WHOIS WEBINAR
TRANSCRIPTION
Tuesday 04 May 2010 at 06:00 UTC

Note: The following is the output of transcribing from an audio recording of the WHOIS Webinar call on Tuesday 04 May 2010 at 06:00 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-whois-webinar-04may10.mp3

http://gnso.icann.org/calendar/#may
(transcripts and recordings are found on the calendar page)

Participants on the Call:
Tijani Ben Jemaa
Paul Fong
Malou Axelsten
Nacho Amadoz
Phil Argy
Walid Nabahin

Staff:
Steve Sheng
Margie Milam
Glen de Saint Gery
Gisella Gruber-White
Matthias Langenegger

Coordinator: Today's conference is now being recorded if you have any objections you

                  may disconnect at this time. Speakers would you like to begin with a roll call?

Margie Milam: Yes please.

Coordinator: Gisella Gruber-White?

Gisella Gruber-White: Good morning to everyone on this morning's WHOIS Webinar. We have

                  Tijani Ben-Jemaa, (Paul Fong). From staff we have Margie Milam, Steve

Sheng, Glen Desaintgery, Matthias Langenegger, and myself Gisella Gruber-White. Over to you Margie.

Margie Milam: So thank you Gisella. And thank you for participating in this Webinar. This is the second of the Webinar series that we've produced at ICANN highlighting the work of Steve Sheng.

Steve Sheng is a Senior Technology Analyst with ICANN and he's put together this report that we'll be discussing today called the Draft Inventory of WHOIS service requirements.

And the way we're going to handle this Webinar this morning or this afternoon or evening for you all is that we will go through the slide presentation.

And if you have any clarifying questions please raise your hand in the Adobe Connect Room and we'll take note of them. And then we'll save the rest of the questions for the rest - for the end of this presentation after Steve has concluded his remarks.

And I'll begin with a little bit of background to help you understand the genesis of this requirements document that Steve Sheng has put together.

And essentially a year ago the GNSO Council requested that policy staff work with the technical staff and the GNSO Council members to collect and organize a comprehensive set of requirements for the WHOIS policy tools.

And these requirements are requested for - from a technical perspective. And they should - the request was that they should reflect not only the known deficiencies in the current system but also possible requirements that might be needed to support the policy initiatives that have been suggested in the past.

The GNSO Council in the past has looked at WHOIS and is currently evaluating other aspects of WHOIS but thought this report was important because it would provide some information on the kinds of technical requirements that would be needed if a WHOIS policy were to be adopted in the future.

And the instructions to staff were to do this in consultation with the various reporting organizations and advisory committees of ICANN including the (SX), the at-large community, the Government Advisory Committee, the ccNSO in the GNSO.

And the request was that we put together a straw man (unintelligible). That is what led to the report that Steve Sheng together with help of others from ICANN staff. And that's where I'm going to (unintelligible) you about the results of that work. Steve I'll hand it over to you now.

Steve Sheng: Thank you Margie for that introduction and hello to all who are joining the call today. I want to move on Slide 3.

So as Margie mentioned earlier the goal is to collect and organize a set of requirements for community consideration.

So these are requirements come from three areas. The first area is current set of features identified as needing improvement, for example in the areas of internationalization and such.

The second area is features to support various past policy proposals. So a GNSO has debated various WHOIS of past policy proposals.

And the goal is to see - is to ask the question if these proposals were to become a consensus policy what kind of technical capability we need to build in to support that? So that's the second aspect of our requirements where their - the compilation comes from.

And the last aspect is features recommended by ICANN's advisory committees and supporting organizations and the community.

So moving on to Slide 4 I want to continue with a couple important notes. First of all the staff interprets the council's intent at the word requirement as technical requirement.

So in other words we're not gathering policy requirements or endorsing certain policies. Let me just use an example to illustrate.

Take the tiered access on proposal that has been deliberated in the past basically providing differentiated access to different parties that use WHOIS.

There's a couple dimensions of this requirement if this would become a consensus policy.

So from a policy perspective there might be requirement for example say a law enforcement should have access to, you know, XYZ data in WHOIS. It doesn't have to be law enforcement. It could be, you know, some other parties.

Then from the operational perspective the requirement will be to answer a list of the questions. For example, you know, WHOIS law enforcement, how to certify law enforcement entities.

And finally from the technical perspective -- and this is primarily the focus of this report and the goal is to find the best technology that can support such access.

So this would include a framework to authenticate users and to give proper access based on their roles. It is the technical areas like this that the report is primarily concerned with.

So continuing on Slide 5 we want to define, clarify our scope. When people talk about WHOIS or deliberate it - deliberate WHOIS in a policy debate they could mean different things.

The technical community generally tend to think WHOIS in terms of the clients and servers that implement the protocol. And I think they are - sorry, the RFC 952, the earlier one. In the later one it's 3914.

The policy communities generally tend to view WHOIS as a, you know, as a data or a single database that serves the registration data. So they're mainly concerned with data accuracy, malicious use of the data.

So since our goal is to look at improving WHOIS, I think it's fitting for us to consider the potential improvements to all of the above areas, the WHOIS clients, the servers, and the data. Primarily we'll be focusing on the technical dimensions.

So on Slide 6 here's the rough outline of my presentation today. First I'm going to talk about the need for mechanism to find authoritative WHOIS servers.

The next four points are concerned with standardizations. You can see the structured queries, standardized set of query capacities, well defined schema for replies, and standardized errors.

Why we need this? The original WHOIS protocol is fairly simple and leaves most of these points, you know, the query and the presentation to the decisions of WHOIS servers and clients.

And there is a lot of variability in these implementations. And such variability creates no uniform experience for users.

And we thought by standardizing some of these features they'll be a more (unified) user experience.

Continuing on Slide 7 we're going to talk about the quality of domain registration data and from a technical perspective how that quality can be improved.

The next point is internationalization, really the elephant in the room. As the WHOIS protocol in today's form cannot - does not and cannot distinguish the encodings that's been used. And it's been - it could be problematic in the IDN arena.

Actually there's a working group called Internationalize Registration Data Working Group actively deliberating on these issues and we will defer most of our decisions to that working group.

Finally I'm going to conclude talking about security, thick versus thin WHOIS and registrar abuse point of contact. So those are my rough outline for today.

I'd like to begin with a mechanism to find authoritative WHOIS servers. The problem is simple. Today it's not easy to find an updated list of domain names and IP addresses of authoritative WHOIS servers.

So there are lists in the past, for example the MIT has lists and even the INA has some lists...

Woman: Excuse me.

Steve Sheng: Hello? Yes?

Woman: Hello.

Steve Sheng: Can you hear me?

Woman: Yes.

Steve Sheng: Oh okay. Can others hear me?

Man: Yes you're coming through fine.

Steve Sheng: Oh okay. Sure. So there - oops go ahead. Let me move to the slide, yes to this slide yes.

So in the absence of that clients use a combination of heuristics, you know, kind of some know the - will know - some of these are - some of these servers are in well known locations. So they use heuristic.

Some others they use the hardware tables to find these servers. And some use DNS records to find the WHOIS servers.

So this is problematic in a couple of ways. It is problematic for the new top level new gTLDs as the top level registries could be expanded significantly in, you know, possibly the subsequent registrars.

And it's also problematic for legitimate automated client of WHOIS as there's - as it's not easily to find where the right authority through WHOIS server to query.

So the requirement is to provide a publicly accessible and machine possible list of domain names or IP locations of WHOIS servers offered by ICANN accredited registrars, gTLD registry operators, ccTLD operators, and regional Internet registries.

So here we want to recognize that ICANN has no contractual power for ccTLDs and the regional Internet registries.

However we hope this improvement will be a good practice so we - that hopefully they can adopt as well.

So moving on on Slide 10 I'm going to talk about structured queries. The issue here is the WHOIS server applications vary with the respect how they expect clients to format their query syntax, their query data.

So the query syntax is generally not standardized. And when you go from different WHOIS servers you may have to use different syntax.

For example, to query a autonomous system member when you go to (errand) you have to use, you know, A6, you know, as kind of as a flag, A as a flag.

However when you go to ripe and you have to use the flag, that's T, AUT dash number autonomous number.

So another example is to control for our IDN responses, you know, different WHOIS servers. For example in DK you use a flag caret set. And, you know, to control the caret set you want the result to be returned.

With .JP you have another flag, a different flag to control.

So as I hope the point is clear is the server applications really vary with respect how they (spec) clients.

And the WHOIS clients, they're different versions and they're not always able to accommodate these expectations for WHOIS servers.

So the requirement is to define the standard query structure the client can implement and that all gTLD registries and ICANN accredited registrars will support.

So user will benefit from a standard query structure. For example a user may wish to submit a list of domain names to WHOIS to check for the creation date of the domain names.

So with the standard query structure he or she can do so without worrying about a specific syntax for each of the WHOIS servers that he's going to query. So that's the requirement for structured queries.

I'm on Slide 12. I'm going to talk about standardized set of query capabilities. Along the lines of standardized query formats past GNSO and (ASAC) reports have called for an expanded query capabilities beyond the domain names.

And some registries already have expanded search capabilities. I believe that I think .mobile already have this capabilities.

So kind of as a - to standardize it, it will be nice to permit users to submit not only domain names as arguments to search functions but they can search on other registration data elements as well. So that's the requirement for this slide.

Moving on to Slide 13, going to talk about will defined schema for replies. I hope this is obvious. Currently there is no standardized format or schemas that registrars and registries return WHOIS queries.

As I mentioned early, the protocol have mostly leave these implementation details to the servers or the clients themselves.

And we have, you know, we have 17 gTLDs. And we, you know, we have, you know, 250 some ccTLDs. And we also have 1000s of ICANN accredited registrars.

And although I mean the variation may not be, you know, in the thousands but there are, you know, quite a number of variations in terms of how the WHOIS output is returned.

And the data is representing differently. And sometimes the terms are used differently. So there's a lot of variability.

So we hope to - the requirement here is to define the standard data structure for WHOIS responses that contain a uniquely identified data elements that must be returned in a manner that ensures no ambiguity across elements, correct syntax and correct semantics.

Basically we're trying to create a uniform experience for the users in terms of the data being returned.

Moving on Slide 16, standardized errors. Currently there's no standardized standard set of error messages identified for WHOIS servers.

And different WHOIS servers may handle errors differently. And the lack of standard error introduces ambiguity and confusion.

So for example let's - so for example if a WHOIS client exceeded the limit, different servers may respond differently to its additional queries.

So for example, some server will not return the result at all. Others will return error message specific to that server provider itself. And still others would simply close the connection.

So there's no way, sometimes it's kind of a puzzle to tell whether, you know, I query the wrong name or just, you know, my quota has been exceeded.

So this leads us to a requirement is to define a set standardized error messages and standard handling of errors for WHOIS. And those are some examples that are shown here.

Very similar other protocols for example, the HTTP protocol, you know, have a standardized errors. For example, you know, the 200, you know, the 300, the 400 and the 500.

You know, most notably the four error means Web page not found. Two hundred means just the query went okay. You know, just along the same lines for standardized errors want to borrow that for WHOIS as well.

Next I'm going to talk about the quality of domain registration data. Obviously the usefulness of WHOIS ultimately lies in the quality of its data.

And by talking about data quality we mean three things. We primarily think about three things. The first thing is the data accuracy. That is, is the data accurate?

The second is the relevancy, the current - the relevancy. Is the data useful or relevant? And finally are the collected data current?

So I want to kind of step from a - from the high level perspective and think about first what are some of the barriers to WHOIS accuracy?

The most obvious barrier is privacy considerations. The originally published report by the National Opinion Research Center -- we'll call it a NORC WHOIS accuracy reports point us at the single - one of the most important factor for WHOIS in accuracy that people are concerned about their privacy. So they may tend to put some bogus information in.

So another factor is obviously, you know, the malicious users, the con artists they use, you know, intentional deception in their WHOIS record.

And the third point is there's currently little or no collaboration of submitted data. You know, there's no tracking not even for, you know, correct syntax or telephone numbers may not be, you know, the - conform to the regular - may not be like ten digit, for example in the US. And there could be user errors in absence of these (unintelligible).

So from the technical perspective to improve WHOIS accuracy, you know, we can address the privacy considerations for these end users maybe by providing (jaded) access.

And from a technical perspective we can also improve for example, you know, corroboration of submitted data to reduce the user error.

So next is the relevancy of the WHOIS data. You know, certain registration data are not as useful today as they were 20 years ago.

And we think, you know, over - we anticipate over in the future maybe there needs to add or modify some of these required WHOIS data elements.

And we think a future WHOIS data model should accommodate sensibility and changeability.

So the requirement to improve the quality of the domain registration data is to adopt - first to adopt a structured data model for WHOIS data that provides a sensibility and changeability properties.

The structured data model can also make some of these checking easier. So for example, you know, it's a more structured and well formatted, you know. Well not formatted per se, but it's well structured so that you can do kind of check in to make sure that user enter in the right information in the WHOIS data.

On slide - moving on to Slide 22, internationalization, the problem here is there's no standards exist today for handling the submission and display of registration data from local language and scripts.

The original - the updated WHOIS protocol cannot - even the updated protocol cannot distinguish the encoding that I used.

And some WHOIS applications or servers in the international - in the IDN arena may not support domain name in your labels. That is the localized labels.

And, you know, some of them they cannot accept or display when characters, characters from sets other than US ASCII-7 are used.

And finally, you know, they - their display is set in local encodings rather than Unicode so terminals must be set to correct encodings beforehand to be able to display the localized responses correctly.

So there's a lot of variations and a lack of capabilities. And currently the - as I mentioned, the Internationalized Registration Data Working Group is actively deliberating these issues and how to fix them.

So for the - and to avoid double the work we want to refer our recommendations to the working group's recommendation and include them here.

Next I'm going to talk about security. Current WHOIS require no identity ascertain, credentialing and oh, authentication. It's more of a kind of anonymous and open system.

However increasingly there's a need to provide security. So for example there's a need to provide mechanism to provide - to protect the privacy of registrants.

There's a need to discourage harvesting and mining. And currently that has been accomplished primarily using rate limit of IP addresses. And, you know, from a technical perspective those countermeasures can be circumvented easily.

And finally there's a need - there have been passed proposals for providing differentiated access, you know, based on the actor's roles.

So there's a real need for security. And talking about the security we want to mention the authentication on access control and auditing.

So how do you authenticate users, how do you give them proper access based on their roles, and finally how to find out if something went wrong and how to correct that. So those are those the three As.

So here the requirement, the potential, the compilation is to define an authentication framework for WHOIS that is able to accommodate anonymous access as well as verification of identities using a range of authentication methods and credential services.

So we want - the authentication framework can still accommodate some form of anonymous access as is current use today. But as more beyond that, you know, can assert identities of different parties. So that's the first point.

The second point is the WHOIS service should support an authorization framework that's capable of implementing granular permissions.

So that granular we primarily mean per registration data object access controls as the tiered access proposal that this is necessary to realize to fully implement the tiered access proposal.

And finally is to define a framework and baseline set of metrics that can accommodate future policy developments for auditing of WHOIS access.

So those are the three requirements addressing the three As questions that have been raised.

For interest of time I'm going to skip the thick versus thin WHOIS. You can read those in the report.

I'm going to conclude with register abuse point of contact. And the requirement is registrars and registries should provide and publish abuse point of contact information as an element of domain registration record.

So the motivation for this requirement is increasingly the domain names has been used as one of the - as one element in the global, you know, electronic crime ecosystem.

And there's a real need sometimes to contact the registrars and registries to take down these registrations.

And there's as I mentioned early, there are - there could be thousands of them. And it would be very hard to find the right person. And thus, you know, the time could be late.

So this has been a (Afac) report. I think (Zac) 40 recommendation and also follow on into the DAG report in the draft acronym guidebook for the new gTLDs.

So we want to put them out here. You know, we think this could improve handling of said requirement abuse.

Briefly I'm going to talk about the next step. When we leave the draft report in March where - this is we're doing the two overview Webinars. And this is the second one tailoring to European and Asian audiences.

We are now consulting with various supporting organizations and advisory committees on a draft report. And we'll incorporate their input.

The deadline for input I want to mention is May 17 of this month. We hope to release a final report in time for the Brussels meeting.

Before opening questions I want to say we really value your feedback, particularly in two areas. The first area is have we adequately identified the origins of each requirement?

And second is did we miss any important requirements or improvements to WHOIS that has been discussed up to date?

So with that I would like to thank you for joining the call and would like to open up questions.

Margie Milam: Thank you very much Steve for that interesting presentation. I think the best way to have questions is if you could raise your hand in Adobe Connect will allow you to (unintelligible) questions report. Do we have any questions?

Ben or - why don't you go ahead?

Tijani Ben Jemaa: Okay that's me?

Margie Milam: Yes.

Tijani Ben Jemaa: Is that me?

Margie Milam: Yes it is you. Go ahead.

Steve Sheng: Hi Ben. How are you?

Tijani Ben Jemaa: Okay thank you. Good morning and thank you for this presentation. I have surely misunderstood but I - if I want to know if the WHOIS data are not accepted or displayed if they are not under ASCII format. Is it right?

Steve Sheng: That depends - that really depends on the client capability. So for example on my client if I - if we only accept US ASCII-7 and, you know, if a WHOIS reply is in IDN format it couldn't be correctly displayed.

And if my client is set to encoding, that is different from encoding the WHOIS server returns, you know, I couldn't display that either

Is that...

Tijani Ben Jemaa: You mean it depends on the server?

Steve Sheng: Depends on the client.

Tijani Ben Jemaa: Client.

Steve Sheng: Yes, depends on the client. I mean there's a - the server is primarily, you know, the capability to take the client input. So maybe sometimes the server couldn't recognize the client's input query for - you know, for example if a client put a new label query in and the server, you know, may not recognize that. So to that end, you know, the WHOIS query would fail.

Tijani Ben Jemaa: You don't think it is a problem because now with IDNs we would have a lot of problems it - this is true.

Steve Sheng: Again that's been discussed in the IOD Working Group. Do I think that's a problem? Yes I think that's a problem. Yes, but that's being actively

addressed in the IOD Working Group. So we are waiting for that, their recommendations.

Tijani Ben Jemaa: Okay good. Thank you.

Steve Sheng: Thank you.

Margie Milam: And it looks like we have a question from (Philip Argie) through the Chat. Yes we have the PowerPoint presentations available online.

And certainly if you have any specific questions that you're unable to ask today or think about after you review the report certainly go ahead and email questions to Steve. Steve, why don't you give them your email address.

Steve Sheng: Yes.

(Philip Argie):   Yes got it. Thank you.

Steve Sheng: Thank you (Philip).

Margie Milam: Do we have any other...

Man: Thank you Steve for the email.

Steve Sheng: Yes. We, you know, we like to hear your thoughts and as we value your feedback. The policy development is, you know, is bottom-up in the consensus process.

But we really want to hear what the community thinks about this report and the ways to improve it.

Margie Milam: (Philip) it looks like you have a question. Why don't you go ahead.

(Philip Argie):   Okay well it sort of came late. I didn't want to jump in and use of other people's time. But as a domain name panelists, one of the issues that strikes me is that we have a need for the ability to quickly look at WHOIS information quite frequently. And we often get treated as if we're trying to infiltrate the WHOIS information as a spammer.

                    I we're just wondering if there was some mechanism by which UDRP panelists could be registered and given some sort of privileged access to look behind proxies and various other people who are trying to conceal WHOIS information for privacy purposes?

Steve Sheng: And that's a good question. Are you - let me see if I understand your question correctly. So you are - you'd be querying WHOIS to find the availability of a domain. Is that right?

(Philip Argie):   Well oftentimes it's to check the correct spelling of the name of the registrant. Often critically it would be to check the date of creation of the domain name.

Steve Sheng: Okay.

(Philip Argie):   These are all quite important issues under UDRP. And oftentimes the parties aren't very good at supplying the correct information. So it would be quite facilitative of a speedier UDRP decision.

Steve Sheng: Okay. Let me think.

Margie Milam: But I can...

Steve Sheng: Go ahead Margie.

Margie Milam: Yes from a policy perspective, one of the things that Steve talked about earlier is that in preparing his report he looked at past policy discussions to

see what kind of technical requirements would be needed to comply with, you know, past policy discussions.

And while right now there is no policy related to tiered access, I think the issue you're talking about really does relate to tiered access.

The idea being that if you're a certain type of - you have a certain need for WHOIS if you're a security company or you're a law enforcement agency or you're a UDRP panelist and you have legitimate reasons to access WHOIS for, you know, valid intended purposes that you would be able to get this priority access.

And the technical protocols could allow this sort of identification so that people that have status could access the full records quickly and not have to deal with some of the burdens of the system right now. For example like you mentioned, there may be rate limits that are imposed by registrars because they're worried about data mining.

And so the issue is really that is, you know, can the protocol accommodate that kind of security authority? And that's one of the reasons that we've included that in the report.

But it really is up to the GNSO Council to develop actual (publicity) requiring that kind of tiered access. And this is more of a technical discussion. And Steve did I answer that accurately from a technical perspective?

Steve Sheng: Yes, yes. Thank you Margie.

(Philip Argie):   Okay. Thank you for that. That's certainly important. But it's interesting because in Australia for example (ABA) has a mechanism that if you can establish you're a bona fide potential complainant under the AUDRP they will supply you with information not publicly available from the WHOIS record

against your perspective respondent. And that's very useful. But that's not a practice that seems to be adopted under other ccTLDs.

Steve Sheng: Right, right, right.

Margie Milam: And if I may follow-up that's the kind of feedback that we would greatly appreciate particularly from the ccTLD communities that may have implemented some of these, you know, these technical aspects.

Because a lot of what we're talking about is trying to come up with something that's feasible. And to the extent that there is experience out there in the ccTLD community that would be something that would be very useful for Steve and could (unintelligible) his final report.

Steve Sheng: Right. So I'll...

(Philip Argie):   One alternative mechanism you might think about because it's more controllable in one sense would be to give the approved providers like WIPO and others some kind of access.

Because WIPO obviously has the means to easily identify its panelists as do other providers. But then you would only need, you know, sort of half a dozen privileged access points that would be easier to monitor and control. But that might be a suitable way of balancing the requirement against the administrative panels.

Steve Sheng: Thank you for that comment. It's very insightful.

Margie Milam: Do we have any other questions? Looks like we don't. I guess we can conclude this call earlier. We do want to again indicate that we're seeking comments until May 17.

And so after you've had a chance to take a look at the report and perhaps think about the kinds of requirements that might be useful for WHOIS please feel free to respond and provide that information to Steve Sheng. That would be most appreciated as we put together our final report.

And Steve do you want to talk a little bit about what may be happening in Brussels?

Steve Sheng: We will probably have a workshop in Brussels on this topic. So look for that if you're planning to attend (unintelligible). The (unintelligible) will be interested in hearing the - kind of the final set of compilations.

Margie Milam: And with that thank you all for participating on this Webinar. And we look forward to further discussion with the community on this important topic.

Man: Thank you.

Steve Sheng: Bye.

Man: Good bye.

Margie Milam: Goodbye everybody.

Man: Bye-bye.

Coordinator: That does conclude today's conference. You may disconnect your lines.

Woman: Thank you (Tracy) and sleep well.

Coordinator: You too.

Woman: Take care. Bye-bye.

Coordinator: Goodbye.

END