ICANN76 | CF – NCAP Work Session
Tuesday, March 14, 2023 – 09:00 to 10:00 CUN

JENNIFER BRYCE: Hello, everyone, and welcome to the Name Collision Analysis Project Discussion Group session. My name is Jennifer, and I am the Remote Participation Manager for this session, together with my colleague, Kathy. Please note that this session is being recorded and is governed by the ICANN expected standards of behavior. During the session, questions or comments will only be read aloud if submitted within the Q&A pod. If you would like to ask your question or make your comment verbally, please raise your hand in the Zoom room.

This session includes automated real-time transcription, and please note that the transcript is not official or authoritative. To view the real-time transcription, click on the closed caption button in the Zoom toolbar.

And to ensure transparency of participation in ICANN's multi-stakeholder model, we ask that you sign in to the Zoom sessions using your full name, for example, a first and last name or surname. You may be removed from the session if you do not sign in using your full name.

And with that, I will hand it over to Matt, the NCAP co-chair. Thank you.

MATT THOMAS: Thank you, Jennifer, and welcome, everyone. It's nice to be here in person. We haven't been in person for a while, so I think maybe we'll

start off with a roll call around the table, just quick introductions. Jeff Schmidt, you want to lead us off and just say?

JEFF SCHMIDT: I think you did it for me. I'm Jeff Schmidt. Nice to meet you, nice to see everybody again.

RUSS HOUSLEY: Russ Housley at SSAC.

JENNIFER BRYCE: Jennifer Bryce, ICANN Org, Office of the CTO.

STEVE SHENG: Steve Sheng, SSAC Support.

KATHY SCHNITT: Kathy Schnitt, SSAC Support.

BARRY LEIBA: Barry Leiba, SSAC.

SUZANNE WOOLF: Suzanne Woolf, NCAP Co-Chair.

MATT THOMAS: Matt Thomas, NCAP Co-Chair.

ROD RASMUSSEN: Rod Rasmussen, SSAC Chair.

GREG AARON: Greg Aaron, SSAC.

MERIKE KAEO: Merika Kaeo, SSAC

JAAP AKKERHUIS: Jaap Akkerhuis, SSAC.

MATT THOMAS: Thank you, everyone. Great to see you. I think our next administrative item is always asking for updates to the SOI. So, seeing no hands, we'll continue right into the agenda. So, we are focused on one item today, and that is continuing to walk through our findings. We're going to pick up with finding B.D. The link there is for your convenience.

SUZANNE WOOLF: For those here who have not been following this work closely, because usually it's the discussion group in the meetings, we have done a great deal of background work or introductory slides on the status of the project in, I guess it was the SSAC intro. But what we're doing now is taking that background work and we're at the stage of trying to turn that into findings and then recommendations. And, of course, that's the part where you find out whether people are really on the same page as far as what they think everything that we've talked about and studied really means. The working draft is very much a working draft, so we're

trying to keep the discussion high level and sort of generally identify, does this finding make sense, is it supported, and what does it imply for recommendations, which we'll be doing later.

ROD RASMUSSEN: For SSAC members who are not part of the discussion group, I've asked SSAC members to—because we're here and this is our meeting—to be here and observe what's going on. And unless you have something really pertinent, please, this is a discussion group meeting, but we do want folks to pick up on that. And I know there are several SSAC members on the discussion group. There's an overlap here. So please take in what's going on, and if you're on the discussion group, please chime in. Thanks.

MATT THOMAS: Thanks for that. And why don't we go ahead and start with discussing our findings. We left off for the risk of CDM data manipulation. As we've done in the last two or three discussion group calls, what we're going to do is put the text, or the rough text up on the screen, give everyone a couple of minutes to read it, and then we will open up the floor for discussion, comments, and then we'll make sure Heather has any other questions or comments or needs more input from us before we move on to the next finding.

So the first one is the risk for CDM data manipulation. And for those of you that don't remember what a CDM is, that is the critical diagnostic measurement. That is the quantitative data measurements that are being collected either—well, throughout the data analysis collection,

looking at typically things such as query volume and diversity, where diversity spans multiple different vectors such as QNAMEs, labels, Q types, and so on. So with that, I'll pause for a couple minutes and let you all read the text. This is actually copied from the sections one through three for the area around data manipulation. So hopefully this text looks a little bit familiar. Jeff.

BARRY LEIBA: As one of the people who was pushing for this text, I like what's here. So thanks.

MATT THOMAS: Thanks, Barry. Jeff.

JEFF SCHMIDT: Thanks, Barry. Jeff Schmidt. And yeah, I like this text and I'm glad we have continued to develop this content. I want to say in general, I think that the risks associated with gaming are material for the next round. keep in mind, we're talking about the people that figured out how to game digital archery in the first five minutes of implementation and invented drop catching.

So I think we need to take very seriously anything that we do will be gamed. Imagine there's 500 applicants for .coin and .crypto and .wallet and whatever, right? There's a whole other universe of possibilities that could go sideways.

So a couple of specific suggestions. First of all, I think we need to delineate between historic accidental protection of collisions and

potential future intentional collisions. And what I mean by that is everything we've done for the past 15 years has been around protecting people that are basically good actors trying their hardest but made a mistake. That's very different than somebody that intentionally tries to simulate traffic for .jeff because of intentions to confuse or add problems later. So kind of delineating good actors from bad actors.

And then the other is related to the CDMs specifically, I think we need to maintain a lot of flexibility and discretion for what CDMs we look at and how we look at them because some CDMs are inherently retrospective. What we did in 2012 with DITL data, looking at things like Farsight data or more generally passive DNS data, resolver data, these are all kind of rear looking, harder to game, not impossible, but harder to game. Things like trial delegations and honeypots and other things are all kind of forward looking by definition, right? We know that .jeff is going to get delegated next Tuesday or we can detect that it's been delegated next Tuesday. And so game on. Thank you.

MATT THOMAS: Thanks for all that, Jeff. And Casey, I see your hands up. Thanks for joining us.

CASEY DECCIO: Yeah, thanks. Just a quick thought. I don't have any general concerns. I think it's good as Jeff suggested. There's one sentence in the middle that sort of seems at odds with itself. And that is the one that starts with at this time. At this time, there's no way to predict or prevent this type of manipulation. And then it says,, identifying the data to differentiate

between legitimate name collisions and fabricated ones requires longitudinal data analysis.

So it sort of says, we can't do it, but here's what you need to do it. And then it goes on to say, but by the way, you can't do it with that data. Or it would be hard. So I get what it's saying and I get that the point of this is to say, hey, this is a problem. I guess what I'm saying is if you want to point out that you may have some avenues forward to try, here's some things you might do, but the way it's written right now, it sort of says you can't do this, but here's how you do it. And that seems to be a problem for me. So a minor thing though, I think it can be fixed with some just adjustments there.

MATT THOMAS: Thanks, Casey. Rod, over to you and then Anne.

ROD RASMUSSEN: Thanks, Casey. I see how you're reading that sentence. Actually, I think you may just need a little more clarification, which I'm sure Heather can clean up. Because it's talking about two different things. It's the prediction of this type of manipulation, full stop. So in other words, we don't know who and what's going to happen. We can't predict this actor, that actor is going to manipulate this thing, other than we think there might be some manipulation. And then separately from that is identifying the data to differentiate between the two things. So they're not really contradictory. They're just talking about two different things. That's what I'm interpreting. And I'm seeing a couple of heads nodding in the room here. Thanks.

MATT THOMAS: Thanks, Rod. Anne, over to you and then Barry.

ANNE AIKMAN-SCALESE: Yes, thanks. This is Anne Aikman-Scalese and I am a member of the discussion group. I agree with Casey and Rod's observations. And I would suggest that we could say at this time, there's no way to prevent this type of manipulation, period, full stop. And then have a next sentence begin with identifying the data to differentiate. And that that would be a clearer way to express it. So I want to suggest that to the group.

And then secondly, and I know this is pretty minor, but I think I previously requested in the second paragraph toward the end of the sentence, it says in the second line, this is a difficult problem that will likely require bespoke data. And I know this word bespoke, we all provided a definition for it, but I just don't think it's in common usage and given language clarity that's needed in the community, we need to find a different word other than bespoke. Thanks.

MATT THOMAS: Thanks, Anne, for all those suggestions. And over to Barry.

BARRY LEIBA: So first on the last thing, it's not bespoke data, it's bespoke data analysis efforts. I don't care about the word bespoke, I'm happy to change it. But just wanted to make that point. The other is I agree with Rod's analysis of the other sentence, that the point is that we're saying

there's no way to predict or prevent the manipulation. And we need to detect it after the TRT's job is to detect that it's been manipulated and deal with it. So yeah, I suppose splitting the sentences may help, but maybe a little rephrasing to make that point clearer is useful.

MATT THOMAS: Thank you, Barry. And I don't see any other hands or... Please go ahead, Warren.

WARREN KUMARI: We can happily tell the TRT that they need to be able to detect this, but I think we've all had discussions and agreed that there's no way that we've been able to come up with that anybody might be able to actually do so, right? Like anybody who's worth their salt should be able to manipulate the data in a way that it's not obvious that that happened. Existence proofs of this are things like looking at data that we have, we suddenly see a large spike in leaked queries which aren't gaming or manipulation. They're just sudden increases in the queries like .console as an example.

So we can say the TRT needs to be able to review these and determine when there's gaming, but I think we all know that that's not really feasible. So, I don't know what we do about making suggestions we know can't be done. That's all.

MATT THOMAS: Thanks, Warren. And back to Barry.

BARRY LEIBA: Yeah, I think you know, Warren, that I largely agree with you. The TRT may need a forked stick or some similar weird device. I think this is the best we can do here. Making sure that the TRT is aware of this that we have considered it. And we say that we know that what we have is fragile, but it's what we have and the TRT needs to do their best. I think that really is the best we can do.

MATT THOMAS: Thanks, Barry. And Anne, I see your hands up. Please go ahead.

ANNE AIKMAN-SCALESE: Yeah, thanks. Do you mind if we could get some real-time comments in the document or real-time edits based on the agreement that is coming out around the table?

HEATHER FLANAGAN: I've been taking notes on the side and thought about putting them in the document. But since it's on the screen, I was a little worried that folks would find all of that distracting. I've already got about a page of notes for what I want to consider in changing the document.

ANNE AIKMAN-SCALESE: Yeah, thanks, Heather. We do do that in other meetings, for example, in GNSO Council. And it does help people to see what the proposed edits are when they're as simple as this one. So I'd like to recommend that to the group. And I don't know if you're uncomfortable with that approach. I don't think it's distracting. Thank you.

MATT THOMAS: Thank you. And Rod, and then over to Jeff.

ROD RASMUSSEN: Yeah, just to agree with Anne on that, the SSAC typically does that too when we're editing documents. So it's just really up to the group itself. So we could decide what we want to do.

JEFF SCHMIDT: I think in response to what both Warren and I said, we could add a sentence or add a caveat that's specifically to look at data generated before, during, and after strings become publicly known.

MATT THOMAS: Thanks, Jeff. With that, I'll ask Heather, I know you mentioned that you have about a page of notes. Is there any other questions, comments that you'd like to get out of the group? Or should we move on to the next tentative finding?

HEATHER FLANAGAN: Let's move on.

MATT THOMAS: Excellent, all right. So the next finding kind of dovetails off of data manipulation and it's talking about timing issue. Just kind of as a general overview, I think this finding really strikes at the core of the sequencing of events that will go on in subsequent procedures versus what happened in 2012. 2012, the applications were submitted and all

the applications closed and then essentially the name collision analysis occurred afterwards. That is not going to be the case going forward. There's also the options of, is it a first come, first serve? Or is it a round? Is there contention sets and auctions? And how does that all play into data manipulation and the timing issues?

So again, here I'll just pause for a couple minutes, let the text be ingested by you all and then we'll have some discussion.

SUZANNE WOOLF: My question here is, is this really a separate finding from the previous one? Or is timing one of the characteristics that we have to point out around data collection and data analysis and potential for data manipulation? Part of what has to be done, just as an editorial matter here, is make sure that we have separated findings and recommendations because we want to be able to draw a logical line between the findings and recommendations, but they are separate. And I'm looking at this and wondering if we can't tighten it into one paragraph of findings that can then motivate a couple of recommendations.

MATT THOMAS: Thanks, Suzanne. And I see some hands have gone up for discussion. So first over to Anne.

ANNE AIKMAN-SCALESE: Yeah, thanks, Matt. And I certainly would not object to what Suzanne has suggested. I want to ask a question regarding that second bullet

point. And the question relates to the history of name collision analysis in the community. It says previous round was unique in a way that name collisions were more of an issue/topic/concern after the application process closed. I don't know if that's very precise in that my recollection—and there are those at the table that I'm sure have an even more clear recollection of mine—is that this issue was raised by the SSAC fairly early on. And I can recall discussions in Beijing about name collisions and a warning to the board about name collisions. And then the issue was more like, when did we start paying attention to that? So there are aspects of this first sentence in particular that strike me as maybe not historically accurate and maybe others with better knowledge can comment.

MATT THOMAS: Thanks for that, Anne. I think that's an excellent point you just raised up. So thanks for crystallizing that. Over to Tom and then Casey.

TOM BARRETT: Hi, guys, Tom Barrett from EnCirca. Yeah I just stumbled a little bit on the second bullet as well. I wasn't even sure the last sentence there was a complete sentence. I think I just, all the parentheses kind of tripped me up. But I think this is a good point to make. And it's not just about, don't forget gaming can go both ways, right? They can try to advantage an applicant. They can also disadvantage an applicant. And so there's a whole series of legal objections that are allowed to applications that may also try to rely on collision data as the basis for their objections. So I'm just curious if we could insert the fact that objections are also are potentially impacted by collisions as well.

MATT THOMAS: Thank you, Tom. Barry in the room.

BARRY LEIBA: I raised my hand for the same thing that Anne did, that first sentence in the second bullet. I would say the previous round was unique in that name collisions became an issue, topic concern only after the application process closed. I think makes that point clearer. And I hadn't noticed it, but I agree with Tom. The last sentence is not a complete sentence. There's something wrong there.

MATT THOMAS: Thank you, Barry. Trying to handle both the hands in the queue online and in person. So I'll go to Warren first and then over to Jeff Schmidt.

WARREN KUAMRI: So sort of more on the meta point of Suzanne's question, I definitely understand what she's saying, but it does feel that this is a big enough and sort of important enough differentiation that it is possibly worth keeping as its own separate finding. It's not specifically just on any of the, can we detect these or not, but it's more on the sort of larger, larger thing.

As for the collision stuff, I think that most of the collision stuff discussions actually kicked off around the time of SAC 57, which was the SSAC advisory on internal name certificates. And that was published on March 15th, 2013. And it was actually sort of embargoed or whatever the right word is until publication. Because it had potential security

impacts. So as for when people started freaking out about it, I think that that might be when it became an issue. So that's all.

MATT THOMAS: Thank you, Warren. Jeff.

JEFF SCHMIDT: Thank you. I think historically, collisions have been around since .CS and SiteFinder and everything in the '90s. But also, I'm not sure that's important. I'm trying to distance myself from that. But I'm also not sure important. I'm trying to distill this. And not the point we're trying to make here. I agree with Warren. I think we're trying to distill this whole section to this issue of before strings are known versus after strings are known, right? That's the finding. And everything else is kind of distracting noise at this point. Thank you.

MATT THOMAS: Thanks, Jeff. Casey, I saw your hand was up originally. Did you want to chime in?

CASEY DECCIO: Sort of, but I guess my concern now is I just, I'm trying to wrap my head around this. And so I wanted to listen to other comments on this. But maybe one lingering, perhaps more minor concern is about the title. Timing issue needs a little bit more specificity and maybe that would help me. I know that there's some more descriptive text below, but, and I'm sure that's on the editorial to-do list, but it would be great to be a little bit more specific about what is actually being said in that title.

MATT THOMAS: Thank you, Casey. Anne, over to you.

ANNE AIKMAN-SCALESE: Yeah, I think that's a very good, it's Anne again, very good observation by Casey. And it seems to me what we're trying to say is that this time around, we're going to look at this issue before applications. And so maybe the finding is risk assessment should be done prior to the application, or not prior to, but prior to award. Risk assessment should be done, yeah, as part of, I think that Heather was working on, as part of the application process was probably even better as suggested by Heather there. And that's probably a finding.

MATT THOMAS: Thanks, Anne. Casey?

CASEY DECCIO: Yeah, and I get the idea here, but we also want to be careful about what Suzanne talked about earlier, the idea that we're making observations about certain things and then making recommendations based on those observations. So using some care and how this plays out so that we have the recommendations based on the findings and not part of the findings.

MATT THOMAS: Thank you, Casey. And at this point, I'm just going to go over to Heather and ask you if you need any other discussion input or you have what we need for right now. Heather?

HEATHER FLANAGAN: There's one thing I didn't understand. That was a comment from Tom about legal objections. That isn't something I'm familiar with or have heard us talk about before. So I don't know what to do with that.

MATT THOMAS: Thanks, Heather. Tom, would you mind clarifying or expanding on that?

TOM BARRETT: Legal objections are the ability for some unknown third party to object to an application due to the fact that the application might infringe on their rights. So typically it might be a trademark order, but it could as well, just as well be someone else. And so certainly the collision data, the presence of certain collision data might be a justification for a legal objection, which of course occurs after the application has been submitted. But it's just a point that the whole gaming issue, which is what we've been talking about and the fact that now that we know that collisions are an issue prior to the application, it really was not the first round, that it raises all kinds of implications about things like legal objections to applications.

MATT THOMAS: Thanks, Tom. And Suzanne?

SUZANNE WOOLF: So what I'm hearing is that maybe the finding here—risk assessment should be done as part of the application process. And Heather, I am not picking on your words. That feels more like a recommendation to me. I think the finding is that the vulnerability of the process to manipulation varies with the timing, the specific data involved, the data analysis techniques involved, and that therefore sensitivity to those vulnerabilities in general has to be part of the picture. But I think the finding is that there is there's incentives and there's vulnerability to manipulation that varies with these multiple factors.

MATT THOMAS: Thanks, Suzanne. Heather, does that clarify what you need for this section?

HEATHER FLANAGAN: Yep, thank you.

MATT THOMAS: Thank you. With that, why don't we go on to the next one? So this is really focused on framing name collisions as a risk management process. I'll just leave the roughed-out bullet points up here for everyone to read, and then we can start having a discussion. Casey, I see your hands up. Casey, Please go ahead.

CASEY DECCIO: Yeah, I'll give my general thoughts on this. And I've expressed this before, but I'll express it again here. I have general concerns with this phraseology here, this idea of a risk management process, and not even because it's wrong per se, because to me, it's just very ambiguous and it feels like a compound statement that builds on several other things when really what this is saying is a few things in my opinion, and that is that the data is limited. Even with the data that we do have, there is some uncertainty, and we've expressed those in two separate findings.

And then the third thing is that then because of that, we're left to make decisions that are based on the best available data that we have and the best available analysis of that data that we have. In my opinion, we should say it that way instead of trying to compound it into this sort of ambiguous phrase that I've never really known what to do with because it feels so ambiguous. So that's what I would do. I would say, nope, make it into three or more simple things, at least two of which we already have, and not try and sort of compound it with this one finding that seems to try to put everything together. That's my opinion on this.

MATT THOMAS: Thank you, Casey, and I see some comments in the chat from Rod suggesting that we switch to Black Swan instead of Unicorn, as well as a question from Warren around, is it risk management or risk assessment? Hands going up, Jeff, over to you.

JEFF SCHMIDT: Thanks, and yeah, I'm with Rod, and I think, Rod, you suggested switching to Black Swan, and I prefer that. I think it's a more understood

term. First of all, pursuant to the earlier point, this feels more recommendation-y than finding-y, but regardless, there's nothing there that's particularly wrong, but I think the objective here is it's a multi-step process. We have to assess the risk, manage it down, and then, and the important point on the bullet that I would like to see there that's missing is the, and then be prepared if we're wrong, right, because it is an imperfect process. We're not going to get them all right. We're going to do our best to assess in advance based on all the criteria we've written about here, but then we have to be prepared if we do something wrong with some sort of an incident response and detection and mitigation procedure, and to me, when you say a risk management process, all of that together is what that means to me. Thank you.

MATT THOMAS: Thank you, Jeff. Any other comments or discussion from the group right now? Oh, Anne, over to you.

ANNE AIKMAN-SCALESE: Yeah, thanks. I tend to agree that maybe this is more of a recommendation than a finding. I do think that the phrase being used is accurate, however, in terms of talking about a risk management process, I think that is a phrase that is understandable to the community and needs to be retained, but perhaps it belongs in recommendations rather than in findings.

MATT THOMAS: Thank you, Anne. Any other comments or questions from the group? I'll ask Heather then. Heather, do you have any other questions that you'd

like to tease out of the group or have further discussion on particular subjects here?

HEATHER FLANAGAN: No, though I would note that we likely won't use either unicorns or black swans in the final text because both are equally unknowable to people around the world. So don't get hung up on that.

MATT THOMAS: Thanks, Heather. With that, I don't see any other hands going up. Why don't we go on to the next finding? Well, actually, we had a little bit more text there. I'll just put that up there for a second. And if any other hands go up, we can continue discussing. Otherwise, we'll go on to the next finding.

All right, well, move on to the next one with the bullet points here. Just pause for a minute, let you all read it, and then we can continue our discussion. Just as a reminder, an RSI is a root server instance or identity, right? And then a PRR is a public recursive resolver, and the RSS is the root server system. Any comments, questions, concerns here? Casey, over to you.

CASEY DECCIO: Yeah, two quick thoughts. One is there needs to be some specific—what do I want to say? When we say visible, that needs to be qualified somewhere because we're talking either about traditional passive collection data locations, for example, the DITL dataset. So that's one thing. I think this needs to be qualified somehow to existing datasets or

proposed datasets or whatever makes sense to put there. There needs to be some qualification.

The second thing is, and I agree that some of this is legitimate, meaning like the idea that, okay, you have, for example, aggressive NSEC caching, this idea that you may not see all queries at root or other servers because of this aggressive caching at the resolver, and so those aren't going to be exposed.

Now, over time, probabilistically, you may see some of those still, and so some of that might be mitigated, but you're right, there's some possibility that they may not make it all the way. But then there's some other things, for example, like this idea that's written here of split horizon DNS.

Well, the question is, and this is a question for the group, if these collisions or potential collision strings are not actually colliding, then do they matter in this particular case? And I ask that because if so, then we need to mention this as sort of potential collisions as opposed to actual collisions, because they are different things. And I'm not saying, by the way, that it's even good practice to be doing that, maintaining your own private DNS space, but all I'm asking is if it's not actually reaching the public DNS, then number one, is it a collision, and number two, does it matter? So those are my two points.

MATT THOMAS: Thank you, Casey. I see Warren has his hand up, so we'll go over to him.

WARREN KUAMRI: So two things, first responding to Casey, yeah, I think it does matter, because even if the names are not leaking into the DNS, if somebody is using a string and therefore cannot resolve the external version of the name, then it matters to both the person using the string and whoever applies for the name.

So for example, if somebody like Amazon is using .internal, if that were to be delegated, even if it's not leaking into the DNS, it means anybody using Amazon AWS isn't going to be able to resolve anything in the real delegated .internal. So I think that it is a significant issue.

And then also, what I was going to say, I just dropped in a link with a bunch of other things on reasons that you might not be able to see the data. So there's QNAME minimization, there's aggressive NSEC caching, etc. And hopefully my explanation of why, even if things don't leak into the DNS, why it's an issue was coherent and made sense to folks, please let me know if not.

MATT THOMAS: Thank you for that. Warren, appreciate it. Jeff, I see your hand went up. Over to you and then back to Casey.

JEFF SCHMIDT: Thanks. I agree with everything written here, but I'm trying to kind of elevate it to what's the essence of what we're trying to say here. And what we're trying to say is the finding is this is an imperfect process. Nobody has full visibility on everything that's happening on the internet all the time. And so it's going to be an imperfect process. We need to try our best given data available and we need to be prepared if we're

wrong. But I think that's the essence here, is that it's an imperfect process.

MATT THOMAS: Thank you, Jeff. Casey.

CASEY DECCIO: Yeah, just to follow up to Warren's comment, and I appreciate that and I agree with that. The one thing that I would add then is just maybe to be careful about the terminology there when we talk about those things. And maybe it's a minor thing, but I guess my question is, and maybe the group still thinks it is a collision. My question is whether it's still considered a collision or if it's a potential collision sort of before the fact, and maybe it doesn't even matter, I don't know. But I agree that Warren's right, the idea that maybe you can't resolve something now. And so it's a potential collision that could be a real collision at some point. Anyway, that's my follow-up comment, but I appreciate that comment from Warren.

MATT THOMAS: Thank you, Casey. Any other questions, comments, concerns from the group at this point? Anne, over to you.

ANNE AIKMAN-SCALESE: Thank you for your patience with the non-technical participant. But I'm wondering after all this discussion, what's actually wrong, if anything, with the title of this paragraph? What everybody said just seems to me

to confirm not all name collisions can be made visible and that that's an objective and neutral finding.

MATT THOMAS: Thank you, Anne. Jeff, I saw your hand went up. Did you want to comment?

JEFF SCHMIDT: I was thinking about what Anne just said. My comment was this is actually a really good additive and new finding that's genuinely additive from this effort. And I think it's important and informative for the rest. That was my comment. I don't disagree with what Anne said. Again, I think we're all agreeing and we've spent the last seven minutes agreeing. We could probably clean up the word a little bit or the words a little bit. I think we're all in agreement.

MATT THOMAS: So with that, I'll ask Heather. Do you think you have the appropriate text input? Do you need anything else from the discussion group at this time?

HEATHER FLANAGAN: The good news is that yes, we all fundamentally agree that there's something here that is useful, which is great. But what I heard a couple of people say is that the word "visible" is actually not specific enough that it needs to be qualified in some way. Visible to whom I think is how Rod put it. And I guess based on everything else we've written so far, I think the answer to that is visible to the TRT. Is that correct?

ROD RASMUSSEN: Yeah, and I almost said exactly that, Heather. That was what I was thinking, but I figured I'd leave it a little more generic for the discussion.

MATT THOMAS: Any other concerns, Heather, or is that... Can we move on?

HEATHER FLANAGAN: That's it.

MATT THOMAS: Excellent. We're making great progress. All right, let's pull up the next tentative finding. DITL data is insufficient as a quantitative measure. I'll just pause here, let you look at the bullet points, and then we can have some more discussion. Yes, Warren, over to you.

WARREN KUAMRI: Actually, because it's an important point, I think can I respond to Rubens on his comments that he made in the chat. Specifically, he has something like .Rubens as an internal DNS, and Rubens is delegated as the TLD. He won't be able to access websites within .Rubens because of that, so it's a collision, but my internal .Rubens service won't be disrupted. That's actually only partially true. It depends on how you're actually constructing those names. So if you are relying on search path processing at any point to expand the name, so for example, if your search path has .Rubens in and you type in foo.accounting, and you are relying on search paths to add on the .Rubens, then it will actually break

by the standard and by definition if a name is entered and contains a period anywhere in it, that will first be looked up in the DNS before any search list processing is appended. So if in the example, you always entered the full name, then yes, he's correct, it would not break anything. But if you are using search path processing at any point, which I'll point out a lot of enterprises do, then it will actually break your internal usage of the name. And sorry to sort of bring that up, but I thought it was important enough for the rest of the discussion and level setting.

MATT THOMAS: Thanks for the technical points there, Warren, appreciate those. Casey, I saw your hand go up, please go ahead.

CASEY DECCIO: Yeah, this is related to the new topic. This DITL data is insufficient. And my only concern with this is that I do agree that the DITL data, we certainly have plenty of evidence to show that the DITL data is limited in what can show in terms of visibility at the root and because of who's contributing and because of affinity for certain root servers and location, everything else. My question is whether that makes it insufficient or simply limited. Because insufficient sort of says, yeah, we already know it just won't work, but it's sort of a blanket statement. And I would really push for saying it has limitations or whatever, as opposed to saying insufficient, because insufficient to me sort of says, yeah, we just can't use it anymore. But I wouldn't say, no, we can't use it, we just have to understand that it might, I don't even know if we'd consider it the best we have right now, but it's among the only things we have right

now. And so expressing that it has limitations is good. Saying insufficient might just be premature because it may not be insufficient. We don't know that.

MATT THOMAS: Thank you, Casey. Jeff, over to you.

JEFF SCHMIDT: Thanks. So I'm on the second topic here as well, [inaudible]. And I agree with Casey, I think this is unnecessarily specific to DITL. And basically I don't understand why we would trash DITL data so specifically, particularly that that second bullet just seems really inappropriate. I think the finding here is no one data set is perfect. Every data set has strengths and weaknesses. And we could make a grid, right? You know, DITL has several strengths. It's what we have, we have history, it's collected in advance, etc. You know, we can make a grid. Null delegation would give you certain pluses and minuses. You know, Farsight data would give you certain pluses and minuses.

And the point is, the reason we have a TRT is to use any and all, plus anything else that becomes available, and make the best decision available with the data at the time. Thank you.

MATT THOMAS: Thank you for that, Jeff. Over to John.

[JOHN:] I'm wondering if it's worth mentioning that it wouldn't be terribly hard for a hostile party to manipulate this stuff. Like if I wanted to make sure that nobody ever delegated .pickle, and so I have a reasonably good idea when people are doing DITL surveys. You know, I'd fire up a botnet and send up a billion .pickle requests.

SUZANNE WOOLF: If only you'd been here 20 minutes ago. Yes, it's a valid point, and we have actually spent a fair amount of time characterizing it. And when you have a chance to read what is there after Heather has a chance to work on it, further comment would be welcome.

MATT THOMAS: Any other questions, comments, concerns from the discussion group at this time? Sure, Rod.

ROD RASMUSSEN: Yeah, so I had a similar concern about that second sub-bullet point, but I believe that's from an actual, that's a quotation from a different thing? Findings from perspective study, or is it not? I don't think it's a direct quote. So if it's not a direct quote, I agree that it's very pejorative.

MATT THOMAS: Thank you. Noted. Jeff, go ahead.

JEFF SCHMIDT: Yeah, thanks. Actually, I mean look, I mentioned Farsight data, right? Or whoever they are now. Domain tools? Domain tools. And then more

generally there may be various commercial implications of any data that's looked at, right? And like, so again, calling out DITL and DNS-OARC here just seems grossly inappropriate. There are commercial implications of any and every data set.

MATT THOMAS: Thank you, Jeff. Suzanne?

SUZANNE WOOLF: Yeah, I think the finding there might be without the pejorative overtones, but there is a finding there that I think Jeff sort of pointed to, that any single data set is subject particularly provided by a third party, for whatever reason, under whatever circumstances, is subject to changes, compromises, issues of any kind. And the recommendation that would seem to motivate is to make sure that there is reliable relationships and reliable ways of continuing to get multiple data sources for any of this to work.

MATT THOMAS: Thank you, Suzanne. Warren, I saw your hand go up. Please, over to you. And then Casey.

WARREN KUAMRI: So, yeah I think that it's going to be almost impossible to stop any sort of gaming. Having a wider view will definitely help some. But I just dropped in a link to a photo showing the root data ranking thing. And that was literally done in as long as a YouTube advert takes, which is what, like four or five minutes? Actually, it's less than that, I hope. But I

think that the only potential way to deal with gaming is to have a very long time horizon and look back and make sure that whatever strings are being considered existed before application times were opened.

However, from research that Matt and Wes Hardaker and a bunch of other people have shown, there is a lot of spikiness in the collision data, right? If you have a look on the list of top strings, they change a huge amount over time. Console being one of the obvious examples, but even just looking at the standard magnitude data, the top couple of names change quite often, even in the top sort of 15 or 20, names come and go.

So I think that the gaming thing is a big issue, and I still don't know if we have any sort of way that we've addressed it. I personally don't really think that it is actually something that is addressable, right? I think that anybody who is willing to spend more than like a few thousand dollars would be able to easily create a system which will be completely undetectable and will be able to create gaming data that is indistinguishable from real data.

And while no one data set is perfect, unless you are able to get a significant number of data sets and have them include things where it is very difficult for somebody out on the Internet to be able to be included in that data set, I think you will have a bad time. So an example of something where I think that that might be possible is if you could get data sets from recursive resolvers, which are not open to the internet, so like ISP recursive resolvers, you could potentially compare that to root data and open recursives and see if there is a delta in the lookup type. But again, I think that really won't work very well because

comparing recursive resolver sets to each other and recursive resolver sets to the root, there's a huge variance. I realize this just sounds like doom and gloom, so sorry.

MATT THOMAS: Thanks, Warren. I have quite a few hands in the queue online and a couple in the room. So Casey, over to you, and then Rubens, Anne, and then Jeff, and just be cognizant, we have four minutes left.

SUZANNE WOOLF: Yeah, I think we'll have to close after this item. Casey, go ahead.

CASEY DECCIO: Yeah, I'll make it quick. I just, I see that we make a lot of points in here about limitations of the data, which I think is very appropriate, but I think somewhere we need to make also the point that the data we have is the best data that we have. And although it's imperfect, we have to rely on that the best that we can with those caveats. And I think we need to point that out somewhere so that someone doesn't look at this and say, well, it's useless. It's like, we can't do anything with this, because that's not actually true. And we have actually some data to show that some of the data sets correlate very well with reported anecdotal experiences through these reports.

MATT THOMAS: Thanks for that, Casey. Rubens, over to you.

RUBENS KUHL: About the second bullet, besides being pejorative, it would also apply to ICANN, because ICANN is a nonprofit and a third party to many others. So it would be basically disqualifying our own work within ICANN by saying that. So we shouldn't go there.

MATT THOMAS: Thank you, Rubens. Anne, to you.

HEATHER FLANAGAN: Thanks, Matt. I very much agree with Casey's observation that no one data set is perfect, but in terms of recommendations, yeah, we need to do the data analysis. I was going to comment on the fact that I thought based on what Jeff Schmitt said, we were going to list some limitations in this paragraph of different types of data. So I wouldn't necessarily just cross out the DITL data has limitations. I think we need to mention what we see as the limitations of some of the data. And I do have a recollection, and this is a question, and it's probably a question to Jeff and Casey, at least. There was some discussion in the group about DITL data being less useful now than it was in 2012. So if I could tease that out from you guys as to what you meant by that, it would be helpful to my understanding whenever this report gets to wherever it's going. Thanks.

MATT THOMAS: Thanks, Anne, and I'll go over to Jeff right now.

SUZANNE WOOLF: Thanks, and I'll go fast. Many pluses to what Warren said a couple minutes ago. I also believe gaming is a material issue, and it has the potential to add quagmire and controversy really to the whole thing. Warren mentioned one of the best defenses we have against gaming is historic data and long live data, and I'll point out DITL is, I believe, one of the longest live data sets we have. Also, and this is the point that I wanted to make, we kind of get into this weird, almost moral hazard sort of argument of do we extend the same protections to somebody that was colliding 10 years ago than somebody that did something that is causing collisions yesterday.

And that's interesting, right? We don't protect people that mistakenly use non-RFC 1918 IP addresses. We say, hey, dummy, don't do that. Do we extend the same protections today to people that are using a namespace they shouldn't use? And that factors directly into gaming, data longevity, historical data availability, and all that.

And then finally, Warren made a point about comparing resolver data to other data sets, and I think Casey did that looking at the Farsight data, again, I'm saying Farsight data, but it's recursive resolver data from a number of closed ISP sources. And so that is a valuable tool, and I think Casey has data on that. Thank you.

MATT THOMAS: Thank you, Jeff. I'll go over to Warren. We need to make this brief, and then we're going to close with any other business and wrap up the session.

WARREN KUAMRI: Thank you. Yeah, I was just going to point it out that the link I put in, other than caching, all of those things are post-2012. So those are changes in the visibility that we have lost since then.

MATT THOMAS: Thanks, Warren. I think we had a good discussion on this finding. Heather, quick question to you. Do you have enough to at least re-edit this?

HEATHER FLANAGAN: Yes.

MATT THOMAS: Great. And then, so why don't we pick back up here? I don't want to cut the discussion short on this finding in case we want to expand on it some more in the next call. But a quick call for any other business. Seeing no hands, and I think I'll speak on behalf of Suzanne and I. Thank you all for a wonderful, productive meeting. And we're adjourned.

[END OF TRANSCRIPTION]