LOS ANGELES – Deploying the IETF's WHOIS Replacement
Thursday, October 16, 2014 – 10:30 to 11:45
ICANN – Los Angeles, USA

UNIDENTIFIED SPEAKER: This is the IETF's WHOIS Replacement, in the Constellation room, on Thursday October 16th 2014, and this session will run from 10:30 to 11:45, local time.

UNIDENTIFIED SPEAKER: Hello everyone. We are about to start. And we have a lot of space here in the front, if you would like to join us here, you're more than welcome.

FRANCISCO ARIAS: Hi. This is Francisco Arias, Director of Technical Services within the Global Division at ICANN. To my left I have Murray. I'm unfortunately unable to tell your last name, I apologize for that. Murray is the co-chair of the IETF working group, developing the protocol, which is going to be the replacement of the WHOIS protocol.

We are going to give an update on where are things in the IETF and what would be the next steps here at ICANN to our future migration to this protocol. So let's start.

This is the agenda for today. So a brief introduction, why we are here on this stage talking about migrating from the WHOIS protocol. This is what other people call the port 43 protocol. In 2010, there were a series of discussions with the community on, once again, I should say, because this has been a topic that has come and gone several times.

But this time, when into a good track. In 2011, the SSAC, the security stability and advisory committee, published an advisory SAC 51, that called for ICANN to work with the community to develop a proper replacement for the port 43 WHOIS protocol. That was adopted by the ICANN Board in a resolution later in 2011, in which the Board's staff to work on developing a roadmap to among other things, and the main thing, to replace the WHOIS protocol.

We publish draft of this roadmap, and we have some interactions with the community. And finally, in June 2012, the final roadmap was published, and described things like, potentially the PDP but also at the same time, working in parallel with contract negotiations with registries and registrars, that were willing to adopt contractual provisions in this regard.

And that's what happened. VeriSign was the first for dot com, when their renewal was up in 2012, they agreed to have provision in their contract to implement the protocol once it's ready, and also to work in their development. Same happened with dot name, biz, info, and org.

And we also contract operations in 2012 registry mandates, new TLDs agreement. And also for the 2013 registry accreditation agreements, RAs also have a clause in their contracts to implement this protocol once it's standardized by the IETF. So in that regard, the idea started a working group in 2012, to develop this protocol, and now they are close to finish, and that's why what Murray is going to talk about later.

The RFCs are expected in the next few months. So this replacement of WHOIS is called registration access data protocol. It's as I said, intended to replace the port 43 protocol. It provides flexibility to support various

policies contrary to what we have with port 43. It doesn't mean you have to implement the different options that the protocol allows you to do, it gives you the option to turn on or off as needed, according to the policy in the TLD.

It is already operating in production for, Lisa [Copal] of the regional Internet registry. These are the IP registries as they are called. We have ARIN, and RIPE. ARIN in North America, RIPE exists in Europe, that have both production implementations. And I think the rest of the areas have at least bylaw implementations, so this is something that is real and already working.

So it provides many benefits, which Murray is going to talk in detail later. And of course, it was a sign and now with the core knowledge of this industry, as opposed to the [inaudible] protocol that was assigned a long time ago, with a very different reality back then.

So Murray?

MURRAY KUCHERAWY: Good morning. Murray Kucherawy. I co-chair the WEIRDS working group, which is the working group that's been developing RDAP for a while now. I'll give you a tour and then I'll give you some status on where we're at. But essentially the problem with WHOIS, as it was presented to us, is that it is effectively unformatted.

A registrar is able to return… A registry is able to return data in any form that it wants, with our without fields that it thinks you might want to see or that it should show you. It's completely ad-hoc. It's unauthenticated, which means pretty much basically all queries are

anonymous. The only way I can tell one client from another is by IP address, and with the advent of NAT, event hat's not guaranteed that the same query is coming from, two queries are coming from the same place.

It's ASCII only. There is no support for internationalization in port 43, as we've heard in the earlier session. And it's completely insecure. The protocol is, has no provisions whatsoever for secrecy.

I'm sorry. I should have rolled through these rather than covering them all. One of the problems with unauthenticated clients is that it is impossible to give preferential service from one client over another. So if we have a query coming from a random person on the net who is interested in scraping details like email addresses for spammers, I can't tell that from a law enforcement query, where I should be given everything as quickly as possible.

It's not possible to do reasonable rate limiting. It's not possible to do many defensive things that are simply not possible if you can't identify one client from the next. I covered this piece. Also WHOIS has no defined extension strategy whatsoever. Sometimes the query supports wildcard searches, sometimes they support ways they do prefix searching.

None of this is standardized. And when you're talking to one registry or the next, you have no idea if they're going to support a search query, so you might get back completely unexpected results, if you try something like that. An interloper can not only see what you're asking about, but see what the answer is. So if someone can man in the middle you

because they can't get the information, they can just listen to your connection until you can get it.

Especially for law enforcement queries where extra data would normally be provided, this is obviously not a preferable situation. So previously, WHOIS actually has its origins in the late 1970s, which is not what you would expect given the RFC number that it has, but that's only because the protocol was never actually written down anywhere until RFC 3912. In '94, we tried to introduce something called RWHOIS, which was more of a hierarchical network structure, sort of nearing the way the DNS works.

But the uptake for this was unfortunately weeks, so there are still some clients, maybe a couple of servers out there, but it's not universally deployed. The IETF tried to take another run at this in 2005, with the CRISP working group producing something called IRIS that tried to be all things to all people, and in doing so, it unfortunately became extremely complex. An uptick for it was, we saw, I think, one open source implementation that didn't really get very far and we, it basically has fallen aside.

So in 2011, we were approached by some ICANN staffers, and at the time, I was working for an email security company. So we were particularly interested in solving the WHOIS problem. There is information that security companies would like to get out of WHOIS that they simply can't get the way things are right now.

And so ICANN staffers approached ARIN to talk about what, if they had any ideas for a new solution. And we started up the new effort using the original requirements for IRIS, which we thought was a reasonably

good requirement set. So this was our, the basis for our work, but set the protocol aside and basically start again from there.

We formed the WEIRDS working group, which is a fun name to say but actually does stand for something useful. There was a birds of a feather session in the spring of 2012. The working group actually started right after that, because it was clear there was a lot of interest. It took us a while to get started because, beyond that, because we were worried that the RIRs already had prototypes for this.

But we weren't sure that the prototypes they had built would actually work for the domain name space. After a lot of internal discussions and non-believers and so forth saying, no, we have to go two separate ways, we eventually decided that we could actually do even one unified solution, and that's what became RDAP.

So the fundamentals of RDAP. The transport is HTTP, which is, which provides us huge advantages. There is a lot of deployed open source web code, both clients and servers. There is essentially a very widely developed and deployed infrastructure for the web. I mean, this is not, shouldn't be a surprise to anyone in this room.

There are already capabilities for security and authentication, which solves some of the requirements mentioned earlier. I can now differentiate one client from another. I can identify, and have a table of, for these clients, I want to give unlimited rate queries. For anonymous clients, I only want to do a certain number per hour, or what have you, now that I can tell them apart.

HTTP already has support for encryption, so I could hide, you could hide the question and the answer inside of a connection. And this solves this requirement. HTTP already has supports for redirects. So if a question lands at a server that doesn't know the answer, but it knows the right place, they can just simply issue a HTTP redirect. The protocol itself does not need extra provisions for redirect.

Replies are going to be JSON formatted. JSON already supports UTFA, so that satisfies the need for internalization. Transliteration, I'm not familiar with how that work. That sounds more like a policy effort that's going on right now, so okay. If I stray away from technology into policy, then I begin to speak gibberish, so I'll try not to do that.

Internationalized domain names are supported in both the question and the answer, work that predated the RDAP development, nicely sets a framework for us to do this sort of work without having to develop anything new, which those two things together basically satisfy all of our internationalization requirements.

The most interesting piece of this lately has been how do we figure out where to send the query. If you have a domain name, you need to make a RDAP query, where does the registry live for dot com or dot co dot UK or anything like that? And the solution is that IANA will begin maintaining a bootstrap registry for each of the three types that we have, network blocks, autonomous system numbers and domains.

The registry will be published in JSON and will essentially, it's our vision that it will essentially become a part of the registration process. So when a new TLD is created, an entry will go in this registry that will say, that new TLD, for that new TLD, you can issue queries to this location.

And the bootstrap protocol explains how to assemble a query based on what's in the registry.

Essentially this is like the DNS root cache, that's a file you download once in a while. It doesn't change very often, only when a new TLD is introduced, or when a new TLD server to be introduced. And for some period of time, you just continue using that bootstrap file until there is a need to do another one.

And again, this, HTTP already has support for redirects. So you could, in theory, have one server, to which you point all RDAP queries, and as long as it knows the right bootstrap details, it can redirect things as needed. This, there is also a registry for not only where TLDs are AS ranges, or networks go, but there is a separate IANA maintained registry for the fields that can be returned, and what will go in them, and what syntax they have.

So this satisfies the need for extensible field sets, so some new data element you want to be able to report via WHOIS, you would send it to the registry, and then clients can learn how to process this new piece of information. And our documents create, they create the field name registry and define all of the current fields we are. There was an extensive survey taken of all of the, I don't know about all, but many, a very large number of the WHOIS registries that we knew about, and find what fields did they typically return.

And so we have this very extensive research about that, and we were creating a first set of fields based on what we found. So there is at least feature parody with what's out there now. The protocol supports the

notion of basic search. It defines a syntax for how you would issue a search query. It does not require that servers implement this.

There is a way to say back, we don't support search queries, or we don't support this type of search query. So again, for feature parody, we want to be able to support servers that currently provide search capabilities, but we don't want to force it on anyone that is not able to, or does not wish to.

So current implementation status. You just need a basic HTTP client to issue a query and receive a reply. ICANN is partnered with CNNIC to produce an open source implementation. The URL is here for where you can have a look at that work. I haven't checked lately on how far they've gotten. ARIN has had an implementation for this for some time, for querying network numbers, and I believe AS numbers. Some time ago, they noticed that the number of queries to the new service, to RDAP, has far surpassed the number of queries they're getting to port 43, so they're, I don't want to put words in ARIN's mouth, but I think they're heading toward being prepared to turn off port 43 all together in favor of the new work.

Several of the other RIRs have either prototypes, or are in private betas. RIPE NCC has their own, and VeriSign and Affilias are doing proof of concept work for the domain name registry side of things. So what are the specifications? There are six documents that are currently in IETF last call.

I think the next slide goes into a little more detail about timeline. The six of them are these: the object inventory just describes our WHOIS research, what led us to the first set of fields. One describes how to do

RDAP over HTTP. And things like authentication are accomplished this way and so forth.

The query format is how to form the question based on the question you want to ask and how to do searches as well. There is one entire document dedicated to security considerations, for – if you need to secure it, here is the way to secure it. If you need to authenticate your queries, or you want to force authentication, here is how you do that. Issues of that nature.

How to form a response using JSON, once you've retrieved the data from your backend, and then the bootstrapping piece, the part that will be managed by IANA. The URL here contains links to all of the current documents, and shows their status. You'll see them all in last call.

Now and show the last call dates and the tele chat date. So the last call will end on the 24th for all documents accept bootstrap, which ends on the 27th. A few days later, they go into ISG review, which is the last of our formal review process. So that's where the Internet Engineering Steering group will have an opportunity to look at it, comment on it, request or insist on certain changes. Assuming nothing serious happens there, it will go into the RFC editor queue pretty much right away, where lately they've been taken about a month to do it. So I would guess early December, unless something serious comes up.

In which case, you know, the end of December is more likely, but I would be very surprised if that, if this goes beyond the end of the year. And it's important to note, I think, that RDAP plays an important role in addressing the ICANN EWG Internet directory services recommendations. I think that was my last slide. Yes.

FRANCISCO ARIAS: Thank you Murray. So now that the work is about to finish now, we need to [inaudible] look on what started a few years ago, and see how we can move to implementation here in ICANN. So I just want to reiterate the difference. This is a simplified anatomy of the WHOIS protocol. We have at the top, the policy, which is perhaps where most of the force are usually here within ICANN.

We have that the data that is transported that sometimes we [out?] secure in ICANN, and define what should be a pass through in WHOIS.

And finally, at the lower level, we have the protocol, which is something that usually at ICANN, we don't pay too much attention. We just use whatever is there that is standardized by the IETF. And here is where the change is coming. This is only on the protocol. Provide a baseline that can be used to implement whatever policies may come up in the future, so that are different initiatives within ICANN, for example, the translation, transliteration, policy development process that is working on in having internationalized data inside the registration data, or the WHOIS, let's say, to use the term that most people know.

However, in order to do that, you need a protocol that supports, and that's where you need something like RDAP in order to be able to do that. And then we have the EWGG, which is steering at even an earlier stage, that may come back as one big PDP, or several PDPs, who knows? But also builds on, something the public needs to be there in order for people to be able to have the sources that are desired.

So I just wanted to give you the message, this is about the protocol, the lower layer in the WHOIS services that we are talking about in this session. So as Murray explained, RDAP provides or enables this short summary of the capabilities that are there. We have, I want to rate also that there is some things here, like differential access that we may or may not have already in agreement within ICANN where this has to be there or not.

The thing with the RDAP protocol, just because the protocol has the capability doesn't mean that you have to implement it. You implement it once, and if there is a policy that says you have to do it, a [inaudible] with internationalization, search ability, etc. So you have the capabilities that are ready to use once there is a policy that says you must, or you should, or you may use that.

And finally, this is, as mentioned, as [inaudible] mentioned in the session previous, the WHOIS session, this is an incremental step forward in the long term view that is described in the EWU port.

So this is, at this point, we don't have any concrete plan, or… The intention of this session is more about discussing what should we do next. This is a high level view of what could be happening here. First, of course, we need to have the protocol to be WHOIS as a standard by the IETF. Then we need to have implementations available, for which is more explained, there are at least two other open source implementations that, at least one of those is already being used in production.

So there are already implementations available, that people can know and start playing with, and once the protocol is finalized, and all the

features the final implementation describes are there, then we can consider them for production environment within ICANN. Then there is a detail that, a set of other details that we need to be defined.

We were using the term operational profile. We need to define certain parameters, or certain configurations on the RDAP protocol, what features should be turned on or what features should not, and so on and so forth. So we need to have a discussion and agreement on what is that need to be in the implementation that is used by gTLD registries and registrars.

And then, of course, we need to have a timeline for deployment of this new protocol. And finally, we need to start talking also about when it would be wise to turn off port 43 service. I have no suggestions here. It's just raising the questions for discussion. And there are more questions here.

So there is currently an opportunity for synchronizing implementation of two things that are in a similar timeline. I don't know if the audience here is familiar here. We started doing, let's say packaging or synchronizing the implementation timeline for things related to WHOIS. Last summer, we started with the first iteration of this. We define as six long implementation timeline. And we included the things that were ready for implementation.

At the time, there were [inaudible], which were a couple of policies that needed to be implemented. And there was this document, the WHOIS clarifications, [inaudible] related to the new TLDs and the 2013 RAAs. So we synchronize the deployment time for those two to be in six months from summer, so that's 4 15 30, I think.

And what we are looking here is the next two pieces that are ready for implementation, or close to ready for implementation, are the thick WHOIS, a policy implementation, and RDAP. And they would seem to have similar timelines that we could perhaps consider synchronizing the two of them. And also, in the case of thick WHOIS policy, there are some questions there about the potential legal issues that may, difficult having the information pass to the three registries that need to do the migration from team to thick.

And there is a potential option in RDAP to use the [inaudible] so that you have the… The way it works is the registry would receive the query for a specific domain name, and suppose that domain name has not been migrated from the register to the registry. Then you who will in turn use that feature in RDAP, to provide the [inaudible], and say, the registry will say, the information that you are looking for is not here.

It's over here, with the register. And then the client will just follow the standard HTTP protocol and would get the information they were looking for, just by querying the registry, which is very close to what the thick WHOIS policy is recommending. Another thing that we may want to discuss is, once all registries are thick one, once the migration has been completed for these three gTLDs that are thick WHOIS, we may need to consider, is there a reason for registrars to offer this service at all? Since the information, at that time, would be available in, at the registries, and they can provide the information?

Is another question that may need to be discussed. And of course, the last one is, how long after we start RDAP deployment for us to consider

turning off port 43 WHOIS service. And I believe that's all I had in the presentation. So I will open the floor for questions.

UNIDENTIFIED SPEAKER: Did you want questions or do you want answers to your questions?

FRANCISCO ARIAS: Both.

UNIDENTIFIED SPEAKER: So I guess, you didn't offer any kind of timeline for when you're going to try to move. Do you have any sense though of when you're going to offer that timeline for when we will transition to this replacement for WHOIS? I have some other things to, but we will share around.

FRANCISCO ARIAS: So I think the most interesting thing we are looking at is potential synchronization of the thick WHOIS policy implementation, and the [inaudible] implementation. So in that sense, supposing once we do the detailed analysis, things indeed are confirmed to be what seems to be now, that there might not be any impact in the timeline that the thick WHOIS policy implementation has.

Then we are looking at the same timeline that [inaudible] has, which is in the order of a couple of years.

UNIDENTIFIED SPEAKER: So I'll offer a comment to add something. One of the things that's interesting to me is the requirements currently that the WHOIS service be on the same server, accessible especially as the registrar site for registries, the website that they have to put up for registrars. That's part of PDT right now anyway. One of the requirements. So I just offer that to you as something to be aware of and be concerned about since obviously you can't run both of these things on the same server, physical server, since they're going to be on the same port.

FRANCISCO ARIAS: Actually, you [inaudible], but yeah, it's a good point. But I think you could do different based on the type of, the mind type, I guess. We're getting into much detail, I guess.

UNIDENTIFIED SPEAKER: Yeah, if you do that, you're forcing a particular implementation. You're forcing registries to implement in a certain way, and frankly, you know, we didn't like the idea that we had to put them back onto the bolt server in the first place, when they forced us to do it for new gTLDs. Because, you know, as a large infrastructure, that's not the way we do things.

You know, different services are on different constellations of machines. And that's obviously the way that one prefers to do these kinds of things. So, yeah I mean, I appreciate that technically the solution is possible, but I would encourage you to, you know, ask around and consider and get some other advice and comments about that, before committing to that.

FRANCISCO ARIAS: Point taken. I think that this could be part of the discussions about developing the personal profile. And I remember, we could even use the bootstrapping mechanism to define, so there is a way to identify what's the server that is being used for this. So we I think we have options.

GREG ERIN: Hello. My name is Greg Erin. So when we make the transition, the registries and the registrars will implement this for the contracts. But other things will start to break out there. Withdraw of the websites and applications that use WHOIS protocol now. And so, working with the rest of the world will be important.

And we should start thinking about how to convey the change through outreach, and make sure that the rest of the world knows about the open source and so forth, because at some point, things are going to break and people will not be able to get the information they need.

So how much do we know about how many implementations are out there, and who we need to reach out to?

FRANCISCO ARIAS: Do you mean [inaudible] implementations? [Inaudible] WHOIS implementation.

Good point. I don't know.

UNIDENTIFIED SPEAKER: We have a comment from a remote participant, Mark Blanchett. In the timeline, there is also some time needed for IANA to implement the bootstrap registries, and their related ICANN processes. In parenthesis, minor but not be ready the next day the IRC is published.

FRANCISCO ARIAS: Yes, good comment, thank you. Chris.

CHRIS DILLION: Thank you very much. My name is Chris Dillion, and I'm co-chair of the translation and transliteration contact information policy development process working group. And I heard you talking about…

UNIDENTIFIED SPEAKER: …title in mind.

CHRIS DILLION: Yup. I was quite worried I was going to stumble as I said that. It's quite a mouthful. I mean, I was interested to hear you speaking about internationalization, and even more pleased to hear the word UTF8 being used. And basically, one of the things that the group has discovered, as you would expect, there is quite a strong demand for the input of put contact information in the non-Latin script.

Now, there is also a possibility, and we are measuring how large that is. There is some demand, and we are still measuring, that you would have not only, and I'll use Japanese as an example, because that's the easiest

for me. So for example, if you had a Japanese address, there is some demand…

Well basically, the idea would be that it would go in, in Japanese script. There is some demand that that would also be transformed, so that could either be transliterated or translated, into ASCII. And so, at that point you get issues about, well okay, you know, which is the main form of the contact information? Is it the original or is it the transformed version.

To cut a long story short, we have a need to be able to relate more than one address. And so the question is, you know, basically, is there some way within the protocol or some other way, of doing that?

UNIDENTIFIED SPEAKER: The short answer is undoubtedly yes, the question is, did we get it right at the starting gate? So I would invite you to look at the registry that we're creating of the fields that can be returned, as part of WHOIS answer. And if the fields that are present in the registry, in a proposed registry, cover your case, then great. If they don't, then it is easy to register additional fields that could be returned and say, if I understand the problem, you would want to say here is the Japanese form of the address, and here is a, one of the trans words form of the address.

And we would just return both of them, and the client would have the option to decide which one it would render. Something like that.

CHRIS DILLION: Yeah, that sounds pretty good. And on top of that, you would also need to have some kind of method for indicating, you know, for example, when the data were put in, because, you know, for example, if you find that you've got English data that was put in three years ago, and then your Japanese data is very recent, it could be that the English data is actually a representation of an earlier draft.

That sort of issue would need to be…

UNIDENTIFIED SPEAKER: I'm fairly certain that what's in there now is just sort of a general, the last time this record was updated in any manner, type of time stamp. You could add a timestamp, a second timestamp that indicates this piece of the record was updated at this time. This piece is updated this time. The data are structured, and so you could define it, you could extend it however you like.

So if the initial version doesn't sort of suit your needs, then come back to us with, here is what you need to add to get to where we need to be, and it's a short document process. I shouldn't make that promise, but it is an one document process.

CHRIS DILLION: Thank you very much [inaudible].

JIM GALVIN: Jim Galvin again from Affilias, just to add to that too. Francisco you had put up there, you know, the idea of creating an RDAP profile along the

way, and I think that addressing these needs and these requirements would have to occur during that period of time.

And then hopefully the working group that, you know, Chris is co-chair of by then, and this particular need will all be documented and laid out. So there are some things to align as part of deploying this if we want to, otherwise it's just going to be an update along the way.

FRANCISCO ARIAS: Francisco here. I would envision that that profile would have to be updated as new policies are developed to account for the new things that are added. But yes.

JIM GALVIN: Okay, so Jim again. The question of… I just want to say once for the record, I mean, I'm sure you know this, but for the purposes of the record. Your question up there about when to turn off WHOIS deployment, you know, if you roll out RDAP. So one of the things that's interesting, obviously, is RDAP is completely backward compatible with WHOIS.

So, the transition, since Greg [inaudible] had earlier, was identifying that a lot of people lose, use WHOIS and a lot of kinds of applications and services, a fairly long transition period would be possible because you could certainly take a RDAP response and strip it down and turn it into a WHOIS thing to push back out the door. And that's just worth observing that that's there. So there is no forced timeline to turn it off. We can do that, whatever seems to make the most sense.

FRANCISCO ARIAS: Just to clarify. I guess you didn't mean to say that is backwards compatible, because they are completely different protocols. I think what you meant is that it is easy to have a proxy, right?

Any other questions or comments on the floor? No.

One more.

JIM GALVIN: Jim again. Should registrars continue to offer WHOIS services? You know, I guess, I would think, I don't know why they would have to except in the context of, how long would you offer WHOIS services if you roll out RDAP? I'm sure that registrars would be pleased to not have to do that anymore, but far be it for me to speak for them. So but I would think turning it off would be a good thing, you know, just as quick as we can, as soon as we get to thick WHOIS.

UNIDENTIFIED SPEAKER: We have another comment from Mark Blanchette, a remote participant. The capability of supporting multiple IAT team and data in RDAP has been reviewed during the study on IAT registration data.

FRANCISCO ARIAS: The IAT team we're referring to is international decision.

Okay. Any more comments? No. Well, thank you very much, with this, we close the session.

[END OF TRANSCRIPTION]