

---

UNIDENTIFIED FEMALE: Hello, everyone, and welcome to the SSR2 call on DAAR, taking place on the 24<sup>th</sup> of June 2020 at 15:00 UTC. And the review team members attending the call today are Laurin, KC, Ram Krishna, Kaveh, and Zarko. Apologies from Kerry-Ann, Eric, Boban, Denise, and Russ. And attending from ICANN Org is Jennifer, Steve, Brenda, Gus, John Crain, and Samaneh. And we have technical writer, Heather Flanagan, on the call.

This meeting is being recorded. Please state your name before speaking for the record. Thanks very much, and I'll turn the call over to our guest, John and Samaneh.

JOHN CRAIN: All right. The call was, of course, requested by the SSR2, so I will leave it mainly to you to ask questions. I think most of you, maybe [inaudible] of you, have met Dr. Tajalizadehkhoob before. So, why don't I let Samaneh introduce herself first, so that everybody gets to know her? She's the principle staff member working on the DAAR system at the moment. Samaneh, why don't you introduce yourself to everybody?

SAMANEH TAJALIZADEHKHOOB: Thank you, John. Hi, everybody. Thanks for having us in the meeting. I don't know if I have met you, some of you, before in ICANN meetings. I'm Samaneh. I lead the DAAR project at the moment. I joined John's team about a year-and-a-half ago. I have academic background in Internet measurement, lead security, and statistical modeling analysis.

---

*Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.*

---

I think, like John said, I will let you guys lead the session and we are more than happy to provide information on DAAR and any other questions. Thanks.

STEVE:

This is Steve. I'm going to jump in just a second, too. I know that this was mostly called because of KC, you had some questions about DAAR and stuff. And I think, Laurin, you're proceeding this meeting. I don't know how you guys want to frame it, so we'll kind of just leave it organic and see how you guys want to do it, or if you have an agenda in mind, we'll take it from there.

UNIDENTIFIED MALE:

Yes. Thank you, all, for coming and taking time out of the week that contains an ICANN meeting as well. That's definitely appreciated. I think, from where I stand, it would make sense if we start with KC's question because this is what kind of created the need for this call and then we can kind of open up to more questions. There will probably be some follow-up questions as well.

KC, you would probably do a better job at asking your questions than I do. If you're happy with that, I'll hand over to you to ask questions. Otherwise, I'll have to try to frame your questions in my words somehow and maybe get it wrong.

UNIDENTIFIED FEMALE:

KC, we can't hear you if you're speaking. We still can't hear you, KC.

LAURIN WEISSINGER: It seems there is no contact with KC right now.

UNIDENTIFIED MALE: Yeah. Laurin, just one second. KC, you're muted. Laurin, in the meantime, why don't you try to frame her questions while she works out. We had her just 40 minutes ago and it was working, so maybe she can work out the details in the meantime.

LAURIN WEISSINGER: Yes, exactly. This was my plan. Essentially, our first question—or KC's first question—relates to slide 38 of the DNS abuse malicious registrations during COVID-19 final. I'm just pulling it up on my own screen.

KC CLAFFY: Hello?

LAURIN WEISSINGER: Oh, hi, KC. We can hear you now.

KC CLAFFY: Wow. That was quite the Zoom failure mode I had. Sorry about that. All right. I didn't want to interrupt, but Laurin asked me to kind of start out here, which I'm actually being driven by the comments to the SSR2 draft report, various folks who on the DAAR-related recommendations are

---

asking for more detail on what exactly we wanted out of DAAR, which leads to the presentation by Dave Conrad on Monday night where he had a slide about what OCTO's plans were for DAAR but it was super vague. It was super high level.

So, the first thing I would like to do is just go through that slide. I don't know if anybody has it handy, if you bring it up. It's on the website. I pointed Laurin at it. Let's scroll down a little bit on what are the proposed changes that OCTO is planning to make to DAAR and what is the timeline for the changes. Let me see if I can find it. Did folks hear that? I want to make sure my audio is working.

UNIDENTIFIED FEMALE: Yeah, we heard you, KC. I can try and find the slide, while I don't know if John and Samaneh, if you want to get started and I'll see if I can take the slide out.

SAMANEH TAJALIZADEHKHOOB: I'm not sure if I have seen the slide, but regardless, I can just say what I know and then we can refer back to the slide when it's on.

KC CLAFFY: Can I share a screen? Oh no, I can't share a screen. If I can share a screen, I can show the slide, but I'll let someone else ...

UNIDENTIFIED FEMALE: KC, I'll give you cohost rights for you to share your screen.

KC CLAFFY: That's a great idea. [inaudible].

UNIDENTIFIED MALE: Conversely, KC, if it's a link, if you want to just post that in the chat and then we can grab it from there and then we can [inaudible] what you're looking at.

KC CLAFFY: I can't tell. Can folks see that? Hello, hello?

LAURIN WEISSINGER: Yes, KC, we can see it.

KC CLAFFY: I can't hear anything.

LAURIN WEISSINGER: Okay. KC, we can see the slide. Can you hear us, still? We can hear you. Okay, KC, we can't hear you. Okay, we seem to have connectivity issues on KC's end. She's just telling me out of Zoom that she has trouble hearing us. She's going to dial in from her phone.

And sorry about this being hard to read. I'm having the same issue. I will try to share in addition. Give me one second what this is. Technical issues.

UNIDENTIFIED MALE: There you go.

LAURIN WEISSINGER: Ah, [inaudible]. Good.

UNIDENTIFIED MALE: So, while KC dials in on the phone, I can give you a quick overview of—

UNIDENTIFIED MALE: Sorry, can we mute KC's Zoom session? Thank you.

LAURIN WEISSINGER: Essentially, what our questions are about is essentially what are the actual plans? Because this is extremely high level and we would just be interested where we're going, what other kind of milestones, the kind of choices, etc. And we [have] no choices yet, but directions, so we have a better understanding of where this is all going in a bit more detail, and so that we can have an idea of what's going on so that we can change our report and adapt things accordingly. So, essentially, we just want a bit more detail on all the points that are related on this slide. I hope KC is able to join us via phone soon.

KC CLAFFY: Can you hear me now? I'm on the phone. Do you hear me?

---

LAURIN WEISSINGER:           Awesome! We can hear you, KC.

KC CLAFFY:                    Yes. I'm on the phone. Gave up on the laptop. Right. We can just either start [inaudible] slide or you guys can just start from wherever you want and talk about plans and essentially—

JOHN CRAIN:                 Yeah. The slide is going in and out, so why don't I do what David did—because David is even a higher level than me—and go one level down. I think [inaudible]. I suspect David had this at the last minute. There isn't much about DAAR on this slide, as you can see.

So, the concept that we have in DAAR—and we're not there yet. I'll start by saying that. This is going to be probably an ongoing path. And I think this is a path we will take with any tools we develop, so it's not going to be unique to DAAR. I know everybody likes to focus on DAAR because it's our most visible tool. But it's more of a principle of we want to work with the community to make the tool better for whatever the community needs and that's something that really has to have some definition from the community and not just as ... Obviously, we have to give input into what is actually possible because some people wanted to paint the sky pink and some people wanted to paint it green. And obviously there are limits to what we can physically do, even in software.

So, what we initially did was we started a mailing list to start discussion. There's not been much discussion on there. We've also been in

---

discussion with the contracted parties, specifically the registries, along some of the questions they had around the data and some of the changes they want. And we've been having discussions, which go very much in line with the discussions with the registries, I must say, at the various ICANN meetings where we've had physical meetings.

So, our long term—and I don't know whether long term is six months or a year or 18 months because we're in the planning stages. And to be honest, Samaneh is doing more of the planning than I am, so I'll let her talk if she has more specific data.

In the long term, our idea is to re-specify the project, if you like, based on community input, and then most likely—because as I said, we're in the early stages—we will put it out for bid or we will rebuild it ourselves to redevelop it in whichever way the community really wants us to. And we've had some discussions. Samaneh, why don't you talk about some of the things that we know have been requested by multiple people? And then maybe also talk about how people can interact with us to try and help us improve it.

SAMANEH TAJALIZADEHKHOOB: Yeah. I can take it from here. So, in general, we've been in discussions with the community about the current limitations of DAAR. I will just point out to some of the most discussed things which were, for instance, the limitations that are imposed by the reputation feeds. The fact that the report at the moment, or the system at the moment, reports aggregations for one day per month. So, the DAAR report—the monthly report—is for the last day of the month.



The fact that, in some parts of the DAAR report—not the system—the abuse types are collapsed all with each other and presented as abuse. One thing as a whole which was criticized because the majority of the data, of the abuse feeds that are currently going into the system are spam feeds.

Also, there were also suggestions on other types of analytics that the community—specifically registries—found more useful to have. Other types of categorization, other latent variables to look at, etc.

What we are going to, at least for now, what we are planning to do, part of it is already in progress, so we are changing the report. Things that are critical to do, like changing the language, clarifying what the DAAR project aims to do and what it does not do.

For instance, it was also discussed in today's session with GAC that DAAR has not been intended to provide actionable data but more to be as a starting point for contracted parties to see where abuse is concentrated and where are the most critical areas that need attention.

One of the other things that we are looking at, and it needs discussion from the community, as John already pointed out, is we are being very selective on the reputation feeds that is going to the DAAR system. They are the most repeatable feeds that are in the industry. But of course there are also problems with them. We are developing analysis to quantify that.

One of the solutions that we thought of solving that issue was, for the second version of the project, which will come later in maybe—discussion with the community of externalizing that point of the project.

---

---

For instance, we develop a system that creates reputation and the input and output would be up to the user of the system. So we do not get involved in what goes in the system and what comes out.

These are at the level of our own internal ideas. It has not been discussed with the community. We don't have input on it. But it will be in the future.

On top of that, something that always gets mixed when it comes to abuse discussion is—John also pointed this out—DAAR is just a tool. We have been and will be doing one-time analysis of abuse, of trends, of prediction analysis, of other variables that can predict abuse, stuff like that. Of up times. This does not necessarily need to be inside the DAAR system as it's just a system and it has its own limitation. For instance, it still has not managed to include registrars in the DAAR system due to [WHOIS collection] [inaudible]. Stuff like that. We've discussed several times in the ICANN meetings.

JOHN CRAIN:

And some of those things, I didn't hear Samaneh specifically state it, but for example the point in time versus averages, that is something we're going to ... Now, that was a good input we got from the community. We're going to implement that. And there are a few of those items that community members brought up in the DAAR discussion sessions that we fully agreed with and are working on making sure we can implement those.

---

UNIDENTIFIED FEMALE: Sorry, John, what were those comments?

JOHN CRAIN: One was the point in time. We're actually also looking at the discussion about data [pass through]. That is more of a legal discussion than a technical discussion, i.e. we have to negotiate with the feeds. Whatever ones were there, Samaneh, that came up that we decided to do. There were some wording issues, I know.

SAMANEH TAJALIZADEHKHOOB: Yeah. There was some clarification on the language. There was suggestions from the community on including other and looking at other relations ... Relations of other variables with abuse. At the moment, as you can see in this example, also we are categorizing registries only based on [inaudible] which is whether they are legacy TLDs or new TLDs.

We looked at other factors. At the moment, our discussion is on embargo, so I cannot go too much into detail of that, but after several sessions of statistical analysis and discussions, we concluded that for now this is the most meaningful and useful information to include. In the report, there will be some new analytics, as in new plots and new statistics in the report.

JOHN CRAIN: The other major one was the no longer clumping of all the abuse types together. So that's another one where we're planning to actually separate those out as quite a few people in the community asked us.

SAMANEH TAJALIZADEHKHOOB: Okay. And someone said about [you] decided this is the most meaningful stat to include. What stat are you talking about?

UNIDENTIFIED FEMALE: The categorization of looking actually at the level of analytics. So, visuals only. The legacy versus new [per] gTLD.

SAMANEH TAJALIZADEHKHOOB: Instead of? What is it that was the alternative? Do you mean dividing it per TLD?

UNIDENTIFIED FEMALE: No, the alternative variables that we looked at were, for instance, TLDs that are aligned with spec 13, spec 11, and some other characteristics of TLDs. Are they brand TLDs? Things like that. There was also some other internal data that was provided by contracted parties, but like I said, there was not much info there and the discussion is still [on embargo].

SAMANEH TAJALIZADEHKHOOB: [inaudible] I anticipate it will be non-embargo?

UNIDENTIFIED FEMALE: We are working with them on a draft recommendation that they provided based on the discussions we had to publish. At the moment, they are revising their recommendations, some of them [inaudible], and

---

---

then it will go through ICANN Org and then ... I think in a month or next two months it will be published.

JOHN CRAIN:

Yeah. This is really with the registries and we've told them quite clearly that if they want us to work based on their recommendations that they have to publish those. They have to be discussable by the community. And to be honest, in a very amiable discussion and they've been very much focused on actually really trying to improve the system. But I think there's more discussion to be had with the wider community.

And also I guess, asking a favor here, KC, if when they publish that, as somebody who has worked in this field for a long time, if you could also have a look at it and see if they make sense to you.

KC CLAFFY:

Right. There's nothing to look at from any of the other [folders] now, right?

JOHN CRAIN:

No, unfortunately not. This is a question we've raised with the ... It's called the Registry Stakeholder DAAR Working Group and we've raised this issue with them multiple times, the need to get this out there so that we can discuss it.

I personally don't think there's anything at all controversial in there but I do suspect that other people may want us to focus on other things as well.

KC CLAFFY: Right. I'm definitely still struggle with the high level of the answers here. I think it's not clear to me when Samaeh says changing the report, changing the language, being more clear about what DAAR does and what it doesn't do, what you mean by those things, beyond the example that she gave about actionable data and DAAR never has been apparently, despite what others have said, intended to provide actionable data. So can you ... Yeah, go ahead.

SAMANEH TAJALIZADEHKHOOB: Yeah. Maybe I sort of assumed that some of these discussions have been discussed in every part of the community. I will be more clear now.

So, there is this confusion that, for instance, the reputation feeds that are going inside DAAR are by ICANN Org. There has been a lot of objections on why do we input such data and this and this list, but the outcome is also basically the reputation score that is the output of the system. It's also created by ICANN Org which is somehow apparently still not clear for part of the community.

So, repeatedly in the sessions that I presented, I clarify that these are just third-party feeds, that we try to be selective on choosing them and they go into a system that the calculation is easy. We collapse them with domains that are in the zone files and then we ... Of course, there are also a lot of corner cases but we created kind of aggregated score, [inaudible] gTLD which is a count of the number of times that a domain [inaudible] is appeared in each of the security threat types.

---

By clearing the language, this language is used in the description of the system, so the link that was sent earlier to you of the methodology paper but also in the monthly reports, also in the presentations.

So, part of the community, the Registry Stakeholder Group asked us to clarify this language and clarify that, for instance, DAAR is not measuring registry actions, mitigation actions, which by their definition we could call it up times, the amount of time that they take for registries to take down a domain because we cannot see that. We only see concentration. We don't see up times.

And stuff like DAAR is never intended to provide actionable data just as a starting point, to start discussions actually with registries and registrars to get them, to look at abuse, to familiarize them with how to measure abuse, and us to help them to develop systems like that.

I hope that was more clear. And if it wasn't, then please ask questions. I will point to specific answers.

KC CLAFFY:

Yeah. We're making progress. Thanks. When you say selective on the reputation piece, what do you mean by that? How do you measure selectivity and how do you validate that a reputation fee should be in the system?

SAMANEH TAJALIZADEHKHOOB: Good question, KC. Initially, like I said, I joined the DAAR project a year and a half ago. Initially, I think the selection was just based on the general reputation of the [inaudible] provider, let's say, in the

---

community. The ones that were used in scientific work and where used by the security firms for blacklisting, whitelisting. I think I summarized all the criteria that was used previously.

At the moment, it has been on pause because of other COVID-related stuff that came out. We were looking at creating metrics and this has been also looked in academia briefly at creating metrics—basically, metrics to be able to evaluate [fees]. Some of these metrics already exist and some of them we are just working on to create them.

Once we have that, then it will be we can quantify evaluation. As of now, we don't have a quantitative metric to say, to evaluate—just more qualitative, based on reputation.

KC CLAFFY: Sorry. And reputation, you get from talking to people, basically, or ...?

SAMANEH TAJALIZADEHKHOOB: Maybe John can comment on that better.

JOHN CRAIN: Samaneh wasn't here at the time we did that back in the day. So, what we did is we looked at white papers and we looked at well-known software vendors, at what they were putting into their blocking systems, if you like--the firewalls, etc., like Cisco and folks like that. Much of this is actually documented in the methodology document.



---

KC CLAFFY: Right. I'm trying to get a handle on because someone said you're looking at creating metrics by which to make this a little more rigorous. What metrics, for example, would you consider using?

And I'm also wondering ... I mean, there are plenty of public—or public-ish, like Spamhaus [inaudible] share with academics no problem—that you could use as a source in cross-validation but there doesn't seem to have any effort to compare the existing commercial feeds that you're getting with what third parties could get and reproduce the work themselves. I'm just wondering why.

SAMANEH TAJALIZADEHKHOOB: On the first ... I didn't quite get the second part of your question, KC. But on the first part, indeed there are some academic works on creating metrics—and by metrics, I mean, for instance, the overlap ... I don't know anymore. How did I call it? But how early the domains appear on the [feeds] in comparison to each other and how comprehensive each [feed] is, from the point that we can see, of course. We don't have the overall view because no [feed] is complete. And the amount of time that a domain stays on a list in comparison to another.

I don't have all the metrics in the back of my mind but these are examples of things that we looked at.

KC CLAFFY: Right. Unfortunately, the academic work I'm familiar with doesn't have a happy ending there. So, it's not clear that's going to be useful.

---

Okay. The second one, the second question is, because of the situation where the raw data and the methods are proprietary, what about a recommendation of or an effort to use more public, that is, clear path to get the data by third parties? And especially academics who may not have funding to go pay commercial feeds. But a lot of the commercial providers are willing to share with academics under specific conditions, including perhaps [inaudible] work.

So, what about having a DAAR where you only use reputation [feeds] that are agreeing to such terms? In the interest of reproducibility, you're only going to use it if there's a clear a path for other academics. It doesn't have to be academics, but for others to get access to the data, specifically for the purposes of reproducing the work.

SAMANEH TAJALIZADEHKHOOB: Yeah. That is a very good suggestion, indeed. Actually, most of the [feeds] that are currently going into DAAR are free, for instance, for research purposes. May not be free for contracted parties. As far as I know, there is only one fee that, at the moment, was paid. But for ICANN, of course, we have to pay for a lot of them.

But we can take that recommendation into account when we go to version two which we plan to put for [inaudible]. I can order it internally. we can, of course, take that and I think it's a good recommendation.

---

JOHN CRAIN: KC, are you thinking of a recommendation from the SSR2 itself or are you talking more generically?

KC CLAFFY: From SSR2.

JOHN CRAIN: Okay. Yeah. I would wonder the balance of what limitations that gives, but it sounds like—and Samaneh is much more of an expert in this area than I am. It sounds like she thinks that the majority of them already do this anyway. So, maybe it's a real easy one, if that's the recommendation.

KC CLAFFY: Well, make no mistake. I'm definitely proposing to modify the architecture, if necessary, in order to make it be the case, in order to make it be reproduceable work, because what you're hearing—

JOHN CRAIN: Well, that was one of the original goals. So, if there is something in there that is hampering that goal ... I fought pretty strongly with some contracted parties who said we should use ICANN's internal data to identify registrars, and I pointed out that if we can't release that data, irrespective of all the issues we have internally of getting to data and what contracts allow, etc., if it was only allowed for ICANN to do that data, how are people going to reproduce it?

KC CLAFFY: Well, since ICANN could do the same kind of thing that the commercial [feeds] do.

JOHN CRAIN: Yeah. Except it's not our data. From an engineering perspective, I'm 100% behind you. The day which we could release every single piece of data we had ... But I'm not the lawyer.

SAMANEH TAJALIZADEHKHOOB: I understand your concern, KC, and I think it's a valid concern, especially in academia. But another point I would like to raise is that the methodology is public. Unless people are interested in recreating the exact number, that at the moment, only registries are getting, so each registry only sees their own score via ICANN API. At both academic level, also any kind of research, any kind of projects. In theory, the project can be ... It's replicable by other sources that are ... By even the data feeds that are—some of it, like Spamhaus. Most of them actually are free for research.

KC CLAFFY: Right. Although, I guess we have to be careful, because even those feeds, the methods of the feeds aren't available and that's where the magic is, right? Just copying what the commercial feeds say is not the sauce that is going to show well. Is it [valid against] [inaudible] [truth]?

---

SAMANEH TAJALIZADEHKHOOB: Indeed. That is also one of the problems that, at the moment, we have, is that we try to normalize the [feds] before we aggregate them in the system, but anyways, the methodology of each [feed] provider is different.

That was also one of the ideas behind the next version is that we only create the system. We let the input be independent. But that still needs to be discussed more.

KC CLAFFY: So, really, the question is you mentioned a couple of times the limits of the [fee] providers on what you can or cannot publish. Can you be explicit about what those limits are? Because I'm also hearing contradicting views on what the [feeds] are not allowing versus what ICANN is not allowing or other contracts that ICANN has, so what is it ... And maybe it's not the same for each [feed], but is it something you can publish what the constraints are of each or all of the [feeds] that you're using?

SAMANEH TAJALIZADEHKHOOB: I think we can publish that. I think John is a better person to respond to that question. Yeah?

JOHN CRAIN: Yeah. I don't know off the top of my head because it's probably a legal thing, but I think that is a completely reasonable recommendation to make, that we publish those limitations. But the limitations, I think most of them—and I don't have all the contracts in front of me—only allow

---

us ... They often use terms like derivative data, etc., and they don't allow us to pass through. And I understand completely why, from a commercial perspective, on their standpoint.

But I think a recommendation that asks us to be more transparent about the limitations is completely reasonable. Then of course we would have to go and look at what we can and cannot do in the contracts. But maybe, just like the one about sharing, with academia or access to academia—just like that recommendation, maybe if ... And I'm shooting off the cuff here a little bit. Maybe the ability to be transparent about what we can and cannot share per contract is a limitation that we place on the feeds. If they don't let us say what we can't share, then we don't use them or something like that.

KC CLAFFY: Right.

JOHN CRAIN: So, I think that's a reasonable recommendation for SSR2 to [send]. If you want to get down to that level of specifics on what should be the, I guess, criteria of the [feeds] ... Yeah.

KC CLAFFY: Right. So, let's go back a little to the [inaudible] stuff. Sorry if I'm monopolizing. Other people should jump in if they have questions. I'm trying to keep an eye on the chat. And this will go beyond DAAR, which you suggested we do at the beginning anyway.

---

When you say never intended to be actionable, my understanding was compliance is also making use of this data as well. Was there ever any intent to do anything that was actionable? I'm still wondering why so much effort is being put into something that seems ... Or maybe it's not that much effort. You can tell me. But it seems like it's pointing to a problem that people are pretty well aware exists. So, what are your thoughts on doing something that is a bit more actionable?

JOHN CRAIN: Do you want me to take that, Samaneh, or do you want to take it?

SAMANEH TAJALIZADEHKHOOB: I will answer the second part of KC's question. You go for the first part, John.

JOHN CRAIN: So, yes, it's sort of well known where these problems were and that these problems were. But when we started DAAR, there was a lot of discussion in the community. I'm not going to call it denial, but there were people saying that a) this wasn't exactly how it was and b) you couldn't measure it. So, we wanted to [throw] data at it and say if you measure it this way, this is what you actually see, so you can actually measure this. And then to give that data as input into the policy process.

And that was the original thought behind DAAR. Can we actually measure this in a consistent way and publish it?

---

The original work was quite a bit of work, but what's happened over time is people have taken their own perception of what DAAR should be rather than what DAAR is and are regularly trying to push it that way.

What I prefer is rather than trying to make DAAR what people want it to be is to figure out what people want and then decide where it's appropriate or us to build that.

Compliance has looked at some of the data, but not purely data out of DAAR. They have asked us, for example, to do other reports that go beyond what DAAR does to actually go and look more in depth at names. If you're familiar with this, with what we're calling the sticker project, there we actually go and look for evidentiary data of reports of actual harms but by well-known vendors, I guess, or corridors. But DAAR itself at all doesn't have that kind of data.

So, if the industry wanted something more actionable or the community wanted something more actionable, I think that's a reasonable discussion to have. I don't know what the effort is. I don't know what that looks like because we've not had that discussion yet. I suspect with will get people who want us to and people who don't. Samaneh, what do you think on that?

SAMANEH TAJALIZADEHKHOOB: I think part of the struggle, at least ever since I am working on this project, is that okay now that we are at this stage? Should we publish or not publish the underlying names?



---

Over time, we see more requests from the community for names to get published, but we still did not have that discussion in which we conclude that the community wanted ICANN Org to go to that road.

In the meantime, like John pointed out, we did other smaller projects in which we did more statistical analysis because, for DAAR, we just do high-level analytics. We don't take into relations and regressions.

For compliance purposes, we did that one time and we're planning to repeat that. And for other future projects, we are planning to do that.

And maybe with DAAR, it goes there or not, but at the moment, this is the state that we are. Like John said, there are other projects that address what you ask.

JOHN CRAIN:

And yeah, when we talk about the difficulty of this, one thing I will say is that we did struggle on resources for a while. We had a couple of people leave. Samaneh had joined us about 18 months ago. But I've actually added two more folks to my team, specifically to do ... Not DAAR because DAAR is just a tool. But to work on these kind of issues.

We have a gentleman called Dr. Carlos Ganan who came from the University of Delft. I sometimes get that wrong, but yes, University of Delft. And Dr. Sion Lloyd who comes from Nominet and has been there for probably two decades doing research. So we have more resources within OCTO beyond our normal research group that Matt runs. I mean, we've basically tripled our research capacity this year. It's easy to triple when you don't have much but we have tripled it.

---

I guess I'm saying we're open to suggestions, not just from SSR2 but just in general from the community about the kind of things we could be or should be looking [at].

KC CLAFFY: Right. That's great. When you say not DAAR but just work on these techniques, you mean your own internal inferences of abuse of activity or something more related to CCDS or do you have any more concrete plans?

JOHN CRAIN: So, the biggest thing that we have on our plate at the moment ... I guess you may have noticed that we had a slight change of how everybody lives recently with this pandemic. So, the [sticker] project was quite a large project resource-wise and quite a good learning. I mean, I'm not a researcher. I'm just an engineer. I'm not that smart with all the numbers and stuff. But I certainly learned a lot watching my guys work on that. The question of where does that go ...

KC CLAFFY: You mentioned [sticker] project. I'm not sure what project you're talking about.

JOHN CRAIN: [Sticker] is our COVID-19 project where we take data, break down the names based on strings within second-level domains, and then do research to figure out whether or not there is any actionable data we

---

can gather on those names. Are they seen as abusive and can we actually gather evidence around that? And then to report that to the parties that can take action, which is basically the registrars. So, that was quite a big project in the last few months.

UNIDENTIFIED MALE: I'm just going to interrupt for a second. KC, that is in relation to a blog that Goran posted a while ago that Denise sent questions to and the review team acknowledged that those were questions that wanted to be answered. So, in the mail list, there's a list of questions and answers that have been previously addressed on that.

KC CLAFFY: Oh, okay. I remember the mail. I didn't know it was called [sticker]. Okay.

JOHN CRAIN: It wasn't at the time. We didn't have a name for it. That's how much of a "we need to do this" it was. [Sticker] is actually a fairly recent name. So, it's not your fault for not knowing that it was called that. Sorry.

KC CLAFFY: Okay, no problem. So, what is the goal of that project?

---

JOHN CRAIN: So, the goal of that project was to put actionable data into the hands of the registrars and registries, so that they could take every suitable action against abusing names. That's all in the blog that was posted.

KC CLAFFY: Yeah. That's one of the things that was talked about on Monday night because there seemed to be quite a disparity between what some registries considered levels of abuse and what the aggregate data showed. [inaudible] as any resolution possible without additional granularity.

JOHN CRAIN: Yeah. So, we were getting down to that granularity. We weren't sending them anything that didn't have a fair amount of evidence, including things like screenshots, etc.

The kinds of things that people have been telling us [inaudible] and we based our reports on the registrar document of what they wanted to have reported in the case of an abusive name. So we took their work and the work of the registries, too, and applied that.

And the numbers that we actually report are quite low, because when you actually go down to looking or the ... There's two reasons for that. One is the ... Well, actually, we don't know the answer, whether it's the registries or registrars or something else. But the names that are related to COVID-19 and COVID-19 related abuse are disappearing faster than we can identify them a lot of times, which I guess is good.

---

And the other thing is once you start actually digging for data, a lot of this stuff ... This isn't really abusive in the first place. There's a lot of park names and names that aren't resolving. They may be names that have risk associated with them but they don't necessarily have evidence of abuse. But we learned a lot from doing that.

Some of those methodologies and tools we still need to document, will probably live on. The only question is: how? And that's really, once again, more of a community discussion.

KC CLAFFY: Right. [inaudible] a little bit watching the time here—about CCDS as a platform. I continue to see complaints about access to that data.

JOHN CRAIN: We put in complaints.

KC CLAFFY: Sorry?

JOHN CRAIN: Yes. We put in complaints, too, when we can't get to it.

KC CLAFFY: Right. So, what is the plan for improving that accessibility of that?

---

JOHN CRAIN:

I'm really the wrong person to talk to that. I know there are plans, but we don't run the operations on that. There were some issues recently. I think they were truncating files. I know there's discussions internally about that.

There is some recommendations already, I believe, possibly from SSAC. Correct me if I'm wrong, Steve. You might not know either. And I know they're being worked on. I think that's a good question to, I don't know, I think Ashwin. I know they're working on it. I just don't know the details, so I don't want to get into them because it's not part of my day-to-day job. But we do rely on CCDS, so if there is an issue, we see it and we also put in complaints and say, "Hey, guys, have you seen this, have you seen that?"

One of the purposes that was not really a stated purpose but is kind of a [inaudible] work is we want to use—not just because of repeatability but also because it teaches us things. We want to use the same tools everybody else is using. So that's why we use CDDS. We have the possibility of access to date those files in other ways at ICANN. But for every time there's a glitch in CCDS, we feel it. Please, if you have recommendations there, go for it.

KC CLAFFY:

Well, I don't know the constraints that are preventing that system from operating more fluidly. [Norm] is actually [inaudible].

---

JOHN CRAIN: I mean, one of the problems I find, being very frank here, KC, that as an engineer, I often sit there when I see these things and say, “Well, I can just fix that.” Because I see engineering solutions to all of these things. But my hand is not on the burner or on the machines there, so trying to figure it out is often interesting.

KC CLAFFY: I see.

JOHN CRAIN: If we ran the world, KC, it would be a better place.

KC CLAFFY: Yeah. Okay. Just the last couple of questions here. If you were—and you are. To the extent that you guys have your plans on DAAR, I’m still not clear on what do you think would be the most important changes to make that would improve its utility to the community? And that’s multiple segments of the community, not just registrars and registries. Is it the splitting up by ... I mean, do you have your own opinion or are you just taking in public comments from others? Is it the splitting up the categories of abuse?

JOHN CRAIN: Well, I think we do have our own opinions but I think the community opinions are probably more important than ours and I think many of them have been voiced.

KC CLAFFY: Have your opinions been voiced? Maybe in the recent report but that was—

JOHN CRAIN: When we have the discussions, I allow my staff to be frank. I think everybody believes that the data would be more useful if you could identify who was behind each string. There were discussions ongoing for a while. I think multiple people in the community have said that.

Personally—and this is just me—I would like a more dynamic interface where people could choose what they wanted to see themselves. In the end, it's just a database, right?

KC CLAFFY: And what [blocks that]?

JOHN CRAIN: We've not decided where we're going yet. The next interface may be more dynamic. I don't think it's [blocked].

KC CLAFFY: It's not on legal or contractual still?

JOHN CRAIN: No, no. We don't have DAAR version 2 yet. At the moment, the output from DAAR—what you see and what the community sees is mainly a

---



---

monthly report, and for the registries, the contracted parties whose data it clearly it, they get to see their daily reports. But I would ... I mean, in my ideal world, we'd have a dynamic interface where people could query the data, right?

KC CLAFFY: Right. Again, what people want to [inaudible] for isn't what you said before which is something that allows them to go to the next step, to reach out to their involved parties and [inaudible] unlikely to be in version 2.

JOHN CRAIN: I don't know where we are on version 2 on that to know if ... I think the licensing issues are what are holding us back on some of what people want and we're working on those. For example, I don't see a scenario where we can just pass through live data. That's just not where the other side of those contracts is seems to be willing to go. I mean, obviously, there probably would be a number you could throw at that, but I doubt the community would be wanting to spend whatever large sums they'd want.

LAURIN WEISSINGER: I just see we have three hands up—Steve, Zarko, and Samaneh.

KC CLAFFY: Oh. Oh, sorry.

LAURIN WEISSINGER: So, maybe because we're nearly on top of the hour—

KC CLAFFY: Thank you.

LAURIN WEISSINGER: We should let them speak as well.

STEVE: I'm going to defer to Zarko and Samaneh first and I'll keep a hand raised for last at the end, but I want them to be able to address first.

SAMANEH TAJALIZADEHKHOOB: Thanks, Steve. I actually also have a question for KC, which I think would be useful input to ask on our people to use your capacity and expertise is that I saw in the email conversation somewhere that you asked about what are our plans to measure the effectiveness start.

I think, research-wise, it's a really interesting question. I have thought about it and I concluded that, because we are only providing aggregation on all kinds of abuse, it is hard to—and we are providing it to registries—it's hard to separate the other factors that can influence registries actions. So, [concentration] of abuse might not be the best metric to measure that and other factors that we don't see, like staff, the [cost] they put on abuse, other stuff.

---

But I think one thing that would be interesting for us and for the project—and I of course have thought of, for instance, effectiveness of the project would be measured if it would have measured up-time. That would be more of an action that is taken by the provider. But would be useful for us if we had your thoughts on that, if you have recommendations on ways you think we could measure effectiveness.

JOHN CRAIN:

Yeah. I think that's a big question for maybe a later discussion.

KC CLAFFY:

Yeah. The problem is the effectiveness metric depends on what problem you're trying to solve and I don't think there's consensus on what problem DAAR is trying to solve. At least from the public comments [inaudible] SSR2 report, there's quite some discrepancy on what DAAR should or could be. Let's let Zarko talk now. I'm happy to talk about that more offline with someone. Zarko had a question. Then Steve.

ZARKO KECIC:

Yes. We're talking about [half of a] domain space [all the time] and I would like to ask Samaneh and John what is the point of us [inaudible] and what benefit this will have joining DAAR? Because there is no actionable items over there and there are a lot of persons in this which are reluctant to give data and [inaudible].

---

SAMANEH TAJALIZADEHKHOOB: I couldn't hear the question clearly but you were talking about ccTLDs, right, Zarko?

ZARKO KECIC: Yes.

JOHN CRAIN: I heard the question if you want me to take it.

SAMANEH TAJALIZADEHKHOOB: Yeah, go for it, John.

JOHN CRAIN: So, the TLD operator has more data than we have. Even though the metrics we give them on a daily basis are not actionable on a per-name basis, they have more data around their registration processes than we have. So, in some ways, it is not truly actionable but it's definitely indicative to them. I know of at least one case of a gTLD before they had access to this data where I had a conversation with them one day about the levels of abuse they were seeing, and then I talked to them the next day and there would have been a spike in the numbers. And because they knew there had been a spike in the reports of abuse in their TLD, they could go back and look at the registration and they realized, in this case, it was related to a sale they did.

---

So, on its own, the data would not be in any way really actionable, but in combination with other data that TLD operators themselves have with the registries, [inaudible] indicator.

SAMANEH TAJALIZADEHKHOOB: And to add to that, ccTLDs [inaudible] to gTLDs. They can see ... If they volunteer and join the project, they can use the most API to receive their daily metrics out of DAAR.

ZARKO KECIC: Okay. Thanks for that. [inaudible] because I expected that we would have face-to-face meeting and I'll meet the two of you and talk about some details. [inaudible] presentation or a webinar or whatever. It's [inaudible] ccTLDs to explain the benefits that we can have by joining DAAR, to help in fighting abuse, but it is still unclear to most of us how we are going to benefit out of just having metrics.

JOHN CRAIN: The answer is yes. we just need to figure out the logistics. Why don't you reach out to me separate or to Samaneh separately, Zarko, and we can work on that? Or maybe include ... Reach out to me and then I'll include the people who need to be included. We'll figure that out.

ZARKO KECIC: Okay, thank you.

---

LAURIN WEISSINGER:           Awesome. So, we're already two minutes over. Steve has a quick comment and then I think we should wrap up, keeping in mind we're already over time. Steve, please.

STEVE:                           Great. Thanks, Laurin. I think this has been a really good conversation, and as I've been listening to it, not being [inaudible] and DAAR and all the other research and stuff that's going on, there's been questions and answers beyond the breadth of DAAR and I've offered to the review team, [the sub-team], that you guys go and look at the recommendations to maybe think about that DAAR is a tool and one of the first tools—arguably CCDS was first—but maybe the recommendations are more about the strategy and the way that ICANN runs its research and analytics in that sense and not necessarily focusing on a specific tool, unless there might be some specifics inside of DAAR itself but the questions that were asked to me felt they were beyond just this tool of DAAR.

So, that's it. I just wanted to offer a reflection. Thank you.

KC CLAFFY:                   Totally agree with that. Thanks, Steve, for that comment. And thanks, everybody, for this call. This is really helpful. All right.

JOHN CRAIN:                 Yeah. Anytime I think.

---

SAMANEH TAJALIZADEHKHOOB: Indeed. Thank you for all the input.

LAURIN WEISSINGER: I'm sorry, I realized I was muted when I tried to wrap up the call. I would like to thank all the members of the OCTO team, and Jennifer and Brenda for doing this for us. This was I think really helpful, good conversation, and yes thank you so much for helping us out with this. Have a good day, everyone. Thank you.

JOHN CRAIN: Thank you.

UNIDENTIFIED FEMALE: Thanks, everyone. Bye.

UNIDENTIFIED MALE: Thanks. Bye.

SAMANEH TAJALIZADEHKHOOB: Thank you. Bye-bye.

**[END OF TRANSCRIPTION]**