ICANN
Transcription
GNSO Webinar on ICANN Domain Abuse Reporting Tool project (DART)
Wednesday, 14 June 2017 at 20:00 UTC

Note: 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: https://audio.icann.org/gnso/gnso-dart-14jun17-en.mp3 Adobe Connect recording: https://participate.icann.org/p5mei4y26f9/**

The recordings and transcriptions of the calls are posted on the GNSO Master Calendar page http://gnso.icann.org/en/group-activities/calendar

(Terri): Thank you. Good morning, good afternoon and good evening. Welcome to the (unintelligible) webinar at the ICANN Domestic Abuse Reporting Tool -- DART -- with presenter Dave Piscitello, taking place on Wednesday, the 14 of June, 2017. In the interest of time there will be no roll call. Recording will be posted on the GNSO Calendar.

I would like to remind all participants to play state your name before speaking for transcription purposes, and to please keep your phones and microphones on mute when not speaking to avoid any background noise. With this I'll turn it back over to our presenter Dave Piscitello, please begin.

Dave Piscitello: Thank you, (Terri), and thank you all for providing me with this opportunity to talk with you about a project that has been maybe a year and a half in the development and five years in the thinking phase. We call the project the Domain Abuse Reporting Tool, and I'll use the acronym DART to save some speaking cycles.

Let me give you a little bit of background before I use the slides. I mentioned that this was five years in the making -- and one of the reasons why, you know, we've been thinking about this for quite some time is that we've, you

know, we've encountered numerous private sector reports and reports from various individual security companies on the state of abuse in the top-level domain and the state of the DNS. And I was frustrated because most of them fell shy in one form or another. Either I wasn't very happy with the quality of the data, or I wasn't very happy with the scope.

And so, I began talking with, you know, some other people from the community and from security and operations and from registers and registries as far back as Dublin. And we were sitting, as is the case very often -- in a bar, talking about all the limitations. And we suddenly said, "Well, let's really do this in a comprehensive fashion."

So that was really the inception of DART and let me give you a kind of high-level overview of what the project encompasses. DART is a platform for reporting on domain name registration and abuse data across TLD registries and registrars. And as I said, one of the goals of DART was to try to do better than many of the other studies.

And so, we wanted to not just cherry pick -- or pick out one or two registries or registrars -- and focus attention on those. We didn't want to create a Top 10 list. We didn't want to create a main and chain environment. What we wanted to do with the complete census of all the TLD registries for which we could obtain zone data and registration data. So, that was - that in itself is a very, very large project, collecting all the data on a daily basis and managing, you know, the database for just those data are – is a big task.

The other exercise that we wanted to do was begin we didn't want to get into a situation where we were relying on one list and one perspective because the community that uses the domain name space -- and also uses reputation data to determine which names to trust or distrust -- use a very large number of reputation or block lists. So we decided to use a very large number of block lists, and I'll show you that list on a later slide.

The other concern that we had and one of the things that always frustrated me was that almost every study is a snapshot. And almost every study takes a day in the time, or day in the life, or a very short measurement period. And nothing provides some sort of historical look, so that you can get a sense of change or clustering or the impact of some external event. So we had – now have data for eight months and we are slowly get - finishing and collecting some of the ancillary data and I'll explain that in a subsequent slide. So we are well on our - on the way to having what I believe is one of the first long, persistent stores of registration data, zone data and reputation data that can all be used to analyze abuse and registration behaviors.

Another challenge that I had when I was looking through the academic community and looking for quality literature was that I could never find a study that looked at more than one security threat. And so we decided that what we would do is we would look at, you know, the (unintelligible) as well as spam. And I'll explain exactly what we do, again, later on in the presentation.

The last thing that -- as a scientist and software engineer at least -- bothered me was that almost all the studies never made their data opened for - available – for review. They never really explained their methodology in a matter where any party could reproduce (unintelligible) we - what the study did. So our goal was to use, you know, commercially available data or open data and to make our methodology known and explain what we were doing in a manner where if somebody wanted to validate our results, they could go and they could implement the exact same project. And so, you know, this is basic scientific approach. This is unbiased data, transparent methodology, and a reducible mechanism.

So I want to put what we're doing in context. You know, this is a research project. This is part of the ICANN Open Data Initiative. And if you're not familiar with that initiative, what we – in the office of the - chief technology officer are txt – are expected to do as part of our, you know, part of our – the

scope of our work is to provide access to data that the ICANN organization or community creates or curates and provide data that will -- in some way or another -- help the community make informed policy decisions.

So as I said, we use data from public, open and commercial sources. So we use DNS zone data, we use WhoIs data, we use certain open source reputation data feeds, or block lists as many people know them. We use some commercial feeds. By commercial I mean that these are feeds for which you pay a license or a subscription, so you can access the feed on a regular basis.

And to the extent possible, in cases where we have no limitation on the new distribution of our DART related data, we anticipate we will make the data reports and publish these periodically and include them in the Open Data Initiative. So in certain cases we are allowed to use the commercial feeds in a manner that - where we're allowed to share derivative data, so the findings of summary analyses that we have. We aren't to just proxy access to those feeds to the other parties. So as we tried to work with the community to understand how should we use this project to better inform the community, we'll have to understand whether or not that's in the scope of our current licensing.

So I've said this at least twice, I'll probably say this on every slide. The purpose of this project is to provide information relating to domain name abuse and abuse of the registration system, you know, to support policy development processes. You know, I personally would love to see the data used in a manner that would allow us to drive abuse to zero. I know that that's, you know, wishful thinking, but you may as well have, you know, a very, very long reach aspiration when you have put together a project of this size. How do we anticipate the data to be used, or reports of the data to be used?

Well we think and we've seen from the data that we have now that we can identify, you know, threats that are reported at a TLD or registrar level for every TLD that, you know, for which we can obtain data, we want to be able to track the security threats over time, registrations over time.

One of the things that we think we will be able to share and work with the community to better understand is what, you know, what part of the name space is being targeted or being exploited, where are the spikes or changes in activity and what event caused those things, what kind of patterns do we see in registration behavior that we associate with abuse? Is there flocking behavior – – meaning is there – do criminals or malicious actors tend to focus their attention on one part of the name space more so than others and what are the causal agents for those kind of activities? That's a very far-reaching and very high aspiring, you know, set of objectives. But you know, once we have these data, I think we have, you know, a very, very promising opportunity to do something fairly dramatic for, you know, for the community and for the name space.

I think that we will be able to help operators understand and consider how they can manage their reputations. We're not going to tell people what to do. We're going to share the data and let them decide what they do, but sometimes -- as I've seen with new TLDs in particular, they are not as familiar with, you know, with name space abuse as some of the legacy TLDs. And even when, you know, we talked about this at the ICANN DNS Symposium there were operators who were very interested in learning more about, well, how did you do this, you know, how did you collect this data?

And I think that this will also help the ICANN community work with the operational security community. The operational security community is often very critical of the ICANN community. And having a common, you know, common set of data to examine and consider and to work together with to understand what, you know, what each community's role in abuse may be very, very helpful.

So I mentioned that we use zone data. We are consumers of all the public means of accessing name space data that the public uses. So we use the Centralized Zone Data Service and we use Legacy Contracts to gather zone files on a daily basis, or more frequently depending on the TLD operator's behavior. We currently collect zones from 1,241 TLDs to use, and we're fast approaching 196 million domains. And the largest operation that we have and one that is very difficult to manage because of some of the constraints, both policy and process in the centralized zone data service is collecting zones.

We also use WhoIs. WhoIs is the means that operational security and business communities use to identify not only the points of contacts and name servers for domain names, but the sponsoring registrar. And so, we obtained the sponsoring registrar from the public WhoIs records for the DTLBs. And one of the things that we focus on -- because we are interested in security threats or abuse -- is name that resolve. So, the names that we use in our database are names that are resolving to IP addresses. The reason why we make that choice is because these are the names that can be beneficial or harmful when visited. You know, if the name can't resolve, it can't actually pose a threat. So that's our philosophy.

We use a very large number of rotation data sets. We used 20 sets. And I'll talk about those in a moment. We spent a fairly long time looking at different reputations feeds, considering the way that they process their data, considering, you know, whether or not the block list is curated or user-submitted. We went through paper after paper from academia to see what, you know, what data they most often employed. We try make certain that we were just biasing toward academic uses, because academians often used free reputation data feeds. But after months of going through literally 86 feeds, we chose 20.

One of the things that we did as a skunk works inside ICANN before we made our final choices is we actually ran through several -- maybe three or

four dozen TLDs on a daily basis -- and we would check names in those TLDs against 86 block lists using a scripting tool that we built in house. It was ponderously slow because it was just the scripting, but it allowed us to get a good sense of the quality and character of the, you know, of the reputation data feeds. And I think with some confidence we can say that we're using feeds that are most popular and most commonly employed. And that is really our goal.

We want to have the ICANN community see the name space the way that people who are applying block lists see the name space. So choosing the names – the block list that are most commonly used, that have the highest confidence between academia and security you know, system manufacturers who incorporate these feeds into equipment like firewalls, or web application firewalls or DNS firewalls was important to us. And so again I'll emphasize that was we try to do is have DART reflect how parties external to the ICANN community see the Domain Echo System.

The other thing that we want to do was, you know, was make certain that we could, you know, we could solve problems that we can might encounter if, you know, we determine that a new reputation feed was a candidate for inclusion, or an existing reputation feed became something that we, you know, something suspect.

We're actually seeing this. I've been looking at repetition feeds for 20 years. We've seen feeds that has started out really, really doing well. And because they weren't very well-funded, they were an academic project that just sort of low steam. The feed got stale. So we want to make certain that our framework is extensible. We're fairly confident that we needed to, we could actually rerun on single day or multi-day efforts our entire engine for parsing security threats using subsets of our reputation feeds.

The two things I want to point out, one is when we say DART in the reporting tool, we mean that we generate reports. We don't mean that users or the

community would actually submit reports to us. The DART project is not a block listing service. We use other people's block listing services. We're reflecting the, you know, view of the consumer -- or the organizations -- that need the name space, not our own generated lists.

So we don't investigate abuse and then identify a domain as abusive and categorize it and put it into DART database. Instead we use data that is being driven by, you know, by industry recommended mechanisms for distinguishing spam, identifying phishing, identifying malware hosting and identifying things like algorithmic generated domains. So by using these multiple feeds, one of the things that we are able to do is not only generate a count of unique, you know, abuse domains -- or domains that have been used in a malicious manner -- but we can also count spam, phishing, malware, you know, hosting, and bot nets. So those are - were for our goals.

We can also create histograms. We currently have, you know, histograms for most of the generic (unintelligible) domains back to January 1. And, you know, and they're very interesting. We are also doing something called a cumulative abuse domain count. This is distinguished from the running count because this is every domain that has been identified as a security threat from January 1. And so what this helps us do is understand, you know, understand churn. You know, how often is a – how often are people going back to a registry and registering new names after the old names have been used or blocked? And so, that will be something that I'm really interested in doing some serious research with. I would like to have some partners to do that.

We do a lot of (unintelligible). We are very careful. You know, were very careful to – in our unique domain count, to make certain that we don't - that if a name appears on the list it only gets counted once. If it appears on two lists or five lists, it only gets counted once. So the - this is a small slide to read. I'm happy to talk to anyone off-line, you can all contact me by email. And I can share with you how we use this list. But most of these lists provide not only a

domain and identify it as, you know, as malicious, but they classification using something called the return code, which identifies the domain of a phishing domain or as a compromise host or as a malware site or as a malware CNC server.

So, I mentioned earlier that we did a lot of research in, you know, in – before we concluded that using a very large number of reputation data sets would be valuable. One of the landmark and most convincing arguments to do this that we found was a paper by (Barnett), (Capp) and (Spring) called, "Block List Echo System Analysis", and they actually wrote this paper and revised it three years in a row since 2012 to 2014. What they found was that it was very rare that a name that had appeared in one block list would then subsequently appear in another block list. And when we ran our tests in house, we came up with the same results. Now every once in a while we would come up with domains that would appear on two block lists or three block lists. But if you have 86 lists and you're trying 86 lists, we never came up with anything close to 86 lists listing the same domain. The reason for that, you know, among the reasons for that is that most of the block list operators generate – or generate their lists using technologies called spam traps, they have large spam, you know, spam trapping networks. And everybody spam trap covers a different space. They also have – some of them have regional biases with their deployments. And so, we feel we're getting a very, very large percentage of the abuse that is being reported.

Well, you know, having said that, obviously no one sees all the abuse. And we don't claim that we have all the abuse. We claim that we, you know, that we are catching quite a bit of it. We're certainly catching enough to make some, you know, some, I think, reasonable judgments about how the domain space is being exploited, who was being targeted and the like. We're not really interested in doing things like ranking or scoring or coming up with a measurement that pits one registry against another.

What we do have a very basic scoring called abuse score. And what it really is representative of a percentage. The goal there is to be able to identify the mean or the average of this score, and be able to see how far from the mean -- what deviation or quartile a registry or registrar is -- to just sort of understand if there being targeted, but mildly, or being targeted, you know, rather extremely. And that's the - that's, I think, important characteristic for people who are interested in understanding, you know, what's happening in the operation.

So the abuse score is a relatively simple calculation. We, you know, we take the abuse listed domains and the TLD on a given day. We divide that number by the domain of the TLD zone that are resolving that day, and we multiply it by a hundred. And to give you an idea what that looks like, for the 1,241 top-level domains, the average abuse score is somewhere around .6. There is a very long tail right now of top-level domains for which there are no reported incidents. So that number is – I wouldn't say artificially low, but it is low because there are –- as of May 31, for example, we had only 356 of the 1241 TLDs actually having reported security event for the month of May -- so we have to a tail almost 900 zeros to calculate that abuse score. And that's fine because we still, you know, have an average, you know, and we use a logarithmic scale representation. I hope to have some of those representations available to show people, you know, shortly.

But what we can then see, you know, just giving you some anecdotal insights into what the data revealed to us is the vast majority of the top-level domains cluster around the average score. There is, you know, there is some number that go from .6 around .8, or I'm sorry 8, which is what we call – what we would identify as the yellow bands. And there are about 25 top-level domain registries that have very, very high scores. And, you know so were trying to understand, you know, why that is so. You know, we want to work with those operators and share the data with them, and say this is what we see. And you know, and certainly that I think is valuable input. And the same score for registrar.

Currently we have been struggling with contending with rate limiting on WhoIs service. One of the things that we that's the big challenge for us is that when we ramped up this project we had to actually start getting 8 or 900,000 WhoIs records to backfill the abuse record that we were identifying. And I was – we finally got caught up this last week. And so, we are now keeping pace with, you know, with the WhoIs data of new registrations. But that could change if, you know, if the numbers change or if the rate limiting changes.

So rate limiting, you know, that everyone else experiences in the operational security and business world is something that we have to contend with. I understand that sometimes registrars have to contend with the same problem when they're asking WhoIs from other registrars. So I think that SSAC is going to bring a discussion to Johannesburg on this subject. Certainly what we do to overcome the rate limiting is what most people do to overcome the rate limiting. We don't necessarily use what you call a bot net, but we use a lot of different machines with a lot of different IPs, all trying to get WhoIs within the rate limits. So it's a very, very big effort. And so, that in a nutshell what we're doing.

You know, we feel very confident at this point that our registrar - registry data is fairly stable, and that our pools of data are reliable. And our feeds are reliable without some of the issues that we have there. I'm optimistic that by Abu Dhabi we have resolved our registrar identification challenge, identifying and sponsoring registers by WhoIs. And we'll be able to generate, generate the same kind of data that we have for registries today for registrars by Abu Dhabi, or earlier.

So the way that we've deployed the system right now is that it is a database for which ICANN staff only have access through a administrative console. And that web interface is something that I am more than happy to sit down with any individual or a small group of people and, you know, share. Obviously there's sensitivity to the information that each registry registrar has

in that information. I want to respect at until policy says, you know, how else we should use the data. But I would love to be about to give people opportunities to see what we do. We will be generating, you know, visuals that are not identifying any of the operators, but identifying the trends. But the access is internal use only. There is none of this data are going out in front of public, except in the form of presentations.

We have – I did give this presentation to the audience at the ICANN DNS Symposium. And since that time we've had interest expressed from ten registries -- CCTLD registries who would like to participate -- and I'm giving this presentation in Johannesburg to be CCNSO on Tech Day. So if you like to come, or you would like to tell people from the DNSO who weren't able to attend today, you know, I'm going to be talking to the DNSO on – I'm sorry, the CCNSO on Tech Day. And I'm happy to have anyone in the off audience. I'm fairly confident that that's an open session.

If you, you know – one of the things that will come moving forward is – I just got our legal review of our Frequently Asked Questions document that goes into what I covered in this presentation more detail, and I have a white paper that goes into even more detail. So the FAQ is about 5 or 6 pages and the white paper is about 25. And the white paper and FAQ are, you know, should be reviewed in short order. They will both be published at the ICANN web pages for the SSR team and I'll make certain that the GNSO staff can post to the GNSO list when those documents are available.

The - moving forward, you know, we have this enormous amount of data. We have this opportunity to go do things for you. And what we're – but I really can't understand from the community is how you want us to use this data? What do you want us to look at? How would you force stakeholders using the data? What kind of access would you want with the data? Or routine reports that you're able to review and then come to the SSR team, you know, or other staff in ICANN to use and consume. You know, is this something, you

know, we believe this is a – the day here can help with the SSR review. It can help with the Consumer Confidence and Trust review.

And there is - there are couple of people to come up to us in Madrid during the DNS Symposium and asked whether or not this was – the data here was sufficient to meet Spec 11 obligations. And I said, "I don't know. We would have to sit and talk with you about what your obligation is and what you need to do." But certainly there's a lot of commonality in the data that were doing processing.

So that's it. I didn't want to use all of my time because I know there's going to be just an enormous number of questions. And – I'm not certain, (Terri), how you want to do this. Do you want to read the questions, do you want me to read the questions?

(Terri): Hi, Dave. Completely up to you, but I'm happy to help out if that helps you..

Dave Piscitello: Well, let's see. You know what, why don't you read the question because I'm on a smaller screen and I can't increase the size of the chat lists, so if you ask the question, I will, you know, I will try to answer it to the best of my knowledge.

(Terri): Certainly, very happy to help out. The first question is from (Mathias Spicer). The question is did the system look for all PLDs or just the new GTLSa and the source data?

Dave Piscitello: All the TLDs for which we can collect zone data. So we get legacy TLD data, for example, from com and net and org using the mechanisms that those operators provide for us. And we uses the CCDS for all the, you know, for all the new TLDs. And so as I said, we currently have 1,241. And you know, and I mentioned that we – my goal – I reach – you can tell that I reach for a lot. And I'm ashamedly anti-abuse and I would like to see us, you know, see us work together to make this better. And so, I would love to see all the CCTLDs

participate. And I think this would be healthy for the Echo System if we were to, you know, if we were to be able to reach out and get the CCTLDs -- and they would see of value in doing this -- then that would be a win, I think, for the entire community.

(Terri): Thank you. And our next question is from Maxim Alzoba. The current ICANN (unintelligible) looks like spam and separate from security threats. Is going to be changed?

Dave Piscitello: So there are a couple of – I'm glad you asked the question because it's always the elephant in the room. Spam is a very misunderstood security threat because most people (unintelligible) who are not really invested in the securing operational community don't – aren't able to distinguish the kinds of spam that are truly unsolicited commercial or unsolicited malicious mail messages.

One of the things that we've seen over time is that from the point at which the (unintelligible) has issued their communiqué I may have said that they want to, you know, security threats such as phishing, pharming, malware, and bot nets all measured and investigated. I and Greg Aaron wrote what we thought was hopefully a clarity-lending article, and if you're interested, I can share that link with you about how the world really measures these things.

And one of the things that we point out is that spam is actually the delivery vehicle, so to speak. You know, not to get too caught up in weaponry, but, you know, in some respects, you know, split spam is the gun and phishing and, you know, and malware are the bullets, because we all see phishing email. We all see spam email. The majority of counterfeiting is distributed – or counterfeit goods, solicitations and advertisements are distributed using spam email. The majority of, you know, ransom ware is delivered as an attachment to, you know, to spam email messages.

And so my understanding is that in (unintelligible), the (unintelligible) made a clarification to their own statement saying that, “We gave those four security threats that everyone has been focusing on the examples, and that we would like to see spam considered as well.” My personal feeling is that you can do this without spam. If you don't see spam, you probably don't understand the problem is. So were counting spam and were happy to make that information, you know, that data available to you.

(Terri): Thank you. Our next question comes from Maxim Alzoba. Are CCTLD tracking part of the project?

Dave Piscitello: Not currently. I mean, so one of the reasons why we did the GTLDs is because the zone data are publically available, it’s part of the contract. CCTLDs these have their own policies. And so, you know, for those CCTLDs for which we were able to - we could go get a zone, for example, .nl on a .se, we can just put them in if we wanted to. But we didn't want to do that and then upset anyone. What we did was we basically said if you want to get give us your zone, we will included it in their system. It's just a matter of figuring out a way for us to pull the zone on a daily basis. The block lists include all the CCTLD abuse. So, you know, so once we get the zone, we’re in great shape to do registry level statistical analysis. But it's not a part of the project yet.

(Terri): Thank you. Maxim Alzoba had asked if it was possible to publish the URL for the side deck. We have answered that we will publish the URL for this side deck on our GNSO Calendar. I didn't know, Dave, if you had another spot as well that you would like to direct us for where side deck would be posted as well.

.

Dave Piscitello: As I said earlier, (Terri), I will make certain that these slides are posted at the SSR page. And when the FAQ is available, I will contact you and others to make sure that the GNSO and all the rest of the SOs and SCs, you know, can keep following our activity, you know.

(Terri): Thank you. Our next question comes from Maxim Alzoba. Is it possible to see which particular list used for DART? Not very registry is happy with all that. I believe it should be not every registry is all of happy with all of them.

Dave Piscitello: So this - if you go to, I think it's Slide 9. The slide I have up here, "The Current Reputation Data Sets", it's flying past, sorry. Yes, it's Slide 9. These are the data sets that we use and I want to make a comment about people being happy or unhappy with a block list. Ultimately the what I care or what I think about the block lists or what a registry operator thinks about block list or any ICANN community member things about the block list is not relevant to the fact the block list is being used in a security system.

And so I said, the goal here for us is to show the community what, you know, the lens through which others are seeing (unintelligible) for the name space. So if you like Spam House, it's fine that you don't like Spam House and it's my understanding that for Step 11 you don't have to use it. But we use it because over two billion mailboxes are protected by it from different security systems. ICANN's own proof point uses it. You know, my ISP uses it. You know, it's pulled into the feed that Facebook and Twitter and other social media used to eliminate comment spam. So that's what they're using. Being unhappy about it is not really productive for the view that we want to take and we want to share with you.

(Terri): Thank you. Our next question comes from (Mathias Spicer). Is it possible to see threads with TLDs lists for used data sources?

Dave Piscitello: I need an explanation of – explanation here because I don't quite understand the question. So perhaps if (Mathias) can ask it, or ask it again in the chat, or verbally.

(Terri): And I see (Mathaias) is typing if we just want to give him a moment. Oh, and we can close that question per (Mathais). So we'll go ahead and move on.

So, the next question is from James Bladel. All this is all rolled up in a GD – in the GTLD second-level domain? For example, a simple compromise five WordPress.com, or Bloggers.com, would it count as a single instance against that domain and a registrar or registry?

Dave Piscitello: Yes. Well it's – so the answer is that we are counting at the registered domain name, not the fully qualified domain name. So, you know, www is counted, you know, www.example.com and (unintelligible) are rolled up into example.com. So if - the domain level threat is what we count, not the individual URL, not the individual subdomain.

(Terri): Thank you. Our next question comes from Maxim Alzoba. Is it going to be – not much time for answers. Oh, could we expect that ICANN staff answers to questions to the registries and registrars (unintelligible)? And Maxim, whatever questions don't get answered, we certainly will go ahead and get that passed along for you. Moving on to our next question is (Mathias Spicer)...

Dave Piscitello: (Terri), can I pause for a moment?

(Terri): Oh, go ahead. Yes.

David Piscitello: Yes. Maxim, I think that's – I'm fairly certain that some of the questions that you've asked we've anticipated and are in the FAQ. And so in addition to the answers here, I think the FAQ will be, you know, will be somewhat illuminating. And anyone that has a question can contact me. You know, and we can talk about what we're doing.

I know the people on the Internet Help Indicators List, that ITHI lists have been asking some questions about DART because ITHI is actually going to be a consumer of DART data. Some of the metrics that have been discussed for abuse in the ITHI program will use our data. So, you can also post it there,

so there might be more general discussion. There's no limit to the amount of time I'm happy to talk about this. You know, I'm that excited about it.

(Terri): Thank you, Dave. Moving on our next question is from (Mathias Spicer). Why ICANN DART will not publish TLD threat relations data.

Dave Piscitello: So it is not a question of why we will not publish the data, the question of, you know, what would you like to publish? So, you know, so we've got terabytes of data, you know, what kind of report would you like to see? So there's something that, you know, this is going to be something that the community has to bear out, because of some people in the community are going to say we would like to see -- and they'll I use the term -- worst TLD registries or registrars, okay. I don't know if that's useful because there's no, you know, measuring worst isn't really helpful. It's like - I feel much more useful to find a way to show trends that everyone can be aware of because my experience in the new TLD programs since I've been tracking abuse and that's there's a lot of tasting, so to speak.

You know, we have seen criminals move from one registry to another. You know, and we've seen registries go and clean up their act and drop the top 25 list somewhere to out of the picture entirely. So the value of having the data is being able to see you trends, like who is on an up tick and why are they on an uptick? Are you now targeted? Who is on a down - downward slope? Are you on a downward slope because you – the measures that you've just implemented are better or because people walked away? And you know, so we need to know – we need feedback from you on what you want us to report, what you want the reports to look like, how detailed you want them, when would you like them? You know, what (unintelligible) and we'll then work to try and provide you with what you need. Yes, just throwing the data out there will cause chaos and rancor and criticism and that's not the goal here. The goal here is to help people become operationally efficient in mitigating abuse.

(Terri): A Dave, I do see (Mathias) is typing a response to your question. He like to see the TLD threat relations, so threats is not a matter of policy or price, for example.

Dave Piscitello: I'm sorry, could you repeat that? I was – I actually was distracted by (unintelligible).

(Terri): Certainly. I should put that question format, I am so sorry. So (Mathias) would like to see TLD threat relations, and his question - his rebuttal is threats are not a matter of policy and price?

Dave Piscitello: I actually don't know how to answer that question yet. Will you be – if you're going to be in Johannesburg, perhaps we can sit down and talk about that?

(Terri): And unfortunately he will not be in Johannesburg.

Dave Piscitello: (Terri) can give you my email address and you know and/or my phone number and we can talk apply because honestly have to think about what that means.

(Terri): And I'll certainly pass that along to (Mattias). Thank you. Moving on to our next question from Maxim Alzoba. Is it planned to actually consult with registrars and registries which kind of data would be useful for them?

Dave Piscitello: So is not just limited to registers and registries, but community. So as I said, I'm talking to all the SOs in SCs and my sense is that different communities are going to want to see different kind of data. Yes, I imagine that the (unintelligible) is going to want to see certain things, how we would visualize and send data to them. We can't obviously just opened a 24 by 7 by 365 consulting service for the data. But what we want to do is understand how to best apply the data for everybody's purpose.

(Terri): Thank you. And our next question comes from James Bladel. Your description of use of WhoIs data appears to violate our – and then he put in parentheses -- Go Daddy's terms of use. Could you obtain explicit permission from registrars to harvest WhoIs data for this project?

David Piscitello: So I'm not certain that speaking to (Ben Butler) and asking him what we were doing and having him say it was okay qualifies as explicit permission. You know, I mean what we do is -- you know my understanding -- we don't try to violate anything. You know, we are simply using WhoIs data that is collected by our provider. And they acquire WhoIs data from the domain tools. They require the data from registries directly and from the registrars. Yes.

(Terri): Thank you. It does appear James is typing, typing in the Adobe Connect Chat, so we'll just wait for his response before we move on. He was alerting us. "I confused, you said you were working to resolve the problem of rate limits." And just a reminder for everyone, audio is connected as well.

Dave Piscitello: (Unintelligible).

(Terri): More than welcome to speak if needed.

Dave Piscitello: So, James...

(Terri): (Unintelligible).

Dave Piscitello: ... said that SSAC is working on a - they have a working party that is looking into rate limiting.

(Terri): And James, I see your hand is raised. Go ahead.

James Bladel: Thanks, (Terri). This is James for the record, and thanks, David, hopefully you can hear me. So what I'm trying to get at here is you mentioned something about using different systems or methods to, you know, to get

around rate limits. I just want to emphasize rate limits are there as a safeguard against bad guys doing exactly what I think you're doing, which is harvesting bulk amounts of WhoIs. And so I'm just trying to get to the bottom of this here. I think that registrars generally would support this effort. And I think that if DART or other groups would come and say, "Look, if you cooperate with us and can you get us some special access?"

I think it would certainly be open to having those discussions. I think the concern is the discussion that says, yes, registrars put rate limits in our way. We got to figure out, you know, if we got to go out and get some Amazon Cloud servers or whatever to get around them. You know, that's – that starts to venture outside the real of good faith and starts to look like abuse of our systems. So...

Dave Piscitello: So James can – will you be in (unintelligible)?

James Bladel: I'm sorry?

Dave Piscitello: Will you be in Johannesburg?

James Bladel: One hundred percent, yes, sir.

Dave Piscitello: Okay, so let's find a time when you and (Jeff Bedzer) and Greg Aaron and I can sit and talk. And if (Ben) is going to be there, then (Ben) as well. What we have to do - we want to do the right thing and to put the data in front of people. And we're not trying to game the system. We're trying to – one – I mean, as I said the problem that we have is we can't do this project without, you know, without being able to identify sponsoring registrar. And what we had chose to do was acquire the sponsored register in the same way that anyone else in the public does.

The mechanism, you know, is one that, you know, that I'm fairly familiar with, you know, because on every block list operator uses the same thing. And so

if we are in a situation of having to ask forgiveness, then I'll forgiveness and say, "Okay, tell us how we can do this better." And I'm really willing to sit down and work that out with any of the registrars because I think that getting the insight that we can now provide to you would be more than worth the value of white listing or do something else for us.

James Bladel: Okay, let's discuss offline. Thanks, great.

Dave Piscitello: Thank you very much.

(Terri): Dave...

Dave Piscitello: I appreciate it.

(Terri): And, Dave, Graeme Bunton would like to be a part of that conversation as well in Johannesburg. And then confirmed…

Dave Piscitello: Okay.

(Terri): To going, and he would be happy to be a part of that conversation.

Dave Piscitello: Okay, I would - we're going to have to - Terri, afterwards let's just sit down and make certain that I get contact information. I know how to get in touch with James, but I don't know Graeme's contact, so let's just make certain that I don't overlook it.

(Terri): Certainly. As a reminder for anybody who would like to ask a question the audio is enabled, so please raise your hand. You can also type you're question in the Adobe Connect Chat.

Currently at this time we have two minutes left of the webinar. We have one remaining question at this time. It comes from Benny Samuelson. How is this

collecting of data threatened and handled with regards to the upcoming GTRP provision?

Dave Piscitello: The data protection. So the only WhoIs data that our project uses is name server information and sponsoring registrar. So I'm - I believe that that's not personal data. There are probably people who -- somewhere in the world -- that have a difference of opinion. We're using the data that are currently public in a manner that everyone is using the data that are currently public.

If the GDRP changes to WhoIs that all the registrars and registries comply with it, we are going to have to adjust how we do this. And so at that point - we're not going to break laws. At that point were going to do what necessary to do to comply with the laws and try to continue to provide data for the project, however that can be done.

(Terri): Thank you. And currently at this time there is no further – actually I do apologize, Benny had a follow-up to that. Okay, so no personal data collected?

Dave Piscitello: For our project - so the project - the project platform is a shared platform. Part of the system, you know, part of the system collects WhoIs data, the entire WhoIs. What we pull in from that data to our data are – is the response and registrar information. I would have to, you know, sit down and talk with the implementers or the developers to understand, you know, what compartmentalization, you know, what the compartmentalization actually looks like in terms of where the data lie. But our project doesn't need personal data.

(Terri): Thank you. And then he concurs that answers this question. Currently at this time of one minute left of scheduled duration, I see no further questions at this time. Dave, do you have any closing comments?

Dave Piscitello: You know, I'm just – I'm delighted that so many people showed up and I think the questions show that there's – there's some interest in, you know, in utilizing the service. I would just encourage people if you're going to be in Johannesburg and you want to see what we do, I'll be happy to show you your data that we see. And if you're a registry or registrar I will be able to show some generic data to different people on different sides.

You know, and I don't want to put this up on a public screen because quite honestly that's the way things get snapshotted and cause a lot of churn. But, you know, in the privacy of my laptop and over a beer I will be able to navigate through the admins just so you can see what we do. Because honestly if you saw some of the the graphing and charting that we currently have, you'd be really impressed. And if there's a way that we can produce this for people, I think that it would be very, very helpful.

(Terri): Well, Dave, we'd like to thank you for taking time to present to us today on the DART tool. If anyone has any further questions or comments, please reach out to the (unintelligible) at ICANN.org, and we can certainly pass those along Dave's information other than the folks we've already – were going to pass information among those mentioned throughout this webinar. So this has concluded. Thank you very much for joining. Operator (Seun) you can please stop all recording to everyone else. Please remember to disconnect all remaining lines and have a lovely rest of your day.

END