALAC comment that we're not already addressing. I didn't feel that they were giving us any direction or comments that we weren't already aware of. I'm interested in just making sure that everyone agrees with that. Does anyone have anything else that they want to call out from the ALAC comment? I do want to be careful not to overlook anything. The question can remain open, if folks want to think about that.

In the notes, Steve, we can see, for those who are not on the call, we can leave the door open for folks who may want to add a comment, based on what ALAC had to say. Okay. Last week we had what I thought was quite a productive discussion. We should continue some of the open questions that we had from back then. Does anyone have a preference for when we go first and what we dig into here?

Okay, so one of the things that we had was the discussion about defining the purpose of the address field. I apologize. As I recall, we had been asked last week to try and get some stuff out on the mailing list, but we're all a little late this week in getting stuff together here, and getting the notes out. I didn't get a chance to follow up on the notes and get my comments into words on the list. I guess I'd like to start with that discussion a little bit and see if we can come to closure on what is meant by that.

I guess I can start by saying what my point of view is. I'll start again with what I was trying to say last week about this, unless someone else would like to jump in. I don't want to feel like I have to be the only one talking. Anyone?

SPEAKER: Apologies should come from me. Sorry that I didn't follow up on that one either. That's all. Please go ahead.

JIM GALVIN: Okay. There's a part of me that feels like we don't actually have to... Let me phrase this differently. I prefer to think about our scope with respect to requirements and the data model as functional work, rather than practical application work. That might be a poor choice of words, so let me try and explain a little bit what I mean by that. Last week we were talking about the purpose of the address. Is it the registrant's address, or technical contact's address? What is it used for?

I believe that we would be best served by thinking about the address field as simply representing... Functionally, the purpose is to be able to deliver a postal package, an envelope or package of some sort. Whether that address represents a registrant or a billing contact, technical contact, abuse contact, or something else, I don't think that that's a question that we need to answer. Our goal is to ensure that the address field has the appropriate set of requirements so that it can be used for the purpose at hand.

Also, the completed [half data 08:05] entered by an appropriate party. We normally think of a registrant as providing all the input, but I want to say "appropriate party" owing to the fact that maybe other things, at some point here in the future, might enter some data. Very quickly, in one sentence, I really think that the only thing that we need to know about an address field – and really for most of the data elements – is its functional purpose.

We should be specifying requirements to meet that need. Additional detail like whether or not it's used for a registrant or something, I don't think is relevant to our discussions. I'd be interested if anyone has an example of why that would not be true. That's my comment on that question. Anyone agree or disagree?

TAKAO SUZUKI:

May I comment? If the address is already valid, and reachable; mails get delivered with no issue whatsoever, that's one thing. If that doesn't work though, or if someone just enters false data, do we care or not? In order for us to say we don't really care, we just need data to be entered,

no issue... If we say, "No, it has to be valid," then all the verification mechanism, what's the [non ASCII 09:52] sent? Who's going to be responsible for translation and so forth? Maybe [unclear 09:59] relevant issue. That's where I'm coming from.

JIM GALVIN:

My response to the validation verification issue is that I don't think that our scope is to deal with that issue directly. The piece of validation verification that I think is in our scope is that we need to be sure that we don't suggest a requirement for a data model that could not be validated or verified. I would also suggest that if it makes sense, if we were to consider a requirement that we think would facilitate validation or validation, then I think it probably would be appropriate to include that.

That would certainly be the something to talk about. I don't think though that we need to take on the verification or validation process head on. I don't think that's in scope, at least not for me. Takao, any response to that?

TAKAO SUZUKI:

The scenario I have in mind is if everything is submitted in Arabic to us, upon registration of a domain. I just don't know what to do. That's where I'm having problems.

JIM GALVIN:

I agree with you. That's a problem. I'm thinking about that problem too. I just don't think that it's within scope for this group to solve that

problem. There was a time, some number of meetings ago, when we were having a bit of a discussion about this. When we were doing our interim report and talking about requirements we were talking about marketing, and the market would figure out what the right thing to do it.

For example, I'm imagining a scenario where... It's possible that as a registrar, since registrars are going to have primary responsibility for validation and verification of data, it's possible that you'll have a website up and, you're right. Someone may come to your website and want to buy a domain name, but validation verification is something that's probably very localized.

This is just my expectation here, speaking personally. It may very well be possible that someone's going to come to you. They're going to live in Saudi Arabia and want to enter an address that's local to there, but you simply have to turn away that business and say, "Sorry, can't help you," because you can't perform the rest of the registrar functions. A registrar is going to have to say, "Gee, I have no mechanism for validating this address. Yes, I'm going to have to pass on this customer."

I don't like that answer. I'm sure that you don't like it. I'm sure a lot of people are not going to like it. I don't know what to do about it. I really don't, but I also don't think that problem is in scope for this group, but you may feel differently. Nishit, go ahead.

NISHIT JAIN:

I think I totally agree with you, on that part of discussing about that that part, about the address, is totally out of scope of our Working Group.

The validation and verification part of the address is an important point, but I think it's totally out of scope of our Working Group.

JIM GALVIN:

Thank you Nishit. One thing that does occur to me though – and Steve, this is probably worth making a note of – this validation verification issue, although we may not answer the question, I do think it's appropriate for us to find a place in our final report to call this question out and make it visible. I'm sure that somebody will want to take up having to answer this question.

Maybe when we get closer to doing our final report, we might actually want to make a recommendation that somebody should answer the question, but we can leave it as an open question for right now that should be documented, because I think it's an important one.

STEVE SHENG:

Thank you Jim. Is it possible to state the question, so that I can note it?

JIM GALVIN:

Well, let's see. Let's try and create something on the fly here. "The Working Group observes that it may be the case that a registrar will be unable to validate or verify an address, if it does not understand the local form of the address that's entered by the registrant." Let's try that for now. I think that's enough to at least key us on the issue, and we can certainly fix that question later, when we add more discussion to it.

I think the exact form of the question is going to depend on the choice we make for the postal address data element in general.

STEVE SHENG: Sure.

TAKAO SUZUKI: That's fine with me too.

JIM GALVIN: Thank you for that. Have we answered the question of the purpose of the address? Is there anything else to talk about? Okay, so I believe then that the answer that we're agreeing to is that for the scope of this Working Group, the purpose of the address is a functional purpose; simply that it needs to be possible to use it to deliver a postal letter or package. Okay, moving onto what else we had here.

Sarmad, you had brought up the question of sequencing last time, so let's talk about that again a little more, and see if we can come to closure. As I understood the sequencing question that you were raising, you were simply observing that as part of your work in translation and transliteration, and the study that you're doing, you point out that obviously context is one of the most important things to help translation and transliteration.

In particular, with addresses, different regions tend to format their addresses differently, and in order to enhance translation and transliteration, the question before us is should we suggest or in some

way document or recommend – whatever is appropriate here – that the order of address elements should respect some standard or something that we’ve yet to identify. I believe that’s the question that you were putting forward, Sarmad. Have I characterized that appropriately, or do you want to add to that?

SARMAD HUSSAIN:

Thank you. I’d like to add just one more thing, and that is that we have a very similar situation with the date period, in that I know of at least three conventions, where in China, they write the year, and then the month, and then the date. In at least British systems they write exactly the opposite; date, month and year, and obviously the American system is the month first, then the date and then the year. I [unclear 19:57] system, which has date fields, that these are handled consistently somehow.

I think it’s probably a good idea [unclear 20:09] that means we can [unclear] at least show that there are [concern 20:20] different [countries], but given the [IRD 20:22] within the data, these things are, in a way, at least [unclear 20:29] that these are [unclear]. Then yes, what you said as well, my idea was that... There are [three ideas 20:38]. One is that we stipulate the process. If we had, for example, a standard way of doing addresses, we stipulate that it has to be done in this way, but that means we are strict, maybe, in some cases.

Another option is that we allow it to be done in multiple ways, but somehow [tag 20:59] it, saying that, “This is [unclear] done this way,” meaning that one way or the other is [tagged], allowing the user to do it

one way or another. Again, I think what I'm saying is that if this data has to be processed automatically, if you know what the [subset 21:28] is, it will be done more effectively. Thank you.

JIM GALVIN:

Thank you for that. The analogy with the date fields is kind of interesting. I hadn't thought about that. That's a good analogy. I have thought about this sequencing a bit myself, and I guess I really don't know for certain where to go with it. There are two things that occur to me. One is, from a purely technical side, there's always great value in having data appropriate marked for what it is.

Having something called a name field "address line one", "address line two", "city", "state or province", "country", all those kinds of labels, and have it broken out into parts, like a date field, as you suggest. From a technical point of view, and the ability to process the data and do things with it, there's great value in all of that, and I think we all understand that. The other side of it though is, one thing that's occurred to me, is isn't this just a display issue?

Maybe that's the way to approach this. Rather than us trying to stipulate exactly how addresses are supposed to be displayed, we make the observation and recommendation that addresses can be displayed according to local convention, and maybe we simply state that when being displayed in an internationalized way, the data is always displayed with its appropriate label.

So you would see "name" and then whatever. You would see "address line one" in whatever, "address line two" in whatever, so if you're going

to present it for presentation issue, in an internationalized form, you always do it with the label, so that it's clear what's going on. If you're going to present it in a localized form, you're free to drop the labels, because you might mix and match the elements in some appropriate way, for the region.

We wouldn't have to stipulate what that was. The local environment could do that, because that's part of what translating or transliterating from international to local form takes. Now, here's the one down side to that, and then I'm interested in reactions. Right now, in registry agreements in particular, when you do display registration information with the WHOIS protocol, it's actually very specific about what's to be displayed and how.

For gTLDs, it's a contractual issue, exactly what that output is supposed to look like. People tend to stick to their contractual requirements. Unfortunately over time here, not all contracts are exactly the same. Now, with the New gTLD Program it's probably all going to get fixed, because everybody's taking the same contract, but you still have that couple of dozen historical legacy TLDs that have to get there.

That becomes an interesting issue for ICANN to consider, if it's going to allow this localized presentation, and allow people to alter it to meet this sequencing issue that Sarmad is raising. That has downstream consequences that I haven't fully thought through, but I observe there are probably some issues there. Anyway, that's my thought on the sequencing, Sarmad. I'm still struggling with how we could actually say something specific about what to do there, and so that's my suggestion on what to say.

Going back real quickly, we could distinguish between internationalized versus localized presentation of the data, and in the internationalized form we'd always put labels on it, and then open the door for localized form to do whatever is appropriate for the region. I'm not sure we could say more than that, so that's my suggestion. Your comments first, and then from anyone else?

SARMAD HUSSAIN:

I [unclear 26:20] agenda. I'm actually not sure what the solution is. As I said, I know that this does present a challenge for translation or transliteration of addresses. I'm really not sure what the answer is. I think a place to start would perhaps be to see whether there are any international recommendations for addresses. I understand that there are contractual requirements, but I think [once always 27:13] the recommendation [unclear] requirement, is that is the case.

So I'm not really... I'm just focusing on that at this time and trying to understand what the problem is, and what a good solution for that problem would be. Then if it goes against the contractual agreement ICANN has, one can always say that this is recommended, or this is [unclear 27:44] even if [unclear 27:44] recommendation you think is too strong. At least have something [unclear 27:52] and see what other possibilities there are.

Maybe I'm not sure how universal [unclear 28:02] this, or the UPU have any recommendations on how to [unclear] addresses across different countries, or does it have to be [unclear] delegated [unclear] others. Then obviously you can't really do much with it. So I think what [unclear

28:28] that we should at least try to see if there is any agreed global mechanism for addresses. Just look at it and see if there is anything there. If not, then obviously we can't really say much more than this.

JIM GALVIN:

Thank you for that. Let me ask Steve really quickly. Steve, do you know, or do you have some mechanism for how we could find out if there are any recommendations or standards for address formats in different regions or countries? Do we have to get the UPU stuff, or is there something else we might be able to do?

STEVE SHENG:

Thank you Jim. I can respond directly to your question, and then I also have a comment on this. In the Adobe Connect I've posted the link to the address template for each country in the world. You can click on that, and then there is a drop-down called "addressing area" where you select the country. For example, you can select the USA, and then you would display the template, the address format, for what symbols are used and how the address is written.

That's the piece of information that I know of. For example, in the US format you would first put the address identification, street number, name, PO Box, [unclear 30:27], highway, postal, and then locality, and then state, postal code and country. That's how, in the UPU system, how they specify address formats for each country. I'm not sure if that's implementable, but that at least is how the standards are being specified.

With respect to your question, Jim, I think the way I see this issue is that it's more of a representation problem than a presentation problem. I think Sarmad's problem is someone would need to take this address to translate or transliterate it, and because there's insufficient information, right now all the addresses are just a bit text field. There's insufficient tagging. There's insufficient meta data that would make the translation or transliteration be not so accurate. In so much, it's a presentation problem, but it's more so a representation.

How do we represent addresses in a way to increase the accuracy of the translation or transliteration. That's how I perceived the problem. That's some quick feedback.

JIM GALVIN: Thank you Steve. Takao?

TAKAO SUZUKI: Thank you. I had a question. When we say "other data", are we talking about the [unclear 32:30] one-field data, or line-by-line, like "country", "state", "address line one", "address line two"? Please can you help me, because I'm not sure if I'm understanding the data format here. In other words, when somebody... Are we talking about storing data as one data, like "address one" obviously as an address, and then you have like four lines, [unclear 33:17], or are we talking about we say "address: country", "address: state", and so forth?

STEVE SHENG: I'm actually not sure I understand your question, Takao.

TAKAO SUZUKI: I'm sorry. I'm thinking the address is not to do with the presentations layer, but from the [unclear 33:50] when the data is stored, one with the same data [unclear] one address data field, and that's multiple lines, and then it's a freeform. That's one way to do it, and then one other way to just saying [data field 34:10] "address: country", "address: state", and then [field end 34:14] separately. Which one are we talking about?

STEVE SHENG: Yes, I got your question. Currently, for example, in the WHOIS, the fields required are address one, address two, city, country and postal code, so that's the breakdown of the address. What Sarmad is saying is – Sarmad, please correct me if I'm wrong – is just having address one and address two might not be sufficient to enable accurate translation or transliteration.

One solution is one could consider to further break down address one into, for example, street, versus [premise 35:10] and building number, and having that text, so that when it comes to translation and transliteration data they can identify, "Okay, that's the street name. What I'll do is I'll transliterate." "Oh, that's a street name. After the transliteration I'll add a translation of the word "street" or whatever."

Then that [richer 35:36] tagging inside the "address one" or "address two" field, that's my perception of the problem. Feel free to correct me if I'm wrong here.

JIM GALVIN: Sarmad, do you want to add to that, or are you okay with that?

SARMAD HUSSAIN: I'm okay with that.

JIM GALVIN: Okay. Takao, did that answer your question?

TAKAO SUZUKI: Yes, it did, thank you.

JIM GALVIN: Okay. Anyone else have any comments on this sequencing issue?

TAKAO SUZUKI: What if the people put the wrong data in the wrong fields? Do we care about those or not?

JIM GALVIN: Well, yes, so this becomes another one of those questions that I think is difficult to answer, and I don't think there's anything that we can say in our scope about it. Arguably, a postal address, for the most part, is free form text, and that's the way we're describing the data model currently. Although what we really mean is the free form text inside certain identified labels, so we have a spot for you to indicate the first line of your address, a spot to indicate the second line, a spot to indicate your city. That kind of thing.

Is it possible for you to fill out those fields wrong? Sure. Within the scope of what we can say, I think it's absolutely possible for you to fill out the fields wrong. Who's to say you don't take address line one and just type all your information right there, one long line, with a comma, which is the usual line separator when you're doing it on a single line? It is free form text.

Strictly speaking, from a syntactic point of view, there's no reason you couldn't do that. However, I think this falls into the category of when the next step is validation or verification of the address, the address would presumably fail in many cases if you don't properly fill out the form. That's where that will get caught. All we can do is specify requirements that a user could syntactically do something that in fact is not valid.

That won't get caught by the data model, or by any IRD requirements that we'll set. That gets caught later on down the road here in the workflow. That's what I think at the moment. Okay. Steve, if it's okay with Sarmad I want to just restate kind of what I said before, and suggest that that's our position here, but it's still an open question because we have people not on the call. It'll give people more time to think about this.

I think the answer to the sequencing question is that we should continue to distinguish between internationalization and localization. For internationalized address, there will always be the labels for the elements that are filled out. You'll always see "address line one" and the data, "address line two" and the data. We will recommend – and we

have this nice reference here to the UPU – that in a local form, an address may be...

Those elements will be pulled together, and they might be presented to a user in a local form, according to these UPU requirements. There is no standard presentation format if you're going to present it in a local form. The guidance would be to do that, not as a requirement of course, but as an option.

STEVE SHENG: Thanks Jim. I'm trying to note this exactly. Internationalized address – do you mean an address in one local language or script, or is it something else?

JIM GALVIN: No. We have not yet answered the question of what exactly is the syntax of the data. Simply commenting on the presentation, which would be that the presentation, in an internalized form, would always be labeled. The elements would always be labeled. In fact, I'm not even speaking to the question of the labels could be transliterated in some appropriate way too, so I'm not even suggesting the labels have to appear necessarily in English or any other single language.

STEVE SHENG: Okay, so what I interpret "internationalized" means is the address in one local language or script. For example, if a Chinese registrant submits both a Chinese and an English address, which one is the

internationalized address? The English one, or the Chinese one? Which one is localized?

JIM GALVIN: We haven't answered that question yet.

STEVE SHENG: That's fine. I'm just noting the question.

JIM GALVIN: Okay. Just a quick double-check back then with Sarmad. Are you okay with what I'm proposing here for the moment, Sarmad?

SARMAD HUSSAIN: Yes.

JIM GALVIN: Okay, thank you. Okay, I'm going to get back to this email message here with the note. The next question is... Okay, now we're left with this question of resolving the issue that we left in the interim report, of three possible sets of requirements for the collective set of address information, postal information in general. Let's see. Well, I'm thinking that it's ten minutes to the hour. I'm wondering if we should get into this discussion.

I'm thinking not, because it's probably going to take five or eight minutes or so to set this up, and then we're going to be at the top of the

hour, and we're not really going to have an opportunity for discussion. I think what I'd like to do, if no one objects, is let's plan our meeting for next time, to focus specifically on this issue. Let's seek to resolve the question of the three address options that we left in our interim report.

Steve, if we can get some notes out here, after this meeting, relatively quickly, that will help prompt me. I'll try to write out some text to set up this discussion so that we can pick this up and devote next week's meeting to this topic, and see if we can come to some consensus.

STEVE SHENG:

Sure. Happy to do that. Jim, one suggestion is because of the complexity of issue, do you want to have a 90-minute call instead of a 60-minute call. That might give the Working Group more time, but on the other hand, it may not be available. That's just a suggestion to think about.

JIM GALVIN:

Anyone have any objections to a 90-minute call?

JODY KOLKER:

I don't.

TAKAO SUZUKI:

Me neither.

JIM GALVIN:

Okay. That sounds good to me, Steve. I don't have any objections either. Let's plan on 90 minutes next week. It would be especially important to get that notice out today, if possible, on the mailing list, to give those who aren't on this call as early of a heads-up as possible. All right. I don't hear anyone crying out in pain here about adjourning at this point, so let's say we're going to do that.

Any other business from anyone first? Okay, then let me suggest that we're adjourned and turn it over to you, Steve, to close this down.

STEVE SHENG:

Thank you. Today's discussion, just a quick summary, we focused on two of the three issues raised. On the issue of the purpose of the registrant address, the proposal was... The purpose of the address serves a function to be able to deliver a postal mail. Whether the address represents the registrant's billing or technical contact shouldn't matter.

On the issue of sequencing, what I gather from the proposal is for the internationalized address to always have the label to associate with it, and for localized addresses, one might consider to use the UPU standards. I have an Action Item to send out a note as well as to reach out to ALAC regarding their comment on the Hong Kong supplemental characters. Those are the quick things I captured. Did I get it wrong? Did I characterize it incorrectly?

JIM GALVIN:

That sounds great to me Steve.

STEVE SHENG:

Okay, so with that, we will schedule a 90-minute call next week from 9:00 to 10:30 Eastern Time, to focus specifically on the issue of the three models that were discussed in the Interim Report. The goal of that meeting is to try and [debrief 48:00] the Working Group, to come to consensus and select one model. With that, thank you everyone. The meeting is adjourned. Can you stop the recording?

[END OF TRANSCRIPTION]