KIMBERLY CARLSON: Hi all, and welcome to today's NCAP Discussion Group call on August 5th at 19:00 UTC. In the interest of time, there will be no role call and we'll take role in our usual fashion based on those on Zoom. I will update the Wiki with the names of the participants as quickly as possible. No apologies have been received. All calls are recorded and transcribed, and the recordings, transcripts will be published on the public Wiki. Also, as a reminder, to avoid background noise and echoing, please mute you phones and microphones when others are speaking.

And, with that, I'll turn the call back over to you, Jim. Thank you.

JIM GALVIN: Thanks very much, Kim. This is Jim Galvin from Afilias for the record, one of our three co-chairs for this august group. So, thanks everyone for being here. I guess we have to do the quick question about is there anyone who wants to say anything about an updated SOI today? I'm not seeing any hands, moving along smartly, and we don't have any new members, so all of that is good.

So, the purpose of today's meeting … Oh, we have a hand. Please, Steve, go ahead.

STEVE CROCKER: Very minor, but my status is changing. I'm becoming a member of SSAC, which is, in some ways, entangled with us.

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.

JIM GALVIN:	Yes, it is. Thank you very much for that. Appreciate that information. And yes, SSAC is certainly entangled with this group.

So, with that, the purpose of today's call is primarily to bring folks up to date. As you know, we skipped over last week's call because the NCAP Admin Committee took some time really to regroup and to figure out where we are and what we're going to be doing and how to approach the work going forward.

So, I'm going to just walk through a little bit about what we've said to ourselves and what our proposal is going to be. And we're taking as an action to build something out here, and I'm just going to walk through what that is, get whatever reactions people have for that today, of course. And before anything moves past the discussion group, folks will certainly have a time and opportunity to review any work, but it just seems like the most expeditious thing at the moment is for the NCAP Admin Committee to take the pen and the action to move things along.

So, with that, if you can, Kim, bring up the Study 2 link, that Google Doc that's there. Folks should have that. And if we scroll down here to Section 3.3.2, just another quick look, we did make a couple of minor changes here into the outline that's there. And I'll just sort of briefly go over this outline again.

So, as a little bit of intro, the approach that we're going to take here is we accept the advice from the contractor and from OCTO with respect to the report, with Work Product One, and the recommendation that the work not proceed as it's currently documented and currently proposed. And we agreed with that.

I think that, even for ourselves, we produced the technical gap briefing document and that clearly highlights that there are differences. The infrastructure or the Internet has changed since this original project proposal three years ago. As well as the bibliography that was Work Product One demonstrates that some amount of work has been done and we do have a little more knowledge, a little more information, about name collisions, and certainly some experience over the last eight years.

So, it seems appropriate to think a little bit about what we really want to do and how to do that. So, that's one bit of our approach towards this.

The second thing is that I don't think that we need to make major changes to this Study Two proposal. We're going to focus on Study Two, we're not going to kind of focus on Study Three at the moment. So, we'll take this into Study Two.

We don't think that we need to make major changes to the overall set of tasks. In fact, there's really only two major changes, if there's any significant changes, to the overall tasks that were listed here. But what we've done is just build out a little bit further some of our thoughts about what's currently Task Four there, the impact analysis. I think it really is very much appropriate for us to think carefully about what the details of the impact analysis really should be, or could be, and that's what we need to expand on.

So, the details of what that was were not really in the original project plan. Clearly what that needs to be is a little bit different than the vision that any of us really had three years ago.

So, let's take some time now to think about what that could be, and that's what we want to do here. We've already talked about this, so we're just going to go with it since there did not seem to be any objections thus far in the discussion group.

So, we're going to go with the six tasks that are here. Actually, it's five tasks, because what's not shown here in the document … We're still going to conduct a root cause analysis. That step amounts to going through the name collision events that we do know that occurred. I mean, there are a small handful of them, and ICANN does have knowledge of what those things were, and we would still seek to want to dig into those and see what we can learn from them.

I realize it is, again, just a small handful, but it does seem appropriate to pay attention to what really happened. Is there anything we can learn from that that might suggest that there's more that can be done in terms of messaging down the road. I mean, is there more that can be done even from a technology point of view? Is there questions that come out of knowing what happened that might be relevant to our being able to provide guidance to the Board to answer the Board's questions? So, we do think it's appropriate to actually have that task done.

The thing which we want to add, which wasn't there before, is the repeating of analysis from 2012, but we want to be really careful about

that. We're going to expand and define that a bit carefully. In particular, it's not about redoing the JAS and the Interisle work. I mean, that's just too much. I don't think that we need to repeat everything that was done before.

However, if we take the approach that we can use corp, home, and mail as a case study, and in fact what we do is just examine those three labels and redo the analysis on those three labels that JAS and Interisle had conducted. And look to grab that information and that data out and ask the same question about them, as well as try to get information from resolver data. We'll seek to see if we can get some of the public resolvers to ask similar questions of their data sets, not just the digital data, not just root server data that we might be able to get, but let's see what we can learn today from the same similar data sets that were done in 2012, as well as try to conduct those questions and that analysis on resolver data to the best that we can get. We'll have to look for some cooperative resolver providers and we'll have to work something out for how to get that done.

And I'll say one last little thing here, and I see your hand, Warren. The idea here is that, overall, what we need to be able to see is the trend from what was true before to what is true today. We need to be able to see what the differences are in what the data gives us because the decisions that were made in 2012 were based on certain metrics and that data was visible to get those metrics. We're kind of assuming here that the same kind of metrics are not going to be available in the same way. So, it's important to see what we can get and what's available and where it is.

And using corp, home, and mail as a case study gives us something really convenient to compare to so that we get that trending information to look at and then we can figure out what that means to us and how to deal with that as we think about the more general problem of what to do for strings that are yet to come.

So, I'll pause there. Warren, you have your hand up. Go ahead, please.

WARREN KUMARI: Sorry, trying to unmute. So, I think a couple of things. I fully agree that simply redoing all of the stuff that JAS and Interisle did is pointless. Redoing the same stuff and expecting to get different results seems interesting. But I'm not entirely sure that we want you to limit ourselves to just home, corp, and mail specifically because they've gotten so much press and media attention, etc.

So, I would think maybe home, corp, and mail, and one or two of the top NXDOMAIN queries hitting the root, or home, corp, and mail, and a random selection of other strings which we knew there were collisions for. It feels like home, corp, and mail in particular are sort of tainted.

JIM GALVIN: So, thanks for that, Warren. That actually is something that we had talked about. We hadn't really made a commitment to which strings to choose. So, at the time that we get a draft of everything put together here, maybe we'll see if we can't get yourself, for example, and your resolver data, and maybe with some luck, since we have Matt Thomas with us here, we can get Verisign to take a look at their root data, as

well as maybe in DITL we can do a quick query, and I say quick as if it's all that easy, but maybe there's a way for us to see if there's a few other strings that we should add and make them part of our case study. It is a valid point. I don't know what those strings are yet today, but we'll see if we can figure out what that might be and get that added in.

I'm going to make a comment here in the document so that we don't lose that point because it's not in here. Jeff, you have your hand up. Go ahead, please.

JEFF NEWMAN: Yeah, thanks. So I mean, I obviously understand the value of redoing or redoing it somewhat for corp, home, and mail, for those strings. But I don't understand what the value of those strings have to do with everything else, except for maybe, as Warren said, maybe the top couple that are having an outlier number of queries.

But everything else was not nearly in the same category as corp, home, and mail, yet there still were restrictions placed on them. And I think at the end of the day, at least for the next round, it's more important to look at what the effect of delegating strings that don't nearly have as many queries.

What's the impact of that and how do we mitigate that in the best possible way? That's what's going to help us for 99.9% of the strings for the next round, right? Corp, home, and mail are what they are, they've always been big and maybe local and a couple of others, but I don't see how that helps us answer the Board's questions unless we're setting the standard at, yes, we're only going to do something if the queries are as

high as .mail because that's I think the lowest of the three. Help me understand that.

JIM GALVIN: So, Jeff, thank you for the question. The critical point here is being able to understand the data. We have to know if the data is still there. That's all that's going on with this exercise at the moment. We have to be able to see what data we can get access to that we're going to be able to use.

So, our objective is to provide guidance to the Board and how to evaluate strings. Part of that guidance is going to be go get the following data from these sources in the following way, and use this as part of your evaluation of the string.

So, for our purposes here, we have to evaluate the effects of the change in the infrastructure, and recall that we identified four key changes in the infrastructure. We have to evaluate those effects of the data that we know about that's there and then consider how can we make judgments on what is available and what's not.

So, using corp, home, and mail seems ideal because we've got clearly solid data from 2012. Let's look at what that data looks like today so that we can see whether or not it's possible to see the same data. We're guessing, like one hypothesis here, is that some amount of what might have been seen in the root servers in 2012 probably won't be seen there anymore. Maybe we're going to be able to draw different conclusions out of what might be available in resolver data, or at least

we have to see what the effect has been over eight years on that data stream.

So, that's really all it is. The Board asked us specifically about those strings, so it would be helpful to …

JEFF NEWMAN: Absolutely, right, but that only gives us information on those three strings, which is fine. And then maybe I'm more interested in Study Three, I know you don't want to go into that, but I am just a little puzzled since another round is imminent on the horizon, and this only helps with the most extreme of examples and not with 99.9% of the strings they're going to get applications for.

JIM GALVIN: Well, we'll have to see, Jeff. I take your point. Your concern is certainly a valid one. I think that we're going to have to see what the analysis shows.

As we get to Task Four here about conducting an impact analysis, we are going to have to consider what kinds of questions to ask, and maybe we will expand our analysis. This is why Warren's point about let's pick a few other strings, maybe we can find out you know what some other top strings are, for some definition of top, and let's begin to ask queries about them in the data stream.

Again, it's all about trying to understand what data is potential to get at and we're going to have to figure that out. So, I can't solidly answer your question except to say that you have a valid concern and we're going to

have to see what the data shows us in order to really understand the next step. That's what I think at the moment. I'm open for other opinions.

And I'm looking in the chat room here for a moment. But, Warren, you had your hand up, too. Why don't we let Warren jump in here while I try to catch up with what's in the chat, too.

WARREN KUMARI: Thank you. So, I think some of the reason I say it seems reasonable to do corp, home, and mail is that they should be really quick and easy to repeat some of the analysis we did before, right? Like corp, home, and mail, but there's a bunch of other high impact ones like—Actually, Rubens just posted a thing to the ICANN Hedgehog.

But yeah, I mean, there's a bunch of well-known ones—OpenStack local, which just appeared, a bunch of the CPE ones which are also right up at the top, local DHCP, Internal, etc. So, it seems like cherry picking a couple of those and the analysis required really shouldn't be a monster huge amount of work. So, I don't really think it needs to slow things down if we get on with it.

I mean, from the initial set of name collision stuff, myself and Jordan Buchanan looked at the DITL data and it took a couple of hours, but it was a couple of hours. People could redo a lot of this analysis using DITL data, which obviously we've said many times is far from perfect, but would still be interesting to do if people would just invest a little bit of time and effort in it.

JIM GALVIN: So, thanks for that, Warren. An important question here really is going to be whether or not we think—when I say we, I'm really referring to this discussion group—are there members of the discussion group as volunteers who are willing to go forth and gather up some data and find a way to do these things? I mean, as Warren says, maybe it's just a couple of hours, maybe it's longer than that. I mean, if you're the right person, and you're connected well, maybe it is just a couple hours. If you hire a contractor, maybe takes a little longer because you sort of have to work out authenticating who's talking to who and working all that out and what's going to happen. There's those kinds of issues that that that come to bear here.

I appreciate that there are details about what is the analysis really going to show, and rather than having a repeat of the conversation that we had a couple of weeks ago, which I'm concerned that it feels like we're falling into that trap, I'd really like to just continue on with the rest so that you have the full picture of what's going to be there, what we're proposing, then let us create a document with this in it. And we'll come back to you with that, and then everyone will get to evaluate the details, and we'll get to see where we are. And there's an important reason for that.

Part of what's going to happen here is, if we agree on what it is that we're trying to do to some level of detail, ultimately, we're also going to have to add some costing information to that. And the NCAP Admin Committee is going to have to take on the job of figuring out how to assign some dollar cost value or effort value of some sort to this work if

we're going to hire a contractor. So that's going to become a consideration.

Let's first figure out what we want to do, and then we'll start to talk about, "Okay, are we going to do it ourselves? Are we going to look for a contractor?" And to the extent we want to contractor, we've got to try to put some ballpark numbers to it because that's what the Board needs to think about this. And there's some process to follow that.

So, again, I appreciate that there are questions about exactly what we're going to do. Let's wait until we have a proposal in black and white, rather than trying to spend a lot of time solution solving here on this call, if we can. And, Jeff, you never take your hand down, so I know that's an old hand. I don't know if you want to talk or if you want to drop your hand?

JEFF NEWMAN: No, I actually did drop it and raise it again, but probably between the time you looked.

Yeah, so when we were initially talking, when we were in Study One, we had talked about the possibility of running studies Two and Three concurrently, and that there was no need, because they were sufficiently different, to have it be sequential or not overlap. And, frankly, I'm a little concerned that Study Two is just, sorry to say this, but an academic exercise because we don't even have a hypothesis or any theory that we're trying to prove or disprove.

It's an open ended exercise, which is fine, I guess, but I'm worried that that could take a lot longer than eight months, and therefore, put a huge delay on getting the Board what I really believe they're asking for, which is criteria by which to analyze whether to delegate strings into the root. And I'm not sure how this gets us there. But I don't think this should stop us working up on Study Three because I think that has sufficiently different areas that we need to look at.

JIM GALVIN: Okay, thanks. Let me let Warren comment here, run out the queue before I respond.

WARREN KUMARI: Thank you. So, first off, huge thanks to Matt Thomas at Verisign for running some queries. I posted a link to an image of queries hitting A and J root, which he kindly shared awhile back. That's [inaudible] and percentage hitting the root. And, as you can see, they started in 2017 and grew a whole bunch. So, I think that that shows that there is still name collisions continue to occur and we haven't solved the problems, new ones come along.

But somewhat related to that, and here I'm tossing Matt under the bus, sorry. This didn't take a huge amount of time, from what I understand for Verisign to have generated. So, it seems that there are people available who could generate this data, or we could hire a contractor to do DITL, etc. But I think that while I agree with Jim that we need to figure out what exactly we want to ask them to do, in the amount of time we have spent discussing what we want to figure out what exactly

we want to talk to the contractor with doing exactly, we could probably have actually found someone, said, "Here's some small amount of money, this is roughly what we want, please go off and build some tooling so we can get a rough idea," then we can look at it and move on with it.

So, it sort of feels like we have reached the point where we have the fairly similar discussions again and again about what we want to have the studies do. And it might be reasonable to just say, let's just do what we think we want so far, get somebody to do the work. It doesn't need to be perfect, but they can give us some initial data. We can then see if it's useful and then make a decision as to whether we want to further refine exactly what the study is. And possibly, hopefully, Matt is on the call, now that I've invoked him many times. Yes, he is.

JIM GALVIN: Well Matt Larson's on the call, but not Matt Thomas.

WARREN KUMARI: Sorry. Yes, that's true. I was looking and I saw Larson. So, I incorrectly invoked [the] Matt.

JIM GALVIN: But that's okay. I think I want to draw a line under this discussion. And I really do want to agree with Jeff and Warren, but let me let me frame it in the following way. I just don't see how to escape being able to ask some questions of the data and do some duplicate analysis. It's not an academic exercise from my point of view, Jeff. And it can't be because

the infrastructure is different now. Warren, you're still unmuted. The infrastructure is different now. And we've already identified that and we know that, so it's important for us to see what data looks like today.

Maybe it is very simple, and it's very straightforward, but let's pause for the moment here, give us a chance to write down what we think we want to ask and do, and then we can have this discussion about whether or not to do it, or how much of it to do is I think really the question that we're having here.

Maybe it's not as deep an analysis as was done eight years ago, and that's fair. We actually don't want it to be quite as deep an analysis. But to the extent that our obligation is to give guidance to the Board on how to evaluate strings, part of that evaluation is looking at the following kinds of data if it can be available about a potential string.

This is how you're going to identify a collision string. This is how new, alleged proposals for strings can be identified as a collision string, go try and grab the following metrics. We have to understand where that data is and what it is, and we need to see how it It's changed from what was available in 2012 to what's available today. Maybe we'll find there's been no change, but the infrastructure change tells me that I have to at least check that the data is still there and shows itself in the same way. And that's all and I think it's important to do that.

So, okay, moving on. And Warren, I'm going to assume that's an old hand for right now.

WARREN KUMARI: Actually it was a new one.

JIM GALVIN: Okay.

WARREN KUMARI: I think what I was quickly going to say is, yes, I fully agree with you, I think we need to look at the data, but I think that in parallel with figuring out exactly what we want to see, we should start making sure that we can find people who have access to the data. We can start talking to people with the data and seeing how we're going to be able to use it. We could even get some people to start looking at the data that we have, because it often happens that until you look at the data, you're not entirely sure what all things you're going to be able to see.

And so, instead of coming up with the perfect plan, and the perfect decision on what all the study will cover, we could start getting access to data, building, tooling, etc., so that once we fully know, we can say, "This is what we want to know, use the tooling and access you have to go figure that out please." But I think I didn't read that very coherently, so maybe I'll drop my hand and stop talking.

JIM GALVIN: I understood that, Warren, and you're exactly right. Once we get some documentation around what it is we're looking for, certainly it would be quite valuable to start doing that at a minimum in an ad hoc way. We've got some good people on this group, we have some good connections. Let's start doing some ad hoc queries and looking because that will

certainly help to refine exactly what it is we want to do. And maybe that'll turn out to be enough, maybe we don't need a contractor for the analysis.

We do need a contractor for the root cause analysis, in my opinion. I noticed that no one's going back and talking about that. I think that's an ideal independent job to assign to someone to go figure out. Let them go do all of the legwork associated with tracking down everything we can possibly get at those collisions that we know that did occur, and document all that up and give that to us so that we have that to work with.

But the rest of this data analysis, agreed, let's try and scope it as best we can. And we'll get there.

What's currently called Task Three there, it's Task Three because it got bumped because of doing this notion of a repeat analysis there. We talked about build a test system for impact analysis and testing mitigation strategies. This gets a little bit to your comment, Jeff, and your question about doing Study Three.

Study Three really is about studying mitigation strategies. And in preparation for that, the idea here, at least in the original vision, was to build up a test harness during Study Two that would allow us to examine in an empirical way some set of mitigation strategies. The fact that we're proposing here to not do that suggests that Study Three is going to be something different than it was. Maybe we can pick that up sooner rather than later, I'm certainly welcome to that opportunity, but I don't want to do two project proposals in parallel at the moment here.

I do agree with you, Jeff. I think that Study Three is going to turn out to be more of a review of controlled interruption and we really have to be thinking about the future. I'm, speaking personally, less inclined to this test system. I'm less inclined to the idea that we're going to test all kinds of mitigation strategies. I think we're going to talk about mitigation more as a tabletop exercise, personally. But I want to set that aside for future work. So, we'll get to that.

Task Four up here is this Impact Analysis. We've kind of talked about some of this, but, in general, you'll see when I get to it down below here, we already had the discussion last week about we took the Board questions and we created a matrix and mapped them to studies Two and Three, and in fact Study One. So, we mapped the Board questions explicitly onto which particular study we thought was going to provide the most information towards answering that particular question.

And so, in this Conduct the Impact Analysis, we're going to be asking ourselves questions and iteratively asking questions of datasets in order to answer the Board's questions.

So, I actually had pulled from Item Two above, there might be some new questions. One of the things if we use corp, home, and mail and maybe a couple of others as a case study, if there is changes in what the data looks like, from 2012 and what it looks like today, that might promote new questions that we want to ask of the data that is available to us. So, part of the Impact Analysis, the real work that we're going to do is we're going to have to be thinking about as we get results, we have to think about what it means and is there more that we need.

We had always imagined that Study Two was going to be an iterative analysis of datasets. So, we had always imagined that we would be able to work something out with those sources of data so that we could have that be part of the discussion group and helping us to understand the data that they're providing to us, whatever results they're providing, and maybe refine questions and give it back. So, a little bit of iteration.

A second part of this Impact Analysis that we're going to be doing is a tabletop exercise. What data do we think we really need in the future? Now that we know what we can do, and we're going to try to test what we can do, let's think about the future here and what we're going to advise the Board about. What dataset evaluations have to happen? Where is the data that that's going to come from? Maybe we don't have to ask that many questions, but if we can get three data sets to ask a minimum set of questions to, that's sufficient. Or maybe one large data set and a bunch of questions is going to be sufficient.

So, we need to think about these things because we need to think about the future. And maybe we're going to iterate a little bit with a couple of the datasets that we have. So, these are questions that we want to ask ourselves about what we're doing. And then there's always obviously how much data do we need? Maybe not a lot, maybe a little bit, we're going to figure out whether or not this is what we need to do in looking at what we have, what we can get, and then think about do we need more or not? And what's the minimum amount of work that we have to put in here to do this?

There's also the retention period question. How far back do we have to ask the question? I mean, if we're going to suggest that the Board needs

to collect a certain set of metrics, about potential strings, how far back do they have to look for that? Is it one week, is it three months, six months, six years? Some data sets are going to have different variability in their retention periods. So, these are things that we have to figure out for the future and what we want to recommend.

And we might try to ask these kinds of questions of the various data sets to see what we get out of them. This is one of the reasons why repeating the analysis is appropriate. Maybe there's just three questions in the analysis to repeat, but we need to repeat that analysis and figure out answers to these detailed questions. Which data sets do we need? How far back do we need to look at the data? Maybe not all data sets are created equally, that kind of thing. These are things we have to do and we have to test out for the future so that the Board has something to work with.

And then the last question that we actually talked ourselves into in detail is this Data Sensitivity Analysis. And this gets to some of what was going on in the chat room here, too. And that is maybe not all of the data is going to be available in general in the future. I mean, in fairness, the resolver operators are all independent commercial companies, even all the public ones. I mean, yes, they're providing a public service, but the data that's behind that really belongs to them. Are they, in general, going to always make that data available to ICANN in order to do its process? I mean, I'm guessing probably not.

So, we have to think about what that means. These are the kinds of questions we have to ask ourselves, we have to consider privacy considerations, we've got to consider access considerations. We can't

just tell the Board to go get these metrics if they're not going to be available. So, we have to figure out what metrics we can actually get at and then how to use them. Even though we might get some additional data right now in our analysis, we have to be fair and recognize that for the future, it's not going to be generally applicable.

And so, this is an important part of what we have to think about in our Impact Analysis as we try to develop guidance for the Board. Warren, go ahead please.

WARREN KUMARI: Thank you. And I guess I'm just going to repeat something I've said on multiple NCAP calls and somewhat in the chat. Pretty much all of the open recursives have incredibly strict privacy policy stuff around them, and getting access to anything approaching something like DITl, I don't think is going to be doable from any of them.

If we do want data, I think it would behoove us to figure out what sort of things we want to ask for. And then I think a number of us in the group will be able to say whether that's even vaguely possibly likely, and then figure out how to reword it in a way that recursive operators might be willing to provide. Rubens mentioned something like DITL. I don't think any of their recursives would want to, nor should they, be willing to provide that sort of level of detail. The level of queries that one gets from an end user is incredibly intrusive. There's a tiny, tiny number of people at Google who have access to those logs, or have access to logs with that info, and even for that, every time they access them, they need to explain why they are looking at those logs.

But maybe things like what the root sees is these strings in the following orders, does that seem correct to you, or sane to you? Or can you give us an idea of the relative percentages of NXDOMAIN queries for things in these strings? Or what are the top things that are currently resulting in NXDOMAIN for you? And that sort of level of aggregation might be doable with the recursives or open public recursives.

But even that is going to require a lot of discussion with people's privacy and legal, etc. So, if we know what we want, we should really get on to starting those discussions early. Like, if I get asked for what is the top N in NXDOMAIN query, I would have to go along and talk to our privacy counsel and explain what the data would be used for, who would have access to it, under what conditions it will be shared, why it's useful, what the privacy implications are for the users. In this case, presumably almost zero, but there's still a long discussion that needs to happen. And so, if we want this access from recursives and also from authoritives, etc., I think that we need to get working on it real soon. Sorry, that was a rant.

JIM GALVIN: No, thank you, Warren. I do appreciate that and you've reminded us of that many times. And I do agree with you 100 percent, obviously.

I do want to separate it into two parts. I mean, this is the whole reason why this Data Sensitivity Analysis is an important part of our Impact Analysis that we have to conduct. I really do think that there are two parts of this. I'm actually hopeful that we might be able to get some analysis now just for our continued analysis in this working group in

developing guidance for the Board so that we want to just make sure that there's nothing interesting that happens in the resolvers that isn't visible in some other way. And it'd be useful for us to be able to ensure that and make that and if we can just do that as a one off, and then we'll just have to go with it in the future, that's useful.

Then there is the question of what happens in the future. I mean, even if resolvers are willing to be cooperative right now in our analysis, and I'm hopeful that we might get that cooperation, whether or not they will cooperate for all time is a completely different question and a different conversation. And we have to consider those questions too.

And that's the whole point of this Data Sensitivity Analysis is to get into that. And I take your point, Warren, sooner rather than later and all of that is true. That's why let's get this proposal on the table, let's see that it's going to get some support going on up the chain to the Board, and then let's just get started and start doing it and see where it takes us and you're absolutely right on all of that.

WARREN KUMARI: Yeah, I mean, one very last quick thing. I can tell you that the view from a recursive is wildly different to the view from the roots. Actually, Rubens put in a good thing, .mail is a prime example of that. There is a massive number of queries which hits the recursives which the root will never see because of caching, but of course it changes based upon the length or number of labels in the query, etc.

So, the visibility in the current data sets is interesting, but is not necessarily representative of what we want to know.

JIM GALVIN:

And thank you for that. So, that's just the part of things we have to understand. Steve, you have your hand up. Go ahead, please. And if you're speaking, Steve, we're not hearing you. Still not hearing you, Steve.

I'm going to go on and let you take your hand down and bring it back up again, give you a chance to rethink, maybe even reconnect if you want to see if you can get your mic working.

So, that's the impact analysis. What I want to do in this proposal, what the Admin Committee is offering to do, is we're going to try to pull all this together and provide a bit more black and white elements for people to look at. Actually write this up in a way that you can see it, and then people can comment on it, and we'll iterate on it here.

The process along the way here is that whatever we come up with here obviously is going to have to get some review by SSAC because the path to putting this in front of the Board is through SSAC. What we're hopeful about is that we'll get the SSAC review at the same time as the Discussion Group review, because they can certainly come to the Discussion Group and participate so that there won't be any time lag with that. And then, of course, the SSAC will have to submit it to the to the Board.

We have actually already alerted the Board Technical Committee to the fact that the Work Product One, which they've received from OCTO with the summary there, we've suggested to them that there will be something coming from this discussion group as a follow on to that. We

did report to the Board Technical Committee that we did take onboard the comment in Work Product One about studies Two and Three not being conducted the way they're currently designed, and that we would be providing some suggested changes to the Board Technical Committee for their review.

So, that's as much as we've said. We gave no technical details about what we might or might not do. We did put some of the questions that we were sort of wrestling with in front of them, but other than that, what we decide here is what we decide here. We haven't presaged what the answer is going to be that we give to them.

And Warren, your hand is still up. I don't know if that's a new hand or not. As long as I took a moment there, Steve, I'll give you another chance. Can you get your mic to work? And I'm not hearing anything. Okay.

So, the rest of those Study Two tasks, obviously producing a report and then public consultation and all that, it should seem kind of obvious. Those are the things that you do. The other things I want to remind you about here are, if you scroll down there, Kim, we do have the chart with the Board questions in it. So, we took the Board questions, we mapped them on to the Study Two tasks, the numbers that are above, as well as noting where Study Three was going to help answer questions. We didn't try to map Study Three tasks, we just mapped the tasks that we know about from above there, and how those tasks were going to be contributing to answering the Board questions. This particular mapping will be included in what we provide to them, and what goes there.

And then I thought there was one other thing here. Yeah. So, if you scroll much further down to the title of the thing that says, "Revised Study Proposals," top of page nine, right there. We have already written some introductory text as a start to what the Study Two proposal is going to look like. We don't want to write a whole lot, we don't want to make this a substantive, material change. We do just want to expand on what the elements are that we'd already proposed, what those details look like so that we can have a much better scope to what it is we're trying to accomplish.

And you'll see there at the bottom that the last line of that introductory text is some kind cost analysis. It remains to be seen exactly what that's going to turn into. We are going to have to give some kind of ballpark number to the Board as they look at the revision, and we'll figure out what that's going to be along the way here.

There'll be an Appendix One, which is the Board questions that we just showed you. And, of course, there'll be an Appendix Two, which is our gap analysis brief that we had previously done. We'll include that as supporting material. The idea here is just try to expand this one page proposal in another page or two, maybe three, to sort of add some detail to what those tasks Two and Four really are up above so that we can support whatever cost analysis we're going to provide to them.

So, obviously a clear consideration for us is whether we think we're going to just do those analyses ourselves, or do we want to get some help from a contractor to conduct them, which would be a really big change from our original vision—certainly an appropriate one, but just calling that out.

Steve, you have your hand up. We'll let you try again. Still not hearing you, Steve. Okay. Maybe if you can type something in the chat room, that would be great.

Otherwise, I think that actually brings me to the end here of what was up. I'll just summarize quickly by offering the following. In the large, our plan here is to provide some additional detail about what we think we'd like to suggest could be done in Study Two. It will be reasonably still closely aligned with what is Study Two, we're just going to provide some detail so that it's more consistent with what we think the world is today. And so, it will deviate a bit from the original vision from three years ago.

And we'll come to some consensus here in this group. We'll run that up through SSAC into the BTC. Along the way, once we decide on what it is we want to do, we'll have to give some consideration to whether or not we're going to do that work as volunteers, if we might seek the support of a contractor and what elements that might consist of, and then we will try to bring some level of effort and costing to that to also include as it runs up to the Board.

I'll add one little legal point here. As we have done in the past, this is an open forum. So, even if there's someone in this group, if we decide that there's work a contractor could do, since this is an open forum, if you're part of this group, you are eligible still to bid on that work if you're inclined to do so if there's work that we identify that can be done by a contractor. What makes that possible is that the details of any costing that we do is done by the NCAP Admin Committee. So, anyone who's part of the Admin Committee and sees that detail is automatically ineligible to participate in bidding for any work. But as long as we're

only talking about the work to be done in a public forum here in this discussion group, it does not restrict you from bidding on the work if that's interesting to you.

And it's just useful to remind people of that. You can feel free to mention that to your friends, anybody who you think. If it looks like we're getting to a place where there might be a third party who could do some work, invite them to come and listen to the calls and check it out because all of that is allowed it. It's all part of the process and that's allowed to happen.

What we will expose is the final total dollar figure, you just won't get to see the details behind it. We've already cleared with ICANN Council in the past that that's the split that we have to make in order to ensure proper protection for bidders and for all participants here.

I had thought this call would be much shorter, but we've actually run up quite a bit here. We've only got a little over five minutes left. That's the update on where we are. Next meetings, I think that we will not meet again until we have a document to share with you. Likely that's not going to happen for a few weeks for a couple of reasons. It turns out that all three of your co-chairs are out for some period of time here over the next couple of weeks. And, in fact, we don't even start to come back until the end of August. Part of that is because one of us is European and has got a lot of off time. I'm American, but I'm still taking some off time. And, of course, Matt Thomas has some family issues going on. He's also taking some time off for a period of time.

So, sadly, until we have a document to review, I would say that we're going to take a pause here in our meetings. It doesn't seem appropriate to continue to meet, there's no one to chair the meetings. So, that's where we're going to be. So, the next meeting has yet to be determined. We'll watch the mailing list. We'll declare the meeting when we have something to distribute. We'll send it around to you, and then we'll schedule a meeting for our usual time slot in the next Wednesday after that. That's the current plan.

So, sorry that you had to hear my voice this entire time. But any additional questions, comments, any other business at all from anyone about anything up to this point? Not seeing any hands. I'm guessing Steve is giving up, he's not going to type in the chat room either.

STEVE CROCKER: I was just going to comment earlier …

JIM GALVIN: Your microphone works. Go for it.

STEVE CROCKER: Sorry about that. I was just going to come up earlier in reference to getting data from recursive resolvers versus the root, that a key point of sensitivity is whether the source of the query is included. Once that's included, it's sensitive no matter where you pick it up. If it's only the question of what's being queried, that's much less sensitive. What string you're asking for versus who's asking for it.

JIM GALVIN: So, yes. Thank you for that distinction. Obviously, we're ultimately going to have to clear that with the resolver providers themselves. I'm hopeful, as you're suggesting, that it's a relatively straightforward thing to get passed, but far be it for me to speak for the privacy attorneys at various companies. And I'm sure Warren's going to stand up and say, "Well, he's not making any promises." Go ahead, Warren.

WARREN KUMARI: Yep, not making any promises. But yeah, I mean, what Steve said is definitely true. The source address of the query is incredibly private, but even if that's removed, there's a lot of information which gets included in people's DNS names. Looking at the privacy policies of people who have put them up for their DNS public resolvers, it's something which people treat incredibly, incredibly sensitively. There's a tiny number of people at Google who have access to any of these logs, and when they do access them, they have to explain specifically what they're using it for for each time they access them.

So, even just stripping out source address and aggregating to second level, I think it's going to be way more than people would be comfortable sharing. And also, apart from the legal side of it, there is a definite cost in terms of user trust. And so, there needs to be a good explanation as to what the value of sharing this data is for the users. Right? How does any recursive resolver sharing this data make my Auntie's life better? And then, even if there is a good and compelling reason, people will need to go through the privacy policies for all of the

recursive resolvers and figure out if it's doable. And if it's not, what would need to change in order for that to happen. And then, those would have to be updated, which is why I keep stressing we need to put out exactly what it is we want to ask and get started on that, and have a willingness to coordinate on what data can be shared.

JIM GALVIN: So, thank you for that, Warren, and thanks to Steve. So, let me bring this to a close here and just say thank you to everyone. The last couple of meetings in particular, we've had some really good discussion, and we really had some good discussion today. I want to thank folks for that. I really do. This is kind of a very specific topic, it can be a little a little tough to get into.

The real meat of why we're here is what's coming. It's always been, in my own personal opinion, that Study Two is the real meat of what we're trying to do. Most of the Board's questions are answered by Study Two. There are certain elements that needed what we had from Study One, those elements that need what we would do in Study Three, but the bulk of them really depend on this work. And this is where we're going to get into the real discussions. What data is there, what does it mean, how do we use it? And what can we tell the Board about how to use that data in making its decision?

So, the real work is coming. This is what we're getting into. And the discussions that we've been having the last couple of times are certainly indicative of the fact that we have a lot to get through. So, we'll try to get started on that as quick as we can here, taking advantage of what

we can do while we're waiting for any formal process and whatever other contracted work might come out of this that we might need.

So, thanks to everyone. One last call for any other business? Not seeing any hands. So, once again, we'll meet again when you see a notice on the mailing list. There'll be no future meetings until we have a proposal to review. And hopefully at that point, we'll pick up from there and start getting into some really interesting discussions. Really appreciate your time. Thanks, everyone. We're adjourned.

KIMBERLY CARLSON: Thanks all, bye.

[END OF TRANSCRIPTION]