Transcript
IRD Outreach Webinar
Wednesday 16 February 2011 at 1400 UTC

Steve Sheng: Thank you. Good morning, good afternoon and good day to everyone. My name is Steve Sheng. I'm at policy staff supporting the IRD working group. And today I'll be giving an update to the community on the interim report. I realize right now the audience is only working group members and staff, but let's go ahead anyway, because more people might be joining later.

So there are two goals for this presentation. First, the working group wants to brief the community of its work so far. And second and more importantly, it's considering different policy options. And in deciding these models and options for internationalizing data, they would really like to seek feedback from the community on these models and recommendations.

So just some background information why and what the IRD working group is talking about. Internationalization, you know, the domain part of the Internet has been increasingly being internationalized. We have guidelines for how to internationalize the domain name part, the label. For example, we have the (IDNA) guideline that provides guidelines how those labels are, you know, to be constructed and, you know, converted in applications.

But so far we do not have standards for the submission and display of domain registration data in directory services. So the goal of the IRD working group is to study the feasibility and suitability of introducing such specifications to deal with the internationalization of registration data. So there are two parts. There's the suitability and there's the feasibility part. I'm going to explain more later on.

What is domain registration data? These are data that registrants provide at the time of registering the domain. So if you register a domain, you would submit this information first to the, you know, either the reseller or to the registrar. And the registrar in turn submits either all or part of this information to the registry, and both registrar and registry maintain Whois services usually, and they display this information using the protocol.

This information includes the domain name, the name server information, the sponsoring registrar information; and more importantly, the contact information for the owner of the name, the administrative contact and the technical contact information. So it's the internationalization of this piece of information we're talking about.

The working group is a joint asset in GNS working group - GNSO working group, sorry. And the first thing it decided is that various elements of the registration data could be separate, be internationalized. So for example, domain names, name server names, you know, all these items listed here -- there are existing standards for internationalization, for example, the telephone and fax number, the email number.

The working group just used those existing standards and applied. Others -- for example, the sponsoring registrar -- the working deliberated and concluded that that piece of information is best to remain US-ASCII to have a uniform, you know, display across the different, you know, internalizations efforts.

So the key part of the registration data, as I mentioned before, is really the contact data. And to that, the working group has discussed four possible models. They do not endorse any particular model, and that's really the part that we wanted to seek feedback from the community.

Now before I go into the details of these models, I want to highlight a couple principles that guided the creation of these models. So the first principle is we

want to avoid the Tower of Babel effect for registration data. So on one hand we want to accommodate, you know, scenarios where a user registrant, you know, provides this information in local languages to the extent they are possible.

But on the other hand, we want to avoid in a situation that, you know, a possible situation where hundreds of languages and scripts in the Whois, the domain registration data, and that registrants from one language, you know, will not be able to use that information because it's in a, you know, completely different language -- for example, an Arabic speaker trying to contact a Chinese domain owner, but unable to do so because all the domain registration data information is in Chinese.

Or, you know, the other way around. So that's the first principle. The second principle that guided us is we wanted to balance between cost and the usability of the registration data. So just following along the previous thought, you know, we have a range of submission scenarios.

So for example, you know, translation would probably be more, you know, accurate and usable. But it's also at a higher cost. So, you know, besides, we considered different sets of scenarios.

And the last one, the last principle, is the need to consider both human users as well as legitimate automation. And that would mean the applications that automatically parse and analyze the registration data. Early days for the Whois services, it was mostly, you know, human users. But these days are currently, you know, there are lots of automation going on in the servers as well.

So what are the four models? The first model is registrants provide the domain contact information data in a (must-be-present) script. It's more like today's status quo where a registrant, you know, in this scenario, a registrant

in internationalized domain labels must present this, you know, data in US-ASCII.

Today it's usually done, you know, registries or registrants would first ask optionally, you know. You know, they'd provide those information in their local languages and scripts. And second, in the second registration page, they have to, you know, put it all in US-ASCII. Here are showing two examples. One is the Russian example. The other is the Chinese example. So that's the baseline.

The second model is registrants can provide data in any, you know, registrar-accepted script. And it's up to the registrar's responsibility to provide a point of contact for, you know, translation or transliteration requests. For example, we provided two examples here.

In the Russian example, a registrar -- me or you -- you know, with, you know, the full number of the contact information, is listed along with the Whois output of the registrant; information which is now totally in Cyrillic and, you know, Russian. It's the same example for Chinese.

The third model is registrants provide the information -- again, any registrar-accepted script -- but registrars (unintelligible) literation tools to publish in the (must-be-present) script. So it's up to the - Model 2 is registrars provide a point of contact.

In Model 3 it goes one step further in saying registrars must transliterate -- provide tools to help the registrants to transliterate the information. Again, the Russian example, the contact information that was previously in Russian has been transliterated into US-ASCII. So that's Model 3.

The last model, it's very similar to Model 3 in terms of the roles and responsibilities of registrants and registrars. But here one critical difference is

registrars provide translation tools to publish in the (must-be-present) language. I guess this would be in English.

I want to highlight the differences between Model 3 and 4. Perhaps that can be best illustrated with the example. So for example, in the Russian example, you know, we have it showing translation and showing transliteration. I want to highlight you to the address fields.

So for example, the country -- in this case it's Russia -- in the translation example, it's translated as Russia. But transliteration is shown as (Rusia). The region is - I can't pronounce it. It's (Vladimirskaya) region. And that's been translated properly into English.

But transliteration is just a phonetic, you know, into the US-ASCII. So in this case of transliteration, a human user who does not understand Russian, probably will have difficulty still interpreting this. But for machine processing, since it's in US-ASCII, it can probably process in terms of automation.

So those are the four models. Just to summarize the roles and responsibilities -- so we have, you know, service to be offered is transliteration or translation. And the responsibilities is either the registrant's responsibility or the registrar's responsibility. We put that in the two-by-two table and summarized the models. Model 2 is not in this table because, you know, registrars merely provide a point of contact to deal with the issues.

The working group has issued a set of preliminary recommendations and would like to seek the community feedback. The first set of recommendations is for a directory service in the (ID) environment. The Whois protocol clients, both Port 43 and WEP, must be able to accept a user query of a domain name in either U-Label or A-Label format.

And second, the protocol clients must be able to display results of queries in both U- and A-Label for the domain names. And third, domain registration data should be include variants of an IDN label in the response as well.

So Recommendation 1 sets out kind of a requirement of the directory service in the internationalized arena, so that a (unintelligible) is requirement put on the line protocol. You know, the second is the sets of things that we've been going through, how to separately internationalize the various different elements for domain names, for name servers, for sponsoring registrar information, the registration status.

There are three questions we seek the community's feedback. The first, which of the four models that we just described for internationalizing registration contact data is most appropriate? And are there any other models that the IRD working group should consider?

Second, which of the preliminary recommendations, if any, are feasible? Are there any related recommendations the working group should consider? And the last one is the issues of language tags. Is there a need for the contact information to be in multiple languages and scripts?

So I want to specifically give a use case for that. By meaning, so we have, you know, administrative contact and technical contact, and the domain name holder information contact information. Is there a need for those information to be in different scripts? So for example, you know, we have a main contact in, let's say, Arabic, but technical contact in Chinese.

Or, you know, how much granularity do we go down? You know, do we have an address in Chinese, a street name address in Chinese but, you know, the city and country name in English? You know, to what degree of granularity, you know, those need to be in multiple languages and scripts? I guess that's the question the working group is seeking feedback.

Those are my presentations. So I think it's fifteen minutes. And as I finish it's, you know, it's still staff and working group members. But I'd like to thank you for your participation, and just welcome questions. Okay, I guess we will wait. We'll have another Webinar at 12:00 in the Pacific time zone. Let's see if we get more participants. But thanks for everyone, and we'll see. Okay, I'll see you in a bit. Bye-bye.

Woman: Thank you, Steve.

Man: Thanks.

Woman: Thanks.

Woman: Thanks much, Steve. Good job.

Steve Sheng: Thanks.

Gisella Gruber-White: Thanks, (Louise).

Woman: Hi, Gisella. Thanks, Gisella. I'll get the recording stopped for you now, okay?

Gisella Gruber-White: Thanks for that. Have a good day. Speak to you later.

Woman: Yes. Have a fantastic holiday. Bye-bye.

Gisella Gruber-White: I will. Thanks, bye.

Woman: Bye.

END