
KIMBERLY CARLSON:

Welcome to today's NCAP discussion group on April 1st at 19:00 UTC. In the interest of time, there'll be no roll call. Attendance will be taken in its usual fashion by those in Zoom. Kathy and I will update the Wiki with the names of the participants as quickly as possible.

We have not received any apologies today. All calls are recorded and transcribed, and transcripts [inaudible] public Wiki. Again, as a reminder, to avoid background noise while others are speaking, please mute your phones and microphones. And with that, I'll turn the call over to you, Jim. Thanks.

JAMES GALVIN:

Thanks, Kim. This is Jim Galvin from Afiliias, one of our co-chairs. We do have Matt Thomas with us, and I am hopeful that Patrik will catch up with us here soon so that we're all present.

Thanks, everyone, for joining. Very happy to have you all here as we move along here. First order of administrative business [two] is to ask if anyone has any updates to their SOI that they have made recently on the ICANN website and they want to just make note of it here. I'm not seeing any hands or hearing anything.

Now, Anne is saying in the chat room that she can't get to the document per policy against the Google account. The document—I don't know if I can adjust the sharing. It's possible to adjust the sharing such that—shouldn't it be possible for anybody who has the link to use it?

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.

UNIDENTIFIED MALE: Yeah, and I think that may be the way it is.

JAMES GALVIN: I'm doing it and I don't see that. Allow external access. Yeah. So anyone with the link. Okay, so I have now changed the sharing so that anybody with the link can access it. It should be the case. Try again, Anne, with the link that you have there. Or here, I'll also, just for the purposes of being certain, copy the link again and put it in here. Let's see if it turns out to be the same. Oh, Kim put it out there again. But it is the same link. Yeah, the link hasn't changed, even though the sharing ability has changed.

I don't think you need a Gmail address there, Anne, so try that. Oh, she got it. Okay, great. We just needed to change the permissions there. New members. We don't have any new members this week, although as is usual for the ICANN thing for those who don't track these things that carefully, people can always put their name up and go to the Wiki page and seek to join and become part of this group and do the SOI thing and all that.

So with that, we only have one item on our agenda for today, and we'll go through this. As we had discussed last week, last week we had gone into our spreadsheet that we had been using to track notes. They were tied and indexed directly to the board questions and what the board was looking for us to speak to, and we had a column in which we had done our own little gap analysis ideas where we tried to capture what we thought were some relevant details that might be important for influencing and providing input to the remaining task that Karen has

with respect to the work product coming out of study one. That, of course, is to make a recommendation with respect to whether or not studies two and three should be funded and that work should continue.

Last week, we agreed that the admin team would take on the task of providing some text to work with here. We would take all of those notes that we had from back then and turn them into just a couple of pages of some text here. Full credit for this awesome writeup goes to Matt, our new co-chair who just jumped right in and produced this document for us. So this is essentially our opportunity as a discussion group to review this and see if we agree with this. And if not, what are the questions or concerns that we need to address and to get to?

Maybe what I'll do is I'll give Matt a chance to take his opportunity. What we did was we took all of those elements that had been written up and basically turned it into four particular points in support of continuing with studies two and three. And Matt, why don't I give you a chance to give an introduction to those things? I think Jeff has his hand up. Let's let Jeff ask his question and then we'll go to you, Matt, to talk about what you have got there. Go ahead, Jeff.

JEFF NEUMAN:

Actually, I'll wait until after the intro. Thanks. I'll defer to the intro and then ask my question after.

JAMES GALVIN: Okay. Thanks much. So Matt, why don't I turn it over to you to just give a quick intro to your four points here? And then we'll open the discussion.

MATT THOMAS: Sounds good. So I took a look at the spreadsheet and the comments that we've been building as a group over the last several months and tried to kind of put them into a rough structure or taxonomy. And it came to me that there were three or four primary areas that I thought we should discuss as "gaps."

The first one was datasets. We've been talking about what kind of data is now available that hasn't been available since the previous analysis, and a lot of that revolves around recursive resolver traffic. Some of that also talks about DITL data and other types of data repositories such as the ORDINAL DNS from Jas and Corp.com as well as certificate transparency records that are now available.

But beyond just the datasets, the second thing that we seem to talk a lot about is just the general evolution of the DNS ecosystem and how the protocol and ecosystem has changed since the last round. And specifically, we've had things like aggressive NSEC caching, QNAME minimization. Now we have things like DOH and DOT coming out that dramatically change the observational capacity of looking at name collisions at various levels in the DNS hierarchy.

So I think it's important for us to be able to take that knowledge and understanding that's happened over the course of basically a decade and reexamine that in terms of how does that influence name collision

risk assessments. That's something that we need to consider as part of that gap.

And then the third item, which we talked about a lot more last week, and that was about controlled interruption efficacy and data analysis. So we do have those anecdotal reports that have been going up to ICANN, and I think we've all kind of reached a consensus last week that it would be worthwhile for us to dig a little bit deeper into those and maybe retroactively take that knowledge in which those name collision events did occur, and go back into previous datasets to see if we can understand, are there other signals that we've missed in those first assessments that would have indicated these risks and events, the name collision events?

And then finally, the fourth kind of catchall grouping I have is more around just the small but very targeted set of publications and vulnerability reports that focus on the name collision problem in general that have come out since the last round and now. I think it's prudent for us to take a look at the understandings of how they're identifying those risks, vulnerabilities and how does that play into future risk assessments of name collisions based off of those. Does that make sense? Jeff, do you want to talk about your question then?

JEFF NEUMAN:

Yeah. Thanks, Matt. So I apologize for missing last week's call. I was not able to be on it, I had a conflict. But a couple of things. This is almost written like an advocacy piece, and I think we need to kind of tone it down a little bit. It makes it sound like there were a lot more problems

than there may actually have been, and we almost make it sound like this is so urgent because [they were] catastrophic events.

I think what we need to do is we can talk about these four things as being needed: datasets, how it interacts with some new [RFCs,] but using terms like—it's a little misleading. And I'll go through it kind of from the beginning. So nowhere in the beginning does it talk about the fact that from the outset, we don't have any evidence—and there's nothing in the bibliography in phase one that says there were any catastrophic events or anything that's of such urgency, nothing that's been found has shown kind of any evidence of a huge problem.

So I think that's number one. But if you read the first introductory paragraphs, it almost makes it sound like we have to do this because there's a huge gap in everything. And I don't think there was that much of a gap. In fact, if you look at since the controlled interruption and this phase two report came out, there haven't been much of anything written, let's say in the past four, five years. Yeah, there's a couple things here and there, but it's very small.

So I think we need to be honest in this, yes, that we can do a phase two. I'm not saying we shouldn't be doing it. But I'm saying that we make it sound like this was such a huge problem and there are so many things that we need to do before we even do a round two. And I don't think we should be putting out a paper that gives that impression.

So while Matt, you talk about—sorry this is long, but in the paper it talks about how over the course of the last decade, numerous peer reviewed articles have been published. Well, that's a little misleading because a

lot of those peer reviewed papers were considered as part of the phase one and phase two reports that JAS had. Now, it may not be as thorough as we all perhaps wanted, but it really hasn't been numerous papers that have been out there.

And also—so you referred to this, and it was interesting for the efficacy of name collision, it's also in there that it talks about—and you said it again, anecdotal evidence or anecdotal reports. Again, that kind of sets the wrong impressions like, “ICANN acted with haste and therefore should never do anything again like that. It was only anecdotal and we should have done something more through to begin with.”

So sorry that was long, but we need to kind of take the subjectivity out of the paper and make it more, “Look, there's some data, there's new datasets that we should review. Yes, we should be more in depth with some of the things from the controlled interruption. Yes, there's new technologies out there we should review.” But I think we should tone it down a little bit. That's just me. Thanks.

JAMES GALVIN:

Jeff, thank you very much for that. Let me jump in here and say, yeah, you get no real argument from me about whether or not toning it down is a good thing or a bad thing. I think that's fine. If you have some suggestions and you're willing to go through this and make the edits in the document—as you're used to hearing, I know, from other places—that would be very helpful in all of this. But certainly, we can also take a crack at this afterwards and see what we can do here.

I do want to point out that what's important to understand about this document that's being produced here is this is not actually going to be—well, I guess maybe that's part of the question that we get to ask ourselves, and maybe that's what Matt Larson was going to get to talking about here in his thing. I'm sure that this'll be a piece of what he wanted to mention.

This is not going to be a formal, published work per se, this document itself. This is going to be something which comes out of this group as a consensus-based statement that'll be given to Matt Larson as the project owner here, and then he will be working with Karen as the contractor producing the work products from study one to produce a recommendation that will ultimately be published in some way as part of the [inaudible] that we have. Because it's their work that's going to go to the board, and we're providing input to that.

I only mention that because the formality of that nuance, I think, is just important to keep in mind. Ideally, we're just going to have our mailing list discussion about this thing and that'll be the historical record, and the official stuff will come out of the report that Karen produces that gets submitted to the board.

I think there was a second point I wanted to make there, but I see some hands up. So let me just pause for a moment and speak to the hands. Jeff Neuman, I'm going to assume that's an old hand and go to Jeff Schmidt.

JEFF SCHMIDT:

Hey. If you hear from one Jeff you've got to hear from both Jeffs. Hi, everybody. My question is kind of along the same lines. First, I want to say thanks to Matt. Everything in here—I'm really glad that you took the time to write it down, and you did a lot of work on this so I appreciate it.

My question is more of a, what do we need to do as a working group to support ICANN and the community in the next steps here? And what I mean by that is, in reflecting on what Jeff Neuman just said, there's a break here between—look, I'm the biggest collision nerd there is. I love this stuff, I've thought about this stuff and looked at it for a long time, as have many of us on the phone. But I think it's hard to say academic interest in a super interesting and super esoteric area equates to danger.

I think the data and our experience at this point with delegating 1600 of these things and the line chart in the first report showing name collisions reports being effectively zero for the last two years is very telling. Interest? Yes. Is this an interesting area? Yes. Would it be good to create awareness, would it be good to educate people? Absolutely. Is that danger? I think the onus is on us to make sure that we're not creating the illusion that there is danger in that point.

I think there's one question that we need to think about when we're going to new rounds. If the purpose of this is to prepare ICANN for another round, there's one question, and that one question is, how do we find other unicorn domains like corp.com? corp.com was and is extremely dangerous. There's probably others. We need to figure out how to find those. That's the question that in my opinion we need to be focused on to support the future rounds here. Thanks.

JAMES GALVIN:

So thank you, jeff Schmidt. I remembered now the second thing I wanted to say in response to Jeff Neuman's thing. Let me reiterate your last comment there with your last sentence there, which was, how do we prepare for the next round?

I don't think that our goal here—at least from my point of view, this is not our goal, and so if the working group, the discussion group here wants to disagree, then people should speak up and say so. Our goal here is not to announce danger. That's what we're not doing. We're not in it for that particular purpose.

Our particular purpose is the board has given us ten issue areas that we need to respond to because the board is in a place where it needs to be able to make decisions about whether or not any given name should be delegated.

So as you kind of phrased it in the end there, Jeff Schmidt, ideally, sort of the conceptual path we're on here is, what set of guidelines can we provide to the board, questions that we can put in front of them, that they will need to have asked and then answered about any given TLD? And then they can subjectively decide for themselves whether or not they're going to delegate a name and what mitigation strategies will apply.

So our purpose here in this study one is not so much that we want to do an academic study of an esoteric space, but it is fair to acknowledge that while we know a lot—or maybe we know everything there is to know, I'll even go so far as to say that with respect to name collisions

and the behavior, because as you say, nothing super dangerous has happened. I do think it's important to make that conclusion, to state that, draw that line, and then in some way, infer, extrapolate or invent—whatever the right word is here—the proper set of questions that can be used to provide as close as we can get to a predictable process for what will or won't be delegated in the future. And in particular, this will speak directly to the .corp, .home, .mail issue for the board as quick as we can get something together for that, and then a longer-term process that can stand the test of time as best as possible.

So I just wanted to kind of reiterate that. I feel like I've said that a few times, and I apologize for being a little long winded, but it is important to keep us on track here about our goal and the issue questions that the board has given us to answer. And we need to speak to that.

Now we have quite a few people, so I will let any other folks have a chat here. So Warren, you're up next.

WARREN KUMARI:

Thank you. I do want to make sure that we remember that absence of evidence is not evidence of absence. Right? The fact that we only got a few reported instances of issue from name collision doesn't mean that there weren't more name collisions. It just means that people didn't report them for a variety of reasons, including they didn't know where, they didn't feel like it, they didn't want to expose the fact that they've been using certain names internally.

I know within my own environment I'd set up a fake root and fake set of TLDs that I could do testing, and I was using .dev and .prod within my

fake one to emulate my development and my—actually I had .dev, .test and .prod to have a development, a testing and a production environment. And those definitely counted as collisions, but I didn't report it because it didn't seem like there was much point.

In much of the original discussion, we've had a lot more requests and requirements that registries need to provide a bunch of reporting on the number of hits that they were getting on things like the 127.0.53.53 [stuff] so that we could actually get feedback and understand how much this was happening, what sorts of names were happening. But that data wasn't collected.

So I'm just kind of concerned that a couple of people had said we didn't get very many reports and therefore we can draw useful inferences from that. I think that that might be jumping a lot further than the data actually—or I guess the real world shows. That is what the data shows, but I think we're somewhat tricking ourselves if we believe that the data actually reflects what all really happened in the real world.

JAMES GALVIN:

Thank you, Warren. Appreciate the reminder about all of that. It's part of what I meant when I said it's important for us to draw the appropriate line about where we are and what we know about what has happened over the last eight years and what's likely to happen next. Matt Larson, you're up next. Go ahead, please.

MATT LARSON:

Hi everyone. I just wanted to go at a little higher level here for a moment and talk about what we mean by gap. And I pasted into the chat a portion of the SOW that Karen is working from, and this is also verbatim text that was copied from RFP documentation that was available to everyone and that was reviewed by—I don't know if this group specifically saw it, if it was sent to this group, but the NCAP admin folks saw it.

So the point is this text isn't new, and this is what is defining gap for the purposes of study one. And as you can see in number four in particular, the gap term is being used specifically in the concept of datasets. And notice the little footnote in number four that then is defined down below, the idea being what we're trying to capture here is that study one is presumably in its exhaustive overview of NCAP research to date, it's identifying all the datasets that went into our current knowledge of name collisions. But we're contemplating these studies two and three which go further than previous work.

So the idea behind this gap analysis is to think, well, knowing what we know at a high level, that study two is supposed to research causes and study three is supposed to research mitigations, what sort of data could we imagine needing for those studies, and then therefore, what are the gaps between the data that we know exists because it's been referenced versus data we think we would need? And then to try to also assess at a high level the potential availability. Like let's say we decide, just as an example, that we needed recursive data from the largest recursive nameserver in the world, Google public DNS. Couldn't do it without that, but Google made it clear that that's not data that they're

willing to share, so we would have to make the conclusion that that particular dataset, though it might be desirable, is not available.

So that's just an example. I'm trying to get at what we meant when we said gaps. And the document we have that Matt Thomas has worked on looks to me—well, it specifically mentions datasets as it's number one. This looks to me more like defining gap in a much larger context, a much more expansive way to basically mean we've done a certain amount of name collision research to date, what are the gaps in the research that needs to happen? What other research would we like to do that would fill out and complete—if I can be so bold as to say “complete”—our knowledge of name collisions?

So while I can see this document that's been written as something that could be valuable going into planning for possible subsequent studies, I need to point out that it doesn't actually specifically address what we're using the term “gap” to mean in the context of study one. I'll pause there.

JAMES GALVIN:

Okay. Thank you, Matt. I appreciate that. I believe that I have a partial response to all of that, but I think I'm going to run out the queue here first and collect some other questions, let people take that onboard and see if we get any comments from other people. I do believe that it's possible to transition what's written here in a way that it reflects three databases and three datasets that we can look at. But we'll come back to that. Ram, you're up next. Go ahead, please.

RAM MOHAN:

Thanks, Jim. There are two things that I wanted to bring up. One is the suggestion of softening language around risk and danger. I agree that we should be prudent in labeling things in general, especially with loaded terms like danger. However, we should not shy away from calling out name collision or name collision issues for what it is. It's a risk and it's a risk management issue.

Stating that name collisions pose risks is not inflammatory. And in fact, stating that it poses risks is supported by the data we have so far. What I think we should be prudent about is in qualifying the risk and kind of calling it immediate danger or something like that. But I think it's hard to argue that it is not a risk issue.

The second point I wanted to make was someone said earlier on on the call that corp.com, that's a clear danger and something like that should have been caught. But I guess one of the questions really is whether we're going to undertake a study that has prevalence in second level or volume in second level and then correlation to the top level.

I guess I'm a bit skeptical as to how you would get metrics or measurement sets to identify in the new round what existing second level domains may pose a significant collision risk in a new round and then publish such a list or bring about the creation of such a list to say these sets of names may pose problems. I'm a bit skeptical about the science behind that. Thanks.

JAMES GALVIN:

Thanks, Ram. Appreciate that. Both excellent points. Jeff Neuman, you're up next.

JEFF NEUMAN:

Yeah. Thanks. For some reason, this isn't posting, but I wanted to post the board resolution that gave rise to all of these studies, because that's what people—yes, I know what the SOW states, and I think that's important, but it's also important to understand the board resolution and why people are looking to this group. And I cannot post it. I'm trying. Well, there's some of it. So yeah, it's better to post the link. Let me do that. Thanks, Matt.

And I agree with Matt too that we kind of go do more on this than identify gaps. And I want to point us as well not just to this board resolution but also to the study one report. Kathy's conclusion says it's not obvious that additional datasets would be needed for studies two and three. Information on previous and recent leakage of .corp, .home and .mail should already be captured in DITL and ORDINAL datasets. And then it goes on to talk about that.

So we're almost talking as if we're ignoring the conclusion in 5.2 of that phase one report, and we shouldn't be doing that, which brings it closer to Matt Larson's comment that there may not be any gaps. There may be areas of interest, but that's not the same thing as gaps. And while I appreciate Warren's statement of the absence of data does not mean the absence of a problem, but on the other side, the absence of data doesn't mean that people couldn't report the problem or didn't know to report the problem, or all the other things that are said.

And finally, on the board resolution, bringing it back to what Jeff Schmidt had said, the board resolution wants us to eventually—and

I don't know if this is studies two or three or both, but—talk about the harm to existing users that may occur if collision strings are to be delegated. That's number three.

Four, the possible course of action that might mitigate harm. Five, the factors that affect potential success. Six—and this I didn't see anywhere—potential residual risks of delegating collision strings even after taking actions to mitigate harm. So there's ongoing—do we have any dataset? Is there a dataset we could look at that would show ongoing harm from TLDs that have been delegated and are past the controlled interruption? I don't know how we'd do that or what kind of data that would be or how you get it.

So I think we need to go back to what Matt Larson said on the gap, which is in the statement of work, but also go back to the board resolution as well. And if there is going to be a study two and three, sort of needs to tie back to that. Thanks.

JAMES GALVIN: Thank you, Jeff Neuman. Let's go to Jeff Schmidt. You have your hand up. Go ahead, please.

JEFF SCHMIDT: Yeah. Thanks. I think maybe I was trying to think a little bit more carefully about what generated my first comment. I feel like a lot of what we're doing is kind of going back to the drawing board and saying, "Oh my goodness, what do we do about collisions?"

So let me propose an alternate kind of thought experiment, which is, what went wrong with what we've done in the last five years. So I would argue that—and there's a little bit of proud papa in this a little bit, so take that aside. I admit it. I would argue that controlled interruption did exactly what it was supposed to do, and we did collectively as a community—capital We—did the best we could at the time and delegated whatever the number was, 1400 strings with no material problems, and we found the three that were a problem and cut those off.

So I would love to take the contrary approach, like, what did we do wrong or what are we looking to improve on in the next round rather than, oh my gosh, what should we do? Thank you.

JAMES GALVIN:

Okay. Thanks very much. And Warren, you have your hand up. I'll let you go and then I want to try and tie all this together before we take anyone else out of the queue here if folks don't mind, use a little chair's prerogative. Warren, go ahead, please.

WARREN KUMARI:

I feel a little bit like this might be beating a dead horse, but someone just said that there was no material harm. And I will fully agree that the sky did not fall, but that's different to people did not see issues. And some of them had financial impact. And we see that if one looks through a bunch of stuff like Stack Overflow. There's a whole list in one of the documents on anecdotal evidence that people posted and did not report through the ICANN process.

So there was no material harm association that was just made. I think there was no major outage that was seen, and there was no material harm to some set of people, possibly including ICANN, but that's different to nobody suffered impact. As I say, this feels a little bit like beating a dead horse, but yeah.

JAMES GALVIN:

Okay. No problem. Thank you, Warren. In this particular case, it's okay to beat that dead horse just a little bit. All right, let me try to tie this together here and try to focus our discussions. I want to tease this out into three separate things. On the issue of gaps, thank you, Matt Larson, for being very specific about—and this of course is obviously his job, but the rest of us should all be onboard with this and understand it thoroughly ourselves. So I apologize for not being very precise about this. But the responsibility that he has and that Karen has is actually identifying datasets.

I would assert—and maybe we'll need to take some time to rewrite this to make this better representative, but I believe that there were three new datasets that exist now that did not exist in 2012 that need to be examined and that we should do some analysis on these datasets to see what they do and don't show, and document that and be clear about it within the ICANN context here. I'm not suggesting we need to do an academic, esoteric study, but we do need to do enough of a look into a set of things to do an analysis, the study two part of this project. And if folks are not up to speed on the project plan, you should go to the Wiki and get a copy of that and get ready and understand what study two is about. But we do need to do our analysis. There are a few datasets.

There is resolver data. I'm going to count as a dataset the collision reports that we have received which I'm calling anecdotal from an academic point of view because there aren't that many of them. But it's not really about volume in this case, it's about doing a root cause analysis. Let's at least learn what we can learn from what we have, such as it is, and see what that helps and how that informs our analysis.

A third dataset that I'm not sure how to characterize, and I don't have a nice pithy name for it, but it seems to me that there's something that we can review looking again at new root server data. And I think that the root server data that exists today is different than what existed in 2012 in that round. And this is important. This gets to the technological change that has happened over the last eight years, this gets to infrastructure changes that have happened over the last eight years. That means that root server data looks different. And it's important for us to see what exists in root server data and also to compare that to what exists in resolver datasets.

And why do we need these datasets? I count these as datasets, and I count them as a gap, and we need them for study two because separate from that, this discussion group and this project is going to continue whether or not the board funds anything. And I think that's an important point to bring out again here to remind people of. SSAC knows this very well. SSAC was asked by the board to answer some questions, and SSAC is going to do that. And we're doing that with this particular discussion group.

So whether or not the board funds any work or not doesn't change the fact that we continue and we do study two. we have to do an analysis

because the objective in our doing an analysis is to draw a line that we're all talking about here. Nothing really bad has happened. Well, that's great. I actually like Jeff Schmidt's idea about let's turn this around. If we accept that controlled interruption did what it was supposed to do, what can we improve on? Can we improve on it? What have we learned from that?

Our objective in study two is to get to a place where we can provide guidance to the board on the way in which it will be able to identify future .corp, .home and .mail if they ever come around again and what it should do about any name that happens to come forward that has any kind of evidence of collisions. Is there a way for the board to think about that so that it can still decide whether or not to delegate that string or if it has to maintain a list of collision strings? And then of course, we have to provide some guidance on how to manage that list of strings, how you get on and off the list.

So we have to do that anyway, and we're either going to do that analysis based on whatever's in the bibliography—and I would assert that we need to do that analysis based on what we've learned over the last eight years, so I'd identified three datasets that I think we can recap the document we have here in terms of them and support them, and we need that. And we need the opportunity, even if we are going to start over as Jeff Schmidt said, in a sense we are doing a little bit of that.

We should look at the data today as it was done then and make sure that everything is as it was, and then provide written, predictable guidance about what can be done in the future about delegating or not. That's our job.

The last thing is the guidance to the board. So we have the datasets to study the issues and the guidance to the board. Oh, and to study the issues, there are two parts to that. That's the way I wanted to do this.

There's the analysis of the state of Internet collisions, if you will, so the state of name collisions, and that's everything that I just said. But there's another piece of this that I didn't want to forget: study three is about mitigation strategies. That's where we get into the idea that maybe controlled interruption is the only answer, and maybe that's a conclusion that we want to come to as we think about it, but our real objective in study three is to consider the question, is it possible other mitigation strategies might be relevant? Is it possible that other mitigation strategies might be helpful to the presence of name collisions that might surface during controlled interruption and thus might allow delegation to happen anyway?

So we have to give some thought to that problem space and provide some advice to the board on how to evaluate mitigation strategies, if possible. We at least have to go through that work exercise of considering whether or not there's advice we can provide, or we're going to throw our hands up, "We need to document that," and everything that got us to that conclusion that there's no way to do that. But that's part of study three.

The board is thinking about the future. How does it manage the presence of name collisions such as they are going forward? And that's our job. And ideally, we get to do it with some fresh data because the world has changed and we should have that as part of our analysis, and

then we go forward and create the guidance that we need to, and then future rounds can continue from there.

So I apologize. I hope that didn't come across too—I'm trying to recap the way that the project plan that has existed for some time, that's sort of our original vision about where all of this goes, and we get to manage that in this group. So in many ways, I'm just sort of laying out as I understand this project going forward. But I do want to be very careful and emphasize to the group that to some extent, we get to manage our own path forward. As long as we answer and speak to the issues the board laid out for this, those ten issues, I'm fine with however we get down that path. That was just sort of my description of my expectation of what I [saw about] the project. Let me pause there if there's any hands, comments, questions, want to agree, disagree. If you want to disagree and you don't know what to say yet, you can say that too. Certainly have time to think about it.

Not seeing any hands. Let me catch up in the chat room here. I see references to risk management. I like that reference too. Yes. Speaking personally, I kind of view name collisions as a risk management problem, so I like what Greg Shatan is writing in the chat room there. Reading that out just for the recording about, "Could be characterized by severity, probability, types of harm, ease of fixing, losses caused by fixes, etc."

I agree with that, and I think at least for me, that's what I kind of get out of the way the board characterized its ten issues, if you will, that it wanted us to look at as it tried to dig in a little bit in that kind of space

and make it specific to this ICANN domain name collision space. Jeff Neuman, you have your hand up. Go ahead, please.

JEFF NEUMAN:

Thanks. And I agree with Greg and with you that, yes, this is about risk management. And in fact, that's what the subsequent procedures had in its initial report: what are the risks, and are there ways to identify the riskiest of TLDs in the future so we can move on with a round two and either have a test of riskiness—if that's even possible—or certain—I think Jeff Schmidt used the term “unicorns”—out there like .corp. .home, whatever else that may cause us to think a little bit more.

So I do agree with all that, but also, we should be tying it back to study one. So we had study one done. Study one did not indicate a huge amount of data out there that we didn't have. Now, comments are still due on study one, so there may be comments that come in with something that we just weren't thinking about, so we need to be open to that as well. But I think, Jim, what you said is a little bit broader when you were summarizing it. I don't think we should go out there and continue to look for new sources of data, except for those that are referenced in study one or potentially ones that we all agree upon that were potentially missed because of comments to that study one.

We all had a chance to read study one and we all had a chance to comment on it. I didn't get from our comments that there's a huge amount of data or much data at all that we don't already know about. So we need to tie it back to that study one as well. Thanks.

JAMES GALVIN: Thanks, Jeff. I'll make one quick comment about tying it back to study one. I think that the notion of asking [for] resolver data never existed in study one. So we can't be bound by the guard rails, if you will, that study one represents. We have to be bound by the problem space, and thus resolver data is a valid dataset to consider and need access to from a security point of view, just to make sure that there's nothing there, because that didn't even exist in 2012. So thank you.

Oh, I thought I saw another hand up there but it's gone now. [inaudible].

JEFF NEUMAN: Yeah, I was talking about this, our SOW, this study one that's out for comment.

JAMES GALVIN: [We] need to be bound by what's there. Well, again, I think that the purpose of the bibliography of study one was to gather up and collect what's there and then, yeah, similarly, I think I still stand behind my comment. I guess I would just phrase it differently and say that if no one really examined resolver data in anything that's there, it's up to use to recognize that that's a valid dataset to look at, as is the dataset of what we know about controlled interruption, such as it is.

We should definitely pull that data in and do a careful and thorough analysis of what's there so that we can see what effect it did have, and then as Jeff Schmidt phrased it, can we improve on that? maybe we can't. Maybe that's the best we can do. And that's fine, but let's at least

come to that conclusion in a scientific way, if you will, and in a methodical way so that we can draw a line under it. Greg Shatan, you have your hand up. Go ahead, please.

GREG SHATAN:

Thanks. I view study one as foundational and a very good, careful way to start a serious study to basically ask ourselves, what do we know, and what do we know now that we didn't know seven years ago or whatever prior benchmark? So it's a level setting. But I don't think it's a boundary. One wouldn't build a house by starting with the foundation and then move in. The point is to move from there. And in this case, even if we have only identified a limited number of new datasets such as with regards to resolvers, that's still important. We're not trying to build up a huge number of new data points, nor are we trying to discover a huge number of unicorns. It's fine that we have a few of each. I wouldn't say that's enough, but if it is what we discover we have, that's fine and it's nontrivial.

I don't want us to be like the head of the patent office who declared in 1903 that it could be closed since all things that were worthwhile had already been invented. We need to continue to be—if not skeptical, at least questioning. And even if we have no new data or only a few new major data piles, data is only the beginning. Data without analysis is nothing. We have to ask ourselves what is the proper analysis of the data and what have we learned about analyzing data and what tools weren't available just a few years ago.

Not that it's a direct example, but look at what Tableau is doing with taking coronavirus data and making it understandable to the general public through the magic of their data analysis engine. I don't want this to become a frolicking detour, but at the same time, we can't close the barn door before the horse has even woken up. Thanks.

JAMES GALVIN: Thanks, Greg. Good points. I agree with you there, and I'm going to read out something that Jeff wrote in the chat room here.

JEFF NEUMAN: I have my hand raised. If you want, I can speak to it.

JAMES GALVIN: Go ahead, Jeff. Please.

JEFF NEUMAN: Yeah. Thanks. So Greg, what you said made a total amount of sense and it was great, but that's not a reflection of what study one was. So if study one were just build a foundation and then leave it to everybody else to finish it, that would be one thing, but the role of study one was to build the foundation and recommend to all—I'd look at it more like a general contractor, and then the general contractor tells everybody else, "Okay, this is what we need when you go forward."

So your analogy doesn't work because it would be like each subcontractor going out on its own and doing whatever it wanted,

saying, “Thanks for the foundation but I'm just going to do this however I want.”

Study one asks a few questions. One of them is what data is out there, and then the second question—which is very important—is what data is needed to do studies two and three.

Study one clearly says that there's not more—that there are some, a couple potential sources, but there's not many, and I don't think we should go out there looking for yet additional sources because we had an expert draft study one, a person who's more of an expert than almost all of us, and I just don't think we should go beyond that and we should need to be narrow and we need to tie it back to study one. Thanks.

JAMES GALVIN:

Thank you, jeff, but I think the conclusion that we're getting to here is even if we say in broad, sweeping terms there were just two tasks in study one, what do we know and what new data can we have? I still want to assert that we have clearly identified—I'm going to call them three data sets that are gaps. These are data that we should go get so that we can continue and do our analysis work, and do the analysis with this new data. And I think we're still at that place and in that conclusion.

I've tried to characterized those three datasets. No one has actually objected to any of them. [Again, I'll speak to them] real quick as we come to the top of the hour and ending the call: resolver data, looking at the collision report data, doing a deep dive study of those, and then three, I do view root server data today as different than root server data

back in 2012, because the DNS infrastructure is a different thing today both technologically and of course, given that we have the first set of resolver data, it's just different. And therefore, it's useful for us to redo those studies just to make sure that nothing dramatic has changed as a result of the technology studies or the change in the DNS protocol and what's gone on there.

So no one is jumping up to comment on that, so I think what I'm going to do is say that what we'll do for next week is take some time to redraft and recast this analysis for you in terms of datasets so that we have something which directly is what Matt Larson is looking for, more directly what he's looking for so that he can make use of that. And we'll have an opportunity next week for that to be on our agenda and for us to review.

Jeff Neuman, I'm going to give you the last word and move immediately to Any Other Business if folks want to get ready for that. Go ahead, Jeff.

JEFF NEUMAN:

Thanks. So Jim, we probably should have said all of this to the contractor when she was drafting study one. I can't remember exactly when the comment period ends, but so the public knows or the community knows what we're thinking about, can you file or can we file a statement or comment during the comment period that says we think these three additional sets of data should also be included in the study one report so we can use them for study two?

I just don't want it to look like we're kind of going out on our own. It would be good to have a good public record on this so that it looks like

we're not just adding things despite what study one said. Does that make sense?

JAMES GALVIN:

It does. Thank you for that, jeff, for calling out the formality of how one does these things. You're absolutely right. Matt Larson, let me just call out your name here for the moment. I'm going to say some things and give you a chance to respond to this. We have actually in the admin committee been talking a little bit about what exactly is the process for this work product and making sure we get community view, [review.] Is this other task going to be a separate document? Is it something that'll just get folded into one work product?

We haven't made a firm decision on the exact details of that side of it, but it is fair to say that this test here and this document, whatever this is that gets produced, will be made available to the community for public comment. It's I part of the timeline that Matt Larson has laid out for delivery of the final package to the board, which is looking like that'll end up getting pushed into the July time frame I think is what we got to. I think that was the last thing that you had proposed, Matt Larson. And correct me if I get any of this wrong. But we are going to make all of this visible to the community. We're not trying to hide anything. And we will, in some way, make all of that happen exactly the way that you're talking about, Jeff. The details of it are just a little ... We're sort of in uncharted space here trying to figure out how to make this work.

Okay, so thank you for that, Jeff, and I'm not seeing Matt raise his hand, so I guess he's comfortable with all of that, which is good. Thank you for that.

Real quick, Any Other Business? I know we're at the top of the hour, I want to be respectful of people. Okay, not seeing any hands go up. Thanks, everyone. Great engagement today. Really appreciate it. Let's continue this as we go along. We'll have a revised document for next week so that we'll meet again and we'll go through this again. So we're adjourned.

UNIDENTIFIED FEMALE: Thanks all. Bye.

[END OF TRANSCRIPTION]