ICANN71 | Virtual Policy Forum – GNSO - (RrSG) - Preview: Registrant WHOIS experience study
Thursday, June 17, 2021 – 12:30 to 14:00 CEST

ANDREA GLANDON: Hello and welcome to Preview – Registrant WHOIS Experience Study. Please note that this session is being recorded and follows the ICANN Expected Standards of Behavior.

During this session, questions or comments submitted in the chat will only be read aloud if put in the proper form as noted in the chat. Questions and comments will be read aloud during the time set by the chair or moderator of this session.

If you would like to ask your question or make your comment verbally, please raise your hand. When called upon, kindly unmute your microphone and take the floor. Please state your name for the record and speak clearly at a reasonable pace. Mute your microphone when you are done speaking.

This session includes automated real-time transcription. Please note this transcript is not official or authoritative. To view the real-time transcription, click on the Closed Caption button in the Zoom toolbar.

With this, I will hand the floor over to Owen Smigelski. Please begin.

OWEN SMIGELSKI: Thanks, Andrea. This is Owen Smigelski. I am head of ICANN Compliance and Policy with the Registrar Namecheap. I am also the vice chair of the Registrar Stakeholder Group for Policy. And prior to my several years at Namecheap, I was with ICANN Contractual Compliance for seven years almost and I spent a lot of time focused on registrar compliance.

Prior to that, I was a trademark attorney, a member of the IPC for a number of years, and I worked on much of the working groups. So, a lot of where I'm coming from in this presentation is based upon my own personal experience and interaction with this issue.

I'm co-chair of the Registrar Stakeholder Group's Registrant WHOIS Experience Study Group, along with Jothan Frakes. However, Jothan was not able to attend this morning, so I'll be handling a lot of the presentation myself. I am joined by Jody Kolker from GoDaddy. I don't know, Jody, if you want to introduce yourself. Otherwise, I'll just wait until it's your time to jump in.

JODY KOLKER: Sure. Hi, everyone. I'm Jody Kolker. I've worked in the domain industry at GoDaddy since 2001 and a large part of what I've been doing over the last few years has been doing WHOIS research on spamming and other nefarious activities, I guess I would say. Thank you.

OWEN SMIGELSKI: All right. Thanks, Jody. Next slide, please.

So, here is a high-level overview of what we plan to cover this morning—or at least for me it's 3:30 in the morning, so that's why I keep referring to morning but understand some people are in a little bit more enjoyable time zone for this presentation.

This is what we're going to go over. We'll have time at the end for questions, so please do hold them. We'll open up the floor for them later. If you are really chomping at the bit to ask a question and don't want to wait, you can put it into the chat and we'll collect them and respond to them later. Next slide, please.

With this, I will hand it over to Jody.

JODY KOLKER: Hi, again, everyone. To get a sense of why we proposed the study, we'd like to present some research that has been done in the past.

In 2007, SSAC released SSAC 23 titled, is the WHOIS service a source for email addresses for spammers? In this report, email messages to addresses that were only used as contacts for domain registrations were collected and analyzed for about three months. Domains were registered in the com, info, de, and org namespaces. Com, info, org registries publish a zone file allowing anyone downloading a zone file to determine when new domains are registered.

De does not publish a zone file. New domains are registered using different combinations of private and non-private domains along with combinations of com, info, org mixed with de domains to determine if the publishing of a zone file along with private domain registrations affected the amount of spam received. The first test registered five com and five info domains without privacy. Com and info registries publish a zone file. For the three months, over 278,000 emails were received for these registrations.

The second test registered five org and five de domains without privacy. Org publishes a zone file while de does not. According to the report, the org registry also may have employed anti-scripting measures of their WHOIS at that time. A little less than 2,900 emails were received for these registrations, one tenth of what was received in the first test.

The third test registered five com and five info domains with privacy with com and info registries publishing zone files again. 284 emails were received for these registrations.

The fourth test registered five org and five de domains with privacy applied to the org domains. Again, org publishes a zone file where de does not. Only 41 emails were received for three months after these registrations. The amount of spam emails received was reduced by a magnitude in each of these tests each time privacy was added to a domain or if a domain was not published to a zone file wherein both options were applied for the domain registrations. I think you could see email addresses return in WHOIS virtually assured spam would be delivered to these email addresses.

In my own experience, we were able to see spam email within minutes of the domain registration. This was caused by spammers who were notified that a domain had been registered immediately after the registration. The spammer would then query the WHOIS for this email address and send an email offering SEO, website and logo design and a lot of other services and products.

In addition to spam, fake renewal notices were and continued to be sent to registrants by companies that are not associated with the registrar of record for the domain.

If the customer would attempt a renewal for the domain through these companies, the domain would be transferred away from the registrar of record to the company that sent the email. This obviously led to customer confusion and frustration. These companies were often resellers of registrars.

Can we go to the next slide, please?

Fake renewal notices resulted in many complaints to ICANN Contractual Compliance and discussions during ICANN meetings. Which in turn led to the adoption of the 2013 RAA which allowed for a compliance enforcement action which was breach and suspension of a registrar for actions of its resellers.

It's been 14 years since the SSAC report has been published. We've had two WHOIS working groups and now registrars are required to verify an email address or a telephone number of the registrant. Spammers have since added using SMS and robo-dialers to their nefarious activities. Which brings us to our purpose. Does unredacted registration data in WHOIS and RDAP continue to harm, enable fraud or increase the abuse that we have. Thanks. Owen?

OWEN SMIGELSKI: Thanks, Jody. Next slide, please.

So, it comes back to why this study and why now. We've seen that there has been changes due to WHOIS publication of data in light of GDPR as well as Temp Spec. And so, why now? Part of it had to do with some registrars and people in the Registrar Stakeholder Group as well as folks online and people complaining about still receiving spam, receiving fraud calls, robocalls, as well as unsolicited postal mail. And so, it got us thinking, why is this still happening? Because for a very large majority of registration data out there, it's redacted or it's not published or it's behind a privacy or proxy service which should mask, and as previously found by the SSAC, it should reduce those things.

So, what we wanted to do was see, is this still a concern or is the contact that we're receiving now, is this based upon legacy registrations or data that had previously been scraped from the WHOIS or is this new data that's being obtained and exploited? So, that was the concern. Is it existing contact information that had been harvested, or is this stuff that's still being mined and utilized by parties for purposes that sometimes is legal, sometimes illegal? Next slide, please.

So, part of the thoughts that we had coming into this was that from the registrant information—I apologize, there's a lot going on, on that section, but there's a lot of things that we have observed and have had complaints about from our customers. There's a number of things that they can get through postal marketing. There still are renewal scams. Other types of letters and correspondence providing services that generally are overpriced or to the extent that they are paid for, may not actually occur. And so, there's concerns with that.

There are telemarketing things as well too. Now with things they did not have back when SSAC originally did the study. There are robo-dialers and ways to mask telephone numbers for calls that go out. There's also some very significant fraud now where you get calls complaining that there's a criminal investigation or that there's something involved with hacking of say a Facebook account or something along those lines which you don't need to have too many people to fall for that for that to be a successful exploit.

There's a lot of business compromised emails that are sent out and also then get on all the spam list and all those other things that—so, the expected result was that with the increase in privacy and redaction of data appearing in WHOIS, as well as RDAP, that these things would go down as that information is not published, and it's not available for harvesting and exploitation. Next slide, please.

So, as you may see up on the screen there, that's one of the—I don't know if you can recall the so-called fake renewal notice. I don't like that term. I didn't like that term when I dealt with this in ICANN Contractual Compliance. However, I haven't yet come up with a good replacement term or one that's commonly known by the industry. So, I'll continue to rely upon that but what you see there on the screen is an actual renewal notice that was received recently by somebody in the Registrar Stakeholder Group.

It was sent to a postal mail address and was a solicitation to renew a domain name. It said it was expiring. It said you need to pay. It does state and you can see it in bold font that says, this notice is not a bill.

But you can imagine that the more of these get sent out, the more that people would be more susceptible to actually go ahead and do that. You can see—it may not be clear on the screen, but the cost for renewing an org domain name is $50 for a year which is significantly more than is charged in the marketplace. So, there doesn't have to be that many people who fall into this trap in order to make this a very, very profitable initiative.

And this went in some back-channel communications with others in the ICANN community. This is not the only one that was received. So, we just wanted to publicize that this is still going on. Perhaps not as widespread or as much of concern about—the reason for the breach regarding the previous round of fake renewal notices was the deception and people were conned basically into thinking that they had to transfer their domain names. There's less of an urgency in these notices. And so, I can't really comment on whether or not it would be compliant with the RAA and previous ICANN enforcement activities.

However, there's concern that people who may not be as sophisticated might receive this and think that they have to do that. So, seeing this, it kind of made us really want to take a look and explore whether or not this is a concern and what can happen.

So, what we've been talking about for several months is trying to come up with a type of way that we can identify this and measure it and get some accurate information and some unbiased information in terms of—we've got some ideas about what we think might happen as a number of the participants in this study within the Registrar

Stakeholder Group are domain name registrants and have seen this. I myself actually noticed an increase as soon as I registered a .us domain name. I almost immediately started getting calls to the number that I had to provide. For those who may not know, .us does not allow for privacy proxy.

So, we wanted to gather this information, see if this is a concern and then share the data and results with the ICANN community so it can help direct and guide current as well as future policy discussions. Next slide, please.

So, why is abuse of registrant data still a concern? And again, as we all know, the GDPR and the Temp Spec have significantly reduced the amount of data in there. However, the data that may be present or may have been previously scraped is—I'm going to use the big air quotes, "better data" for marketers at least from their perspective because of the 2013 RAA which has significant improvements from previous versions of the RAA in terms of the quality of the data.

Registrars are required to verify either the telephone or email to make sure that that information works. And if it does not work from the verification process or whether it's identified as inaccurate, then there are some specific required steps that registrars must take to either correct or suspend that domain name.

So, from that perspective then, any data that can be found would be a lot more valuable as it's going to be a telephone number that works, an email address that works versus previous versions of the RAA where there could have been a little bit more wiggle room in terms of the accuracy of that data. And again, even at post the GDPR Temporary Specification, data is still there and can be there in non-redacted format whether it's by choice or by accident. Hopefully, registrars are gaining consent before putting information into public registration data services. And then, again, as I mentioned before, some ccTLDs require unmasked registration data and you can't use a privacy proxy provider.

I wanted a .us domain name. And actually, in part, that's one of the reasons why I ended up getting a PO Box, much to the surprise of my wife wondering what kind of nefarious purposes I was doing. I was just trying to register a domain name and not have people start sending lots of spam or unsolicited postal mail to my house. Getting this information, scraping from WHOIS is very low cost. It's not a lot of overhead to do this. It could be done through some type of automated process that can crawl and go through various zone files and things like that and look up.

So, it can also leverage existing databases when you're looking at sending spam. If you have 500 million email addresses which is not too—could potentially be something you could obtain from scraping and going through a large amount of domain names. That's not that much in terms of cost to obtain especially when—even if you get a 0.0003 response rate through some type of scam, that's still pretty good return. There are sources out there that have these databases that are continually being updated and being resold and repurposed and retooled out there.

And people who are collecting this data are certainly not doing so out of public interest. They're doing it for purely financial purposes and really there's no concern on that. So, wanted to see what's going on and does that still exist? There are some—I want to say legitimate, perhaps legal purposes for using this data solicitation for search engine optimization or transferring to the registrar. I won't go about whether or not that's violating terms of service that may be present in disclaimers in WHOIS or RDAP or otherwise. And also, as previously highlighted, this data is also being used for illegal purposes. That is also a significant concern. Next slide please.

So, right now, we're in the very beginning phase here. This is kind of an overview of what we're going to be doing. We're in the process of doing a study design as well as testing, doing a proof of concept. We have done a couple of tests registrations to see whether or not something pops up or we start receiving postal mail, email, phone calls, etc. So, what our plan is—we'd be ramping that up after this ICANN meeting—what we're going to do is create and test clean personas.

What this means is we're not going to be using telephone numbers that have currently been used. We're going to not use postal addresses that had been used for registration data purposes. We're going to create new email addresses. And this way, we can ensure that the only place that this data is being placed anywhere even remotely online is via registration data through either WHOIS or RDAP. And so, for this, we're going to have to create new email addresses and telephone numbers, and to the extent possible, we're going to remove spam filters or telephone number blocks.

I'm very, very happy to see now that my cell phone provider will actually label calls coming in as spam risk which makes it very easy for me to decline those calls. We try and remove those just to make sure that we're able to see this data and we can get these types of inquiries in there. Again, I use a postal address that's not currently used for any type of domain name registration or is not used elsewhere.

Of course, this is all going to be valid information. We're not going to go astray of having inaccurate data or anything like that. So, we'll be able to have this stuff there that we can use and share in there. And so, what we plan to do is once these personas are created and this contact information is created, we're going to test for two weeks to confirm that it's clean to make sure that we don't receive phone calls randomly in there or at least to measure there to an extent if there is a random call from the robo dialer, etc. Just to kind of measure it and see what's there. See if any spam arrives. See if there's any mail that does arrive.

Then the next step, what we're going to do, have several people on a team. I think we're about six or seven. We're going to register a number of domain names through several registrars and either choose to not have privacy or to immediately remove privacy. We're also going to utilize registrars that have implemented the part of the Temp Spec that allows a registrant to expose complete registration data in the public WHOIS/RDAP.

That way we can ensure that this is a place where these registrations are actually—all of the data is present in WHOIS/RDAP in its complete, unredacted and complete form. Not sure to the extent of whether all registrants have done this. Speaking from Namecheap, I know that we have implemented that, but we'll be sure to choose registrars that allow us to do this. We're going to select several different TLDs as well as registrars. And we are going to not just be US-centric as I think a number of people on the team are. But we're going to try, to the extent possible, use registrars from around the world just to ensure any type of bias that might be there as well too.

And then, the next step is we're going to document the contact. We're going to wait and see what happens. We suspect that we are probably going to start receiving spam or calls or postal mail, but we're not sure if we will. We're also not sure exactly what type of contact we're going to receive. Is it going to be for legal purposes? Is it going to be for less than legal purposes? Again, our suspicion is that we will probably receive this but we don't want to bias our study or our approach. And so, we're keeping an open mind in terms of what might happen.

And then, as we see—I'm not implying that us on the Registrar Stakeholder Group are like Mr. Bean, but I just wanted to find a humorous image of somebody doing desk work. So, what we're going to do then is we're going to collect that and then we will do an analysis of what our results are and then be able to collect and report back hopefully. We plan to have a session at ICANN72 and present the results of this. Thanks. Next slide.

So, our next steps is that we're holding the session here and we're welcoming participation and feedback from the community with suggestions [inaudible], [inaudible], etc. Whatever people might have there. What we'll do then is we'll take that information and feedback and then use it to finalize our methodology and approach.

Then we're going to go ahead and create our various personas. Then, the fun starts. We are going to register domain names and collect the data. And as I indicated previously, it's our goal to have this ready to present a report at ICANN72. Next slide, please.

So, that is it in terms of presentation here. So, I'll open this up to Q&A. I see there is some stuff in chat although I have been completely and willfully ignoring it. So, I'll have to scroll through here and see what's going on here. So, scrolling through here, I see a question or a comment from Susan Payne. Are you going to completely use different set of registration data? Yes. Our plan is to not use the same data. We're planning to create different personas. Multiple personas so that we can use different ones for different tests and see whether we get different results or how that works.

One may be a Gmail address. One may be a Google voice telephone number. But we'll certainly vary that. We may use a non-common email provider as our own mail provider, so that it's not one of the Google, Hotmail, etc. One of those. We use our own just to even out variables in there in terms of how or why so many may come across that registration data.

I see in the chat also from Chris Lewis-Evans. Are you looking at geodiversity of registrants as well? In the team that's doing this, we have participants from North America as well as Europe. So, our goal is to hopefully be able to use diversity from those two regions there so that it's not just the United States that's being present in there. We're limited in terms of our ability and our funds to have, say, a Chinese address or something along those lines. So, to the best that we can, we'll do that.

I'm reading a comment from Ching Chiao. "Sorry if I missed anything. I thought the session was about registrants' experience in providing/inputting contact info. It seems that the purpose is to initiate a new honeypot project in which lots of ccTLD and many service cybersec companies have repeatedly done so. "

Yes, I know some of these efforts may have been duplicated out there elsewhere by others. But we wanted to do something in terms—specifically looking at the redaction versus not redaction there. There's been a lot of talk about abuse and what's being done to combat abuse and that registration data must be out there and published so that people can view it and see it.

Part of the concern at least on a number of registrars is that this data is still being used and exploited for improper purposes. And so, we wanted to just—none of us at least in the Registrar Stakeholder Group have done this and we wanted to take our own approach and look at it and do that.

So I think going through the chat—oh, I see a question from Chris. Go ahead.

CHRIS LEWIS-EVANS: Thanks. And I thought I'd talk to save you talking through the questions and everything else. So, just one other further clarification. You said about using a number of different registrants to register the domain names. Are you going to be utilizing resellers as well? Because I know some of the extra sort of layer that that puts upon or the difficulty in getting contact can sometimes change the data that you actually see.

OWEN SMIGELSKI: Thanks, Chris. I don't know if we've specifically put that down as something that we've planned to do but it's certainly a good way in terms of ensuring that we have diversity of registrars to use more than just a retail registrar. So, I'm sure that we can add that to our list. And I think we've got somebody taking notes here, so we'll be able to take that back and include that in our design to ensure that we do have that type of increased diversity of type of registrar in there as well too.

Let's see. There's a question in chat from Syed. Discussion on WHOIS data is continuing since many years and ICANN also launched different programs and studies for this particularly in terms of completeness, accuracy, and privacy. Can you tell us two to three main key reasons that still domain registrants are facing challenges due to WHOIS data problems?

So, Syed, I do know that a lot of the ICANN initiatives, and I speak from a personal experience, when I was with ICANN Contractual Compliance, I was on the steering committee for the WHOIS ARS which I think that acronym stands for Accuracy Reporting System although don't hold me to that. It's been a little while since I've dealt with that. And that was a system and a process that ICANN came up to—and this is based upon directions and recommendations from a review team if I recall correctly. And that was to look at the accuracy of registration data that was present in the WHOIS which was the only service being used at the time to provide that data.

And so, part of it was to measure and then test the completeness and the accuracy. So, to check to see if the email address present in WHOIS worked. Check to see if the telephone number present in WHOIS worked. And then, check to see if the postal address that was provided in WHOIS was compactable. Didn't actually send things to that, but it was reviewed in terms of whether it appeared to be [inaudible] that would be deliverable.

And so, a number of rounds of this was conducted by ICANN and reported to community. So, if you're curious about that, you could just do an internet search for ICANN WHOIS ARS. There were I think six rounds of the WHOIS ARS. And found that overtime that there were increasing levels of accuracy. And some of that was in part to the efforts of ARS. Incorrect records were provided to ICANN Contractual Compliance for follow-up and correction.

And then, as well, as more registrars went to the 2013 RAA, those types of data became more accurate. And then, also, as new registrars were onboarded and new registries were onboarded, there were some, let's say, growing pains in terms of how data was being collected and reported and that improved over time from ICANN outreach efforts.

So, that was the concern, was that there was—this information was in there. The WHOIS ARS has been paused pending the GDPR and Temp Spec as there are some concerns with the GDPR about whether or not this is something that ICANN is allowed to do. So, the reason for checking this is again, just because information may be redacted, either not present, there is still the ability to have that registration data present whether or not somebody completely understands what is meant by putting the registration data in there or it's possible to use a privacy proxy service.

[inaudible] still have data there in—have some type of contact information. There may be an email address as opposed to a link to a form. So, there may be some types of data that are coming through. And again, one of the reasons why this has been triggered is, people in the Registrar Stakeholder Group still continue to receive these types of—the so-called fake renewal notices, robo calls, telephone numbers. We're still seeing the types of things that happened prior to the GDPR and elimination of the Temp Spec. We're still seeing these types of exploits happening.

And so, it was just kind of more of a curiosity about whether we have done this. So, again, we've kind of done a proof of concept on this. But moving forward, we're going to ramp it up and do some more checking on that.

Next, it's actually a question in chat from Mason Cole. It says, "How do you take into account whether spam filters, block list prevent the spam from getting to the registrant?" Yeah. Again, Mason, that's a good question. So, part of that is utilizing either—if it's possible to remove or reduce filtering if you're using, say, Gmail. That's possible too. You can do that. And then, also, that's another opportunity for us to be able to utilize a different—not widely used email provider or our own that we set up. Some of us are tech guys that are involved in this. And so, we could set something up where [inaudible] bypass.

We have seen some significant amounts of spam already associated with this. Even utilizing, if I recall correctly, Gmail. I then wasn't able to put it in the slides only because of the inbox has so much spam in it that the graphic we had was almost unreadable in terms of the massive list of literally hundreds of messages of spam. A lot of the Bitcoin scams as well as some other exploits that we commonly see. So, I hope that answers your question.

Again, we're going to do our best to try and reduce that. But again, not everything is perfect and we may not be able to avoid all of that.

So, question from Werner Staub. "Is the study limited to the crude publication as opposed to redaction of the registrant's day-to-day email address? Crude publication is going to be a bad idea in most cases. So, studying that will not bring a lot of new insights." We were not planning on doing any type of redaction at all. We wanted to see, is this still a thing? There is to a degree a large amount of these databases that have already been created and generated by the various—I don't want to say service providers but the people who provide these types of services.

And so, that was the rationale for utilizing brand new telephone numbers, email addresses, and postal addresses that had not been present there to determine whether or not these lists are still being added to and whether or not it should be a concern that registration data is being exploited still and collected and harvested for these types of legal and sometimes illegal opportunities.

Comment from Chris Lewis-Evans. "It may also be interesting how much spam would have been blocked by standard tools deployed by email providers, etc." Yeah, that and as somebody who works at a company that's got an email filter, it's not perfect. The way that spam filters work is generally, you're looking for an IP address of the sender that's known to be "bad" and that's in reports. You're trying to filter out keywords as well as content, length of message, etc. But they're imperfect. It's not possible to always prevent spam. So, it's possible that some of that might be prevented, but that's also one of the reasons why we would intend to have no ... removal of spam filters from that.

So, I think I have gone through everything. Either comments and/or questions that were in the chat. If I've missed anything, please do let me know or if anybody has anything else that they are interested in bringing up, please do so now. Okay. I'll move on to the next slide.

So, I don't see any more questions or comments here. I'll wait another minute or so to see if anybody has anything. Do appreciate some of the questions and feedback that have been provided during the session. We'll incorporate this in as well too. If anybody doesn't feel comfortable jumping in now or something else comes to you later on, our email addresses for myself as well Jothan are there in the slides. So, you can feel free to reach out to us and provide feedback as well too.

We'll probably be kicking this project off in July. So, if you do have anything that you want to provide, please do so sooner as opposed to later so that we can make sure that that's included in there. And then, we'll be launching this and collecting data and hopefully be able to report back to you at ICANN72 whether that's virtual, in person, or however that ends up being.

So, thank you everybody for coming. And unless there are any other questions, we can, I guess, end the session early and I'll give you a little bit back of your time.

[END OF TRANSCRIPTION]