UNIDENTIFIED FEMALE:	Thank you. Welcome to today's NCAP Discussion Group call on 15 May 2019 at 21:00 UTC. In the interest of time, there will be no roll call. Attendance will be taken based on the Zoom list. We have apologies from Matt Larson, Jaap Akkerhuis, Merike Kaeo, and Danny McPherson. All calls are recorded and transcribed and will be published on the public Wiki.
	As a reminder, to avoid any background noise while others are speaking, please mute your phones and microphones. With that, I'll turn the call over to you, Jay. Thank you.
JAY DALEY:	Great. Jim, over to you.
JIM GALVIN:	Thanks, Jay. Jim Galvin here for the record. Welcome, everyone. Glad you could be here. I'm just going to do a couple of administrative things here up front and then we'll dig into the meat here with Jay. The first question is just to ask everyone if anyone has any update to their SOI that they wish to speak to and alert folks to here. I'm not hearing anything. This is good. We don't have any new members this week. We normally track this, so we already know the answer to that question. We normally ask folks when they're new to introduce themselves the first time that they're on the call. Actually, Jay, from last week, since I missed last week, is there anyone? I'm seeing a few people here. Is there anyone here who has not been on a call before who is one of our

discussion group members? Not hearing anything, then we'll assume we're all set there. So, let's jump into the work items that we had for this week. Jay, we'll go back over to you for item #4 on the agenda.

JAY DALEY: Great. Thank you. I'll just bring this up. Right, everyone. So, this is the document we have been working on for study one. What we're going to do now is just go through the final changes that we have put into these so that this can then be handed over to Matt Larson and OCTO for them to then put through the procurement process. Hopefully, everyone is comfortable with that. I'm just going to make sure I've got the chat up as well in case anyone needs to interrupt.

JIM GALVIN: I'll keep an eye on the chat, Jay.

JAY DALEY: Okay, great. Thank you. First of all, we're going to announce task one. Task one was originally properly defined name collision. This has now been changed to finalize the definition of name collision. And then what's listed below is the definition of name collision that the SSAC working party produced before the discussion group setup which we discussed about two or three meetings ago in the discussion group.

So, I don't plan to go through it all but it's got the things that are in scope and subject to data studies, things that are in scope but will be addressed with general advice not data studies, and the things that are out of scope.

The reason that we are not finalizing this is because we want to just ensure that the contractor has a chance just to think through any additional things or check [inaudible] with this because we've had one piece of feedback from someone that requires a little bit of adjustment to this which is about unintended consequences of broken pieces of equipment and they were able to provide a specific example which we don't think is then captured in there.

So, do I have any comments or questions about study task one within study one at all? No hands noticed. Great. So, moving on then.

Study task two has been slightly amended to say undertaken in formal public consultation of the finalized definition of name collision. The reason it says informal, because if it didn't say that, then a public comment would be required and that is very lengthy and detail process and that's unnecessary [inaudible]. This has also already been partly through a public comment process.

So, then onto study task three, which is to review and analyze past studies and work on name collision. This is the criteria and the specific prior work that we have discussed previously and has been circulated on the list and agreed. Is everybody comfortable with that still? Good.

JIM GALVIN: Do you want to scroll down on the screen to show the rest of four there?

JAY DALEY: Yes. I'm coming to four. So, that's only three. I'm just checking there. So, four, then. Four is producing a written report from the reviewed material. That did say solely from the reviewed material but Jim and I discussed it and I removed the word solely, as me being a little bit too tight on things.

The reason for this specific point there is that we don't want the contractor going beyond the criteria or other things. We've had a relatively good consensus decision about that, those criteria, and we want to be sure that's followed.

So, the written report should, first of all, provide an explanation of the issue. I think we're all comfortable with that. And then these are the bits that we've added following last week's conversation. Summarizes the known, in brackets, evidenced harm of name collisions. Any comments or questions about that one at all? No? Moving on.

List all the relevant previous work on the subject using the criteria in task three. Looks good.

Then, D is document any mitigations/actions taken so far, specifically including controlled interruption and the technical impact of those mitigations only. Then with huge brackets, no examination to be undertaken of the non-technical impacts such as [resource at all costs]. Rather than explain this, does anyone have any questions or comments on D here? No? Good.

Then, E is the work that Jim added in which includes any important points that should be brought forward for this project. Important points include, but are not limited to, questions about the data use,

methodology applied, any technical gaps that should be considered and any competitive or opposing recommendations that may be identified. Is everyone comfortable with that text as well or have any other things to say on that?

Then, six is identify data sets used in past studied and determine if those data sets are still available and any constraints there may be regarding access. Any comments on that?

So, overall then, for study one for this tasks, is there anything that anybody feels is missing or is there anything that I haven't covered, anyone feels should be changed at all?

JIM GALVIN: Jay, a couple of things there in the chatroom. Starting with Ram's comment at the top here, raises an interesting point about the informal council, information public consultation. I think that the way that I would characterize that … So, going back up to item two, informal public consultation of the finalized definition, it seems that – and I agree with this point – we want to be careful not to restrict ourselves such that if we learn things along the way here, we might still want to come back and update things and revise things.

We've got the use of the word "finalize" there, but lets be careful not to restrict ourselves to not being able to move this forward after we've done studies two and three.

JAY DALEY: Okay. And revise if required for this or do we want to have a look at the … I'm a bit concerned that the suggestion is that we would revise in study two or study three the definition. Is that the intent or is it simply that we should be able to revise it as a result of the public consultation?

JIM GALVIN: Well, that's what I would allow for, but we've got a hand up here. Why don't we see what Eric wants to suggest?

JAY DALEY: Yes. Please, Eric, go ahead.

ERIC OSTERWEIL: Just as a suggestion, I think if the definition is treated as being able to be evolved but with the understanding that in order to evolve the definition, there needs to be a clear justification. That would be kind of in keeping with the scientific rigor. You come up with something based on evidence and it's a theory and theories can only ever be disproven which means they evolve.

My personal independent opinion is I don't think we should worry too much about down the road in study two, someone might come up with a reason to change the definition. The definition as a theory may always need to be held up for scrutiny and evolution would be my two cents.

JAY DALEY: Right. Thank you, Eric. Does anybody want to suggest a different word than finalized? Is that part of the issue? Do we need to adjust that?

JIM GALVIN: I think, from my point of view, I'll take what Eric said and just go with that. I think that our goal here is following some scientific rigor and I think that, as Eric said, that allows us to evolve things if evidence surfaces. Then the right thing will happen. We'll take that evidence and we'll act on it. So, I don't know that we have to do anything in particular here, just acknowledge to ourselves we're going to follow ordinary process in that respect.

JAY DALEY: Yeah. So, there are already a number of things that have appeared in the last few meetings that are push-forward things, so I think at some point I need to make sure I've documented those properly so that we have those recorded so that we can remember that and that's one of those push-forward things. Okay, good. Does anybody else have anything else they'd like to add to this particular, to this study one …

JIM GALVIN: I just want to call out two other comments, then, from the chatroom here. Rubens and Steve Crocker are both identifying some other work that's out there. Rubens is talking about data sets. Do we want to call out the data sets particular? Steve Crocker is reminding us about the site finder work and other kinds of things. I guess my response to those is just that I don't think we need to be complete here. Part of the goal of

study one is for them to make sure that they go find these things. If people want to suggest particular words, I guess we can consider that but those things are certainly not out of scope and I would expect our contractor to make sure to include those things.

JAY DALEY: Okay. Rubens and Steve, are you comfortable with that response from Jim?

JIM GALVIN: We also have a spot on our Wiki page. We have a spot where we're collecting references, right?

JAY DALEY: Okay. Yeah. Sorry. I've just asked a question. Steve, you were going to reply?

STEVE CROCKER: I guess so. I remember vividly the reactions to site finder and it included some very pointed and specific collections of data and demonstrations of the kinds of things that go wrong when people get back responses that they weren't expecting. I don't have any strong feelings about whether you include it in the wording here or make it available later or whatever. I just was trying to be helpful at saying that would be one of the things that I would hope that they would look at it. What they do with it after they look at it, that depends on everything else they're looking at.

JAY DALEY: Yeah. I'd like a little conversation about this, actually, because I'm inclined to specifically include it because it is so old that it may be forgotten about otherwise. Is there anybody who would object to any of that being included?

JIM GALVIN: Jay, by included do you mean listing it here or just that it gets put on the Wiki so it's there?

JAY DALEY: Listing it here under B. No objection from anyone to that then? Okay. I think we will put in there an analysis of site finder that meets the criteria above. Yeah?

UNIDENTIFIED MALE: I would imagine Jim is in the optimum position to dredge that up.

JIM GALVIN: Sadly, yes.

JAY DALEY: Analysis of the impact of site finder.

JIM GALVIN: And you should look at the bottom of the chatroom there, Jay. Rubens clarified his comment about data sets.

JAY DALEY: Yeah. I'm unable to see the chat while I'm sharing this, actually. I think I've lost the window and hidden it and therefore can't get it back.

JIM GALVIN: He says his point is not to mention DIDL but to mention that some studies used DIDL 2012 and the data sets available might be newer additions of those.

JAY DALEY: Yeah. [inaudible]. That's an admin point, I think. Yeah. Okay, good.

JIM GALVIN: One last thing here. There's one last comment about Ram clarifying his comment about information public comment. Let me read this out here, too. "Let's be careful to define "informal" public consultation. ICANN has certain norms for public consultation and I would not want our informal process to be pointed out as a reason to doubt the outcome."

JAY DALEY: That's fine. I think this is going to be managed by OCTO. It's not going to be managed by us. I'm sure OCTO will be doing that within the correct ICANN framework for these things. I don't think it's appropriate for us to tell OCTO to do those things within that framework.

JIM GALVIN: Well, I don't want to overly put words in Ram's mouth here. He'll respond in chat if he wants something different. I think the point that he's making here is to use the word public consultation, even if you stick informal in front of it, public consultation means something to ICANN and this is not just about what OCTO might do. We're going to get called out on this kind of stuff. So, I think that we need to be a little … We should be a little more deliberate here about what we mean by informal. Otherwise, we're going to get stuck with what ICANN expects it to be.

JAY DALEY: Right. Can somebody put some wording here? So, Rubens suggested appropriate nature public consultation. I think we just put in public consultation as decided by OCTO or something like that and leave it at that if everyone is comfortable with that. I think this is a bit of a [inaudible].

JIM GALVIN: Ram says that's fine with me.

JAY DALEY: Okay. I'd like to move on. This one is a little bit of a [inaudible].

JIM GALVIN: Yeah. I guess it's fine. I'll just observe it still is … To me, it's still a little touchy because OCTO, in still being the ICANN Org, I feel that they're probably still going to want to stick with ordinary ICANN processes. They're not going to want to just invent something new. That's the problem here.

JAY DALEY: Well, okay. So, I'll just put in a public consultation. I'll just leave it. Remove the word "and informal" [and left there] public consultation. OCTO can decide what that means later, I think.

JIM GALVIN: Fair enough.

JAY DALEY: Okay, great. Can we have the agenda back up, please? Going to move on from that one now. Marvelous.

So, we're onto item five now which is planning for study two. This is some little bit of fun stuff for you. We have some questions down. [inaudible] ICANN meeting. Item five and study one tasks. There is a requirement in study one tasks about presentation of the report. Right.

So, moving on to item five, what we are looking at here is this is just a general workshop. This is beginning of a conversation that we're going to have now about what level of … Well, I'll take it back – sorry – and start this again.

As part of the project plan, we are probably going to be commissioning a third-party to develop some form of tool that takes raw DNS data and anonymizes it in some way that the end result is still usable for the purposes of the data analysis but certain privacy concerns that people have are dealt with through that process.

What we are considering is the production of a tool that will do that on data sets that are given to us or that people can run themselves before they give the data sets to us. Steve has taken his hand down.

So, this is a conversation for us to have about what we think the requirements of that tool are, what we want to see from the data at the end of it. I'm happy for people just to throw that out and let's start recording those things so we can start working this up over the next few meetings or so. Anybody have anything they'd like to start with on this one? This is a technical conversation. Any technical people like to throw in about what they would like to see? Okay, I'm going to have to start making this stuff up, then.

Some of the criteria that we have talked about, some of the things that we want to see—

JIM GALVIN: Jay, we lost you. I hope you can hear us, Jay. We're not hearing you at all.

JAY DALEY: Right. Okay. Sorry. I pressed the wrong button on that. So, the first thing I think that we would like to see is complete anonymization or sources

of traffic, so that sources cannot be re-identified so that we do not know who was making a specific query. Is that something that people generally agree with? Steve, go ahead.

STEVE CROCKER: I'm trying to get my head around exactly what we're trying to accomplish with this. This is a way of protecting the privacy of the data and still having it be useful I guess is the overarching statement.

So, let me recommend two approaches. One is, yes, let's do indeed do some anonymization and so forth and we can have the discussion about how to do that. At the same time, I think all of us know that in the extreme this is a really, really, really hard problem if you're going to depend upon that completely because you could do a hell of a lot of computing and untangle the anonymization by looking at different things.

I think the right thing for us is to say, hey look, let's do all of this because it's good practice and so forth. But on top of that, let's treat the anonymized data as sensitive in its own right and protect it just as if it was the real stuff and that layer of protection is sensible and appropriate and will keep it from being subjected to [inaudible] cryptoanalysis or [inaudible] analysis or whatever people want to do. I see Rod twitching there.

JAY DALEY: That's very useful, Steve. Thank you. So, that is similar to the view that we had discussing this so far, that we need to have the data, once

anonymized, doesn't mean that it can then be published, that there is then a separate process or methodology required for publication which is where we're going to come on to the reproducibility which is the only reason we think that may be required.

So, Eric has his hand up. And Rubens, it would be very useful if you could make the comments not just in the chatroom because I can't keep up with two different streams that well. I'm going to ask Eric to go first and then I'm going to come to you, Rubens, and explain what you've written. Eric, over to you first.

ERIC OSTERWEIL: Yeah. In line with what Steve was saying, potentially there's an ocean here and a Bunsen burner. I think we want to be really careful. I think there's lots of different types of analyses that may or may not wind up being the things that the contractor wants to do in study two. It's probably really hard to figure out the optimal way to do general purpose anonymization without knowing what the analyses are going to be.

So, I think it may wind up being difficult to do all this ahead of time. Certainly, having the conversations is a good start now. But one thing that also might be worth considering is something that I think I heard, implied by what you said at the end there which is if someone is going to come through and come up with a set of analyses that they find are helpful for study two, it's possible at that point to figure out what the sensitive data is and a proper way to anonymize it. For example, if you have IP addresses that you don't want to disclose but part of the

analysis is to look at [collated] IP addresses in a prefix, whatever your anonymization technique is, it needs to preserve the locality of those IP addresses. You can't just do a [inaudible] 256 hash of them because IP addresses next to each other will be very far apart from each other at that point. It will be anonymous, but it won't preserve the property that you may be wanting to measure.

On the other hand, if someone comes in and they do a measurement study and the goal is to have that study reproducible, you could then come up with an anonymization technique after the analysis is done that preserves anonymity and also preserves the fidelity of the measurement. That one is unfortunate in the sense that you kind of have to do the measurement first to know, but you could say that the contractor is going to get access to sensitive data and we're all going to agree on the proper way to anonymize it while still having reproducibility. I'd hate to kick the can down the road but that may be important.

JAY DALEY: That's very useful. Now, the reason that we are looking at this is because we've been informed by some data providers that they're not going to give us raw data but that if there is a process that they can give us some form of redacted data available, then they will follow that process and give us that. So, that's why we're trying to pre-guess it. But as you said, perhaps it's useful for, in study two, for someone to take some readily available data, begin to play what form of analyses they might do on that and then understand the form of anonymization,

redaction that's required for them to move on to those people that have more sensitive data to be provided. Steve, back to you.

STEVE CROCKER: Thank you. Just to take this one step further, there's something that's been bothering me about this. It is this. No matter anonymization you do up front, it may turn out to be, in keeping with the discussion we had, something that interferes with the analysis that you wanted to do, so you're going to find yourself in a position of saying, gee, that was reasonable to anonymize it that way, but that isn't going to serve our needs, so please go back and re-anonymize it and here's some stronger criteria to adhere to.

So, access to the original data, not necessarily by us, not necessarily by the contractor, but by the people who provide the data, they may have to provide it yet again under some agreed-upon new anonymization process. That might turn out to be [inaudible].

Hopefully, if that's the case, it would not be a huge deal. That is, the anonymization should be a relatively strict forward pass through the data after one sees what the issues are and agrees upon the new algorithm.

JAY DALEY: That's very useful. Rubens, I'm going to read your stuff out then, given the toddlers and toys issues in the background there. I only have a quiet sleeping dog in the background there, so I'm very happy. Rubens has written, "I would imagine changing IP addresses to ASN if ASN is ISP but

replacing it by end user for all end user ASN [original] IP address, but not destination IP or the query itself." Effectively, that's the locality preservation that Eric has been talking about.

Also within that, I think we are … The end user query isn't changed is something I would challenge there. We don't necessarily need all of the labels of a query. We potentially need to know that all of the labels are the same, but actually we only need to know the collision labels of a query and that makes something quite different there.

Matthew, you have your hand up. Go ahead, please.

MATTHEW: I apologize if my audio is awful. But I would [inaudible] that there are very many times where the contextual awareness of all labels, not just the collision TLD itself, gives the extra insights that provides avenues for additional research or contextualizing what is going on, what is causing the actual leakage or the suffix [inaudible] list or whatever the underpinning is. I think keeping all of those labels is something that is very important for this.

JAY DALEY: Okay, thank you, Matthew. So, an example you mean is if there's an underscore TCP dot underscore, whatever it is, so there's that type of thing for an SLP record, then we would need those labels in order to be able to understand the nature of the [SLV] query.

MATTHEW:	Exactly. It's that kind of context that puts the leakage into more taxonomy that you can understand.
JAY DALEY:	Okay. So, what we're potentially talking about in terms of redaction of labels is redaction of private, non-collision labels but we still want any standard labels or labels that help identify the product or system that may be generating the collision. Eric, you have your hand up.
ERIC OSTERWEIL:	Yeah. I think Matt is spot on. I think, just to further underscore it, when we get to the point I'm talking about, like search list processing, at that point it's very hard to know which label to start looking at because the search list gets appended and, depending on how big the search list, how long the search list is, you're not sure how many labels to chop off, especially if you're looking for something that was prepended down lower. So, it gets really complicated.
	On top of that, this might be part of – a good explanation for why trying to figure this out before we know what the analyses we're going to do is could wind up really hampering the consultant down the road. For example, if we said it's just the top two labels and [we went from there] and then someone said, "How can you do search list processing that's in the RFP?" then we'd sort of put in place a redaction technique that makes the measurement that we've asked for impossible because we were trying to be very proactive.

So, I'm not sure I have an answer but I just wanted to point out that I think we should be really careful here.

JAY DALEY: Right. Just to be clear, the purpose of this conversation is to tease these issues out, not to come to a decision and not to try to define an anonymization routine at the end of this. This is a really very helpful conversation. Has anybody got any other particular elements here about this that anyone would like to pull out at all?

Some of the dimensions, we've already talked about specific labels that identify things are useful for us. We've talked about locality. Are there any other bits and pieces that we think are important here? Do we need to know when something was done, whether it was done in 2018 or 2016? Does time dimension matter to us? I have a yes from Eric. And Matthew, you have your hand up again.

MATTHEW: Yes. I was just going to agree. We should absolutely include the time stamp in the contextual awareness of also where that query was destined for. I think that would be important as well, especially if we're talking about various different levels of the DNS hierarchy or if we're only looking at root servers. That gives you a different understanding of what might be going on.

JAY DALEY: Okay, excellent. Thank you. Would anyone like to contribute anything else to this conversation so far? Okay. The way I can see this going is

that study two is going to require the contractor to do some pre-analysis of existing data sets or data sets we can get a hold of easily in order to then come up with some thoughts about anonymization in order to get some of the harder-to-get data sets. I think it's quite straightforward. Okay, good.

So, the second bullet point here is about reproducibility. It should be straight obvious, but if it is isn't, that is at the end of this project, what is the ability we're going to provide for any third-party researcher to be able to reproduce the research that has been undertaken by the contractors using the same data?

Does anybody have any thoughts about this, about what level we should be aiming for or we should be targeting? Just to pre-note that in the proposal that has gone to the board on this, we have actually stated in this that we don't think this can be guaranteed. It's only something that we can target. But I still would very much like us to have the conversation. Steve, you have your hand up.

STEVE CROCKER: Yeah. I like your last comment, because in line with what I was saying before, making the data available, even anonymized, to some third party opens the door to possibly piercing the vale. It's a challenge. You really do want to make it reproducible. You really do want to show how this was arrived at and make it as easy as possible for people to check and satisfy themselves about what was done.

On the other hand, if you're really trying to protect the privacy of the data, you've got a challenge there. If I heard your words, I wouldn't

have raised my hand. It's not a guarantee that you're going to be able to accomplish the reproducibility and the protection of the privacy as straightforwardly as it might seem by just putting down as an extra bullet point.

JAY DALEY: Thank you. What we had discussed previously in the SSAC was that we may have some form of research panel that takes – sorry, a panel that has specific proposals put to it for reproducibility and that those are agreed as one-off terms with one-off sets of contracts and things with those people who wish to do the reproducibility. Is that something that we think that could work in this context or do we think that … Anyone have any concerns about that?

STEVE CROCKER: Let me just chime in. Basically, you'd want the people who are going to do that have access to that data to adhere to the same rules of protection that the contractor is going to adhere to. That is, even though it's been anonymized, treat it as sensitive data and be responsible for holding that in confidence, etc.

JAY DALEY: Yeah. The other thing we had talked about … Sorry, does anyone else want to say anything about that, about the idea that potentially, at the end of this, we can allow reproducibility through some form of panel that analyzes each individual specific request?

Okay. The other thing we had talked about was whether the data providers themselves, if they're ones that have specific concerns about this, should be allowed a veto on individual reproducibility proposals. Does anyone have a view on that?

STEVE CROCKER: [inaudible] raise my hand.

JAY DALEY: Go ahead, Steve.

STEVE CROCKER: Yeah. It was a conflict of interest issue that comes up. I doubt that it will happen for real but I can easily imagine somebody would raise it as a matter of form. If you give the data provider veto power, then that raises the question of whether or not they have the either purposefully or incidentally would be biasing the results and steering things in one direction or another. Particularly, if they can preclude some other party from getting access to the data, then it's a question of whose interests are being protected. I don't want to go too far down that rat hole but it ought to be … You're going to have judgment about who can have access to it. You want to take into account the data provider's wishes but you don't want to give them absolute control.

JAY DALEY: Okay, great. Thank you. Anybody else like to comment on a specific point? No? Okay. So, the final question on this one of reproducibility is

perhaps more of – not such a policy question as perhaps a perceptual question is: if we are unable to provide any reproducibility as a result of the level of protection we put into the data of this, what impact do we think that would have on how people accept the results of this project? Are there any other mitigations that we need to put in place should that be the case? Any thoughts anyone have on any of this? Everybody is required to have at least one thought per meeting. Eric, you have some thoughts. Go on, please. Thank you.

ERIC OSTERWEIL: Yeah. I can't help myself. I'll take the bait. I think one thing we could think about – and this is very high level and wish-washy, but I want to make sure I make my thought quota this week – is consternation over the results, somebody that takes exception, it could be that … It might be useful to flip it around and say anyone is certainly entitled to take exception but that there should be some kind of evidence behind what's prompting the exception. In other words, if the data is not transparent and reproducible and someone calls shenanigans, it should be more than just calling shenanigans. It should be shenanigans because here is some data that refutes it or something. It's just a thought. I'm not used to having that level of concern but maybe that's one way to put it.

JAY DALEY: That's very helpful. Thank you. Anybody else like to comment on this point at all? No? Great. So, we're going to move forward now to number six. Number six is simply I'll throw it open to people who want to say here is a specific type of data set I think we need or here's a

specific data set we ought to talk to someone about or anything like that. It's simply a free-for-all for us to make sure that we've got some understanding of the coverage of the range of data sets that people think would be useful. Eric, I'm going to assume that's an old hand. Would anyone like to start off with throwing some stuff out here about what range of data sets they'd like to see made available? Matthew, go ahead, please.

MATTHEW: I think the obvious answer here would be root data, obviously, and then all data available through the DNS OARC products and their [inaudible] and their one-offs would be an excellent starting point.

JAY DALEY: Right. So, we've got the OARC stuff. Now, when you talk about the root data, how many root servers? How much from root servers would you say?

MATTHEW: One of the papers we presented at the London conference did an analysis looking at when you look at the name collision data only using subsets of root data. Your results can vary drastically. Of course, as a researcher, I'm going to say the more, the better. I don't think there's a hard and fast line I want to say here in terms of meet X or Y, but …

JAY DALEY: Okay, so just to be clear, we do actually need to ask for multiple roots, as many roots as we can get because you think there's a good case that just using one or two roots is insufficient.

MATTHEW: Yeah. At least we presented some data that showed root [affinity]. I'm including more data here, you're obviously going to get a better perception of the world [inaudible] collisions that might be in a specific contextual or geographic area.

JAY DALEY: Right. That's very, very useful. Thank you. I think that's something we can start to work on early, then, about trying to get as much of that data out as possible. Let's talk about … Dmitri, you have your hand up. Go ahead.

[DMITRI BURLOVSKI]:   I think when [inaudible] request some data from major public resolvers, well, I think that Cloudflare, Google, and [inaudible] will be enough. Thank you.

JAY DALEY: Right. Now, a question on that for you, Dmitri. Do we think that if we got it just from one of them – say, Google – do you think that would be sufficient or do you think again there's the possibility of some affinity there that means that we need to get it from multiple?

[DMITRI BULOVSKI]: Well, I think that it makes sense to get this data, especially from Cloudflare because using the [inaudible] technology in [inaudible] which is set up directly to Cloudflare.

JAY DALEY: Okay. Right. Anyone else like to comment on this about major caching resolvers at all? Rod?

ROD RASMUSSEN: Yeah. I think you actually changed it already. I was going to point it out. I think Dmitri mentioned public resolvers which would be OpenDNS, Google, and those folks. There's also ISPs which would not necessarily be open. They have large ISPs that may be willing to share some of their data as well. I think your large caching resolvers covers both those categories and potentially others.

JAY DALEY: Right. Rod, can you – any idea of the sort of scale of ISP or the numbers or anything like that, do you think?

ROD RASMUSSEN: Well, if you take a look at large consumer providers, you're going to have pretty large data sets to look at potentially. It really depends on what's [inaudible]. Then they get into things like sampling and the like, so you may or may not have "full visibility". Then there's what is the

customer base of that ISP. A lot of ISP data would be fairly skewed either towards a home maybe SMB user versus somebody who may provide enterprise level type transactions.

One interesting bit which I actually need to do some exploration around is some of the providers [inaudible] info blocks has customer data available because of work that they do for them where the data is shipped off. There are a few players that might be able to have that to consolidate different types of data which may or may not be willing to share, obviously, but that would be interesting because it gives you different types of data, especially if you're looking for some of the collisions that were suspicious of happening on corporate environments that we wouldn't necessarily see in a public ISP environment.

JAY DALEY: Right. Thank you. Is there a generic type of data there from security companies who hold data? Is that another one as well we should be looking at?

ROD RASMUSSEN: Yes, definitely. [inaudible] people running DNS firewall type products and things like that.

JAY DALEY: Right. Okay, fantastic. Matthew, you have your hand up.

[MATTHEW]: Yes. I see Rubens made a comment in the chat about [inaudible] and that made me think of something, that we should also bring in longitudinal data going back not just current data, if possible, since so many protocol changes to the DNS are impairing our observations [inaudible] to the root, things like [queue name minimization, root on loop] [inaudible], NX domain cut, those things are going to influence maybe some of the studies, so being able to quantify over time how those deployed and are changing, our measurements for these studies could be very helpful.

JAY DALEY: Okay. I'm just making a note of that. Wonderful. Does anyone else have anything else they would like to add to this? No? Great. Okay. Thank you for that. That's the end of item six. Jim, over to you.

JIM GALVIN: Thanks much. Okay. This is the point when we get to ask any other business, if anyone has anything else that they want to add. I will make a comment while I'm waiting for all of you to jump on your hands and put up your note and ask something. To comment about our Marrakech plans, we are meeting weekly at the moment but we were just talking earlier, Jay and I, and I think that we're not doing any urgent work at the moment, so it seems appropriate to not meet during the week of the ICANN meeting. I just want to alert folks to that now. I realize it seems like a month away is such a long time but I thought it would be useful to let people in on that, so you can plan your schedules accordingly.

We, also, this time won't meet on the following Wednesday which would be July 3rd. That does have the potential to be a significant US holiday. But in addition, SSAC doesn't normally hold any of its meetings the week after an ICANN meeting, so we're for the moment just following in that tradition. So, that's two weeks during which we won't meet. Otherwise, we're going to continue our weekly meetings. As it says there on item eight on the agenda, our next meeting will be next week at this same time on Wednesdays. So, one last call for any other business from anyone.

JAY DALEY: No. Could I just say for the next meeting if anybody has any specific things they would like to discuss, please let us know. We're potentially coming to a bit of a hiatus while we have the work going into study one, so it would be useful for us to see whether we actually need these meetings or whether we can reduce the frequency or something like that. We can still sure think of interesting things to talk about but we'll need to make sure it's productive. Jim, you have a question in the chat.

JIM GALVIN: Yeah. I'll speak to that. Thanks for your comment, Jay, because [inaudible] to everyone to keep that in mind. Maybe our discussion point next week might be to decide to take a hiatus until we actually start to get some real work coming from a contractor here.

In fact, it's part of the response to you. The question in the chatroom is: doesn't our schedule call for an all-day meeting in front of an ICANN meeting? Yes. The project plan actually says one of the ways in which

we were seeking to … It does call for face-to-face meetings of this particular discussion group and the project proposal is to have those the day before each ICANN meeting.

However, again, we're not prepared at this time to have that meeting for this particular ICANN meeting. It's just a timing thing. We kind of figured that we wouldn't have a contractor on board yet and it doesn't look like we're likely to have that at this point, so we're not likely to have any real work or even any preliminary results out of study one.

However, folks should be thinking in terms of the fall meeting in Montreal, assuming everything progresses here and we do actually get study one underway and we get some work product out of a contractor. I would expect that this discussion group will be in full swing and we'll be looking to have a face-to-face meeting at the next ICANN meeting at the end of the year, the Montreal meeting at that time, and it'll be a one-day meeting in front of the ICANN meeting, so folks should please make a note of that as you're planning your travel to keep that in mind that that's a strong potential.

With that, if there's nothing else from anyone, then thank you very much. We're adjourned. See you all next week.

JAY DALEY: Take care, everyone. Bye.

UNIDENTIFIED FEMALE: Thank you.

UNIDENTIFIED MALE: Thank you.

[END OF TRANSCRIPTION]