AUTOMATED VOICE: This meeting is being recorded.

FRED BAKER: And we're being recorded, in case you wondered. John, could we start with you?

JOHN KRISTOFF: John Kristoff. I'm a research fellow with ICANN.

PAUL VIXIE: I am Paul Vixie from Cogent.

KEN RENARD: Ken Renard, ARL.

HOWARD CASH: Howard Cash, ARL.

MATT LARSON: Matt Larson, ICANN org.

PAUL HOFFMAN: Paul Hoffman, ICANN org.

BRAD VERD: Brad Verd, wireless ticket maker.

LARS-JOHAN LIMAN: Lars-Johan Liman, Netnod.

HIRO HOTTA: Hiro Hotta, WIDE and JPRS.

SURESH KRISHNASWAMY: Suresh Krishnaswamy, RSSAC Caucus.

SUZANNE WOOLF: Suzanne Woolf, USC ISI.

UNIDENTIFIED MALE: … RSSAC Caucus [inaudible] to RSSAC.

DUANE WESSELS: Duane Wessels, Verisign.

JESS OSBORN: Jeff Osborn, ISC.

FRED BAKER: Fred Baker, ISC.

KARL REUSS: Karl Reuss, University of Maryland.

NAELA SARRAS: Naela Sarras, IANA liaison.

ROBERT STORY: Robert Story, USC ISI.

DANIEL MIGAULT: Daniel Migault, IAB liaison to SSAC.

TERRY MANDERSON: Terry Manderson, ICANN org.

ANAND BUDDHDEV: Anand Buddhdev, RIPE NCC.

KEITH BLUESTEIN: Keith Bluestein, NASA.

PETER DEVRIES: Peter DeVries, RSSAC Caucus.

MATT WEINGBERG: Matt Weinberg from Verisign.

ABDULMONEM GALILA: Abdulmonem Galilia, RSSAC Caucus.

FRED BAKER: We also – I can hear noises online. Those on line, could you please say who you are?

UNIDENTIFIED MALE: We have no one [online].

FRED BAKER: You don't have anyone? Then I'm hearing – I don't know. We also have various staff in the room.

Brad, maybe you want to go over ground rules?

BRAD VERD: Real quick, just a few logistics. Again, welcome, everyone. Breakfast will be here after day. We'll have lunch also. Bathrooms, for those of you who haven't been here, are out this door, directly beyond the exist. Hang a right and bathrooms are down on the left. Your badges basically allow you access to the restrooms in this whole executive briefing center. So please don't wander beyond that. Otherwise people will find you and bring you back.

Wireless credentials were e-mailed directly to your e-mail. The e-mail came from me. Those credentials unfortunately or fortunately,

depending on what your point of view is, are reset every day. So we'll have to do this again tomorrow. The e-mails that they were sent to were the ones given to me from ICANN. If, for some reason, you can't get them or they don't work, you need to let me know and we'll send it somewhere else.

Does everybody have wireless?

[LARS-JOHAN LIMAN]: I haven't tried yet, but [inaudible].

UNIDENTIFIED MALE: Just checking. Does anybody need a power splitter? Does everybody have power? Any problems with power?

Okay. Never mind then.

FRED BAKER: Ozan, are you – Ozan was around here.

OZAN SAHIN: I'm here.

FRED BAKER: There you are. You want to go over the agenda for the next three days?

OZAN SAHIN: I'm trying to [inaudible]

FRED BAKER: Oh, okay. You're busy. Carlos, can you do that?

CARLOS REYES: Hi, everyone. Before we get started, I did want to take some time to introduce Danielle. She just joined the support team. This is her first workshop. Danielle joined ICANN a little more than a year ago, right?

DANIELLE RUTHERFORD: Mm-hmm.

CARLOS REYES: She was working with the team that processes all of the advice from the advisory committees. She just joined the policy team about two weeks ago, I think.

DANIELLE RUTHERFORD: Mm-hmm.

CARLOS REYES: All right. So …

All right. So please say hello to her. Like I said, she'll be here this whole week. You'll start seeing her on all RSSAC-related calls and sessions at ICANN meeting. We're really excited to have her.

Danielle, would you like to say anything?

DANIELLE RUTHERFORD: Yeah. I'm really looking forward to working with this group.

CARLOS REYES: All right. Ozan is putting up the agenda here. I'll stall for a few seconds as he gets it up.

Okay. So just to quickly go over the next few days here, thanks, everyone who joined the reception last night. We'll have obviously more networking throughout Tuesday and Wednesday as well with dinners, as Brad mentioned.

Today we're starting off with the working breakfast now. The first session is an RSS Metrics high-level overview of the document. Duane and Russ circulated that.

UNIDENTIFIED MALE: [inaudible]

CARLOS REYES: Yeah. Ozan … So the first session is going over the document. I think that was circulated last week. We have a coffee break after that and some meta questions regarding thresholds, lunch, and then, in the afternoon, we're discussing availability, latency, and publication latency. That'll be continued throughout the afternoon. In the afternoon, we

have a wrap-up, and then later tonight we have dinner. We'll meet in the lobby and then walk over. Dinner is at American Taproom.

I'll pause there and late Duane and Russ say anything, if you'd like to say anything about the schedule today. But that's the overview of today.

DUANE WESSELS: I don't have anything else to add.

RUSS MUNDY: I guess the only request I would put in that – I know it's hard to remember – is to please say your name when you start to speak because we're going to have these recorded. If we have to go back and make reference to what was said, it's really helpful to know who said so we can follow up with you need to. Thanks.

FRED BAKER: [So then] the continuing agenda tomorrow and Thursday. Tomorrow we're going to be on metrics most of the day again. Thursday we're on metrics in the morning. In the afternoon, we have the RSSAC meeting (the monthly meeting), and then we disappear and go to the airport.

Questions on that? Agenda bashing?

Okay. So—

[NAELA SARRAS]: Sorry. I have a quick question. What time is the afternoon meeting? The RSSAC meeting on Thursday.

FRED BAKER: For that, we have to see the agenda.

[NAELA SARRAS]: Yes.

UNIDENTIFIED MALE: 10:45.

[NAELA SARRAS]: Oh, 10:45 in the morning. Okay, thank you.

FRED BAKER: I think we're now waiting for 9:00, which is 11 minutes from now, unless Russ and Duane want to move on.

DUANE WESSELS: Did you do the ground rules?

FRED BAKER: Brad told us not to go to places we shouldn't go.

DUANE WESSELS: All right. Got it.

BRAD VERD: I'll just add to that. I think we all know that the main topic here will be metrics. I think it can be a heated discussion. So let's be cordial to each other. Be civil. But, if you have an opinion, please share it. Don't withhold. This is the time to get it out and get the document right. So let's put it on the table if you've got some. That's all I'll add.

FRED BAKER: Russ and Duane, let me give you the floor.

DUANE WESSELS: Okay. Ozan, are you going to present the slides, or do you want me to present them?

OZAN SAHIN: [inaudible]

DUANE WESSELS: [inaudible]. For the first session, there's a PowerPoint slide deck that I thought we'd go through. Ozan will put it up on the screen and in the Zoom room, I guess.

Yeah, the status slides. All right. Thanks, Ozan. This slide deck goes over the current status of the Metrics Work Party and the document that we've been working on. You don't have to follow along in the Google

Doc, although I suppose you could if you wanted to. You'll probably lots of chances to do that in the coming sessions.

Next slide, please. This outline just gives a general structure of the work party document so far. It has these general sections, which we're just going to go through one by one. There's an introduction, which is pretty short, a background and scope. There's a section that talks specifically about vantage points. Then there's a section that's general comments about metrics and measurements – things that apply to all the measurements. Probably the most meaty part of the document is the section on RSO metrics – these are metrics specific to individual root server operators – followed by a section for metrics on the root server system as a whole. There are some recommendation. Then there's some boilerplate stuff, and then an appendix with example results.

Next slide, please. A note on BPQ. In the statement of work, there was an item that says something like, "The work party should refine the concepts of BPQ from RSSAC037." In the work party, we did talk about this quite a bit. In the end, we decided that it didn't really fit and that it should be removed from this document and published separately. There is, as far as I know, not yet a separate document created for that. So this is still a to-do item. Just so you know, this document does not reference the BPQ stuff anymore.

LARS-JOHAN LIMAN: Thank you. But if it doesn't fit here anymore, is it relevant at all then? Do we need to create this separate document for it?

DUANE WESSELS: I don't know.

LARS-JOHAN LIMAN: All right. I suggest for someone to come up with a reason to create a separate document before we put an effort into that.

DUANE WESSELS: Right. That's an open question. It's an open question whether it needs work it all, whether it wouldn't need work in this party or a separate work party, or not. We just felt that it didn't really fit with the other things that we're talking about in this document.

BRAD VERD: I'll just add, I think, when the statement of work was written, it was a nice-to-have. I think it says that in the statement of work. It was like, "If possible, if the opportunity arises" type of thing. So having that removed, I think, is fine.

DUANE WESSELS: Okay. Next slide. The introduction section has essentially a copy of the statement of work or the relevant parts of the statement of work. It goes over the document structure. Then I think it has some boilerplate text, which all RSSAC documents have, that describe how this is a product of the work party and represents RSSAC's [inaudible] recommendations and so on. So that's pretty straightforward.

[inaudible]. So there's a background and scope section, and it has these subsections. It opens with a purpose. It describes the purpose for this work and this document, specifically referencing minimum levels of performance. It also mentions that this work is motivated by the RSSAC037 document. In that document [there's what's] called the performance and monitoring and something else function. It also includes the acronym or the phrase "service-level expectations."

As a whole, the document does not use the phrase "SLE." I don't think it uses it outside of this background section. I think that's the only place. We do, from time to time, talk about levels of performance, but SLE? This is the only place that occurs. SLA definitely does not occur in this document. At least not at this time.

There's a section which captures some of the discussions that we had a while ago. There was a time where this whole document was trying to be much broader in scope. We agreed to narrow the scope. So there's a section that says there are certain things that are not in scope for this work. For example, out of scope are research into performance trends. We are not collecting this data or describing these metrics for that purpose. So they're not for long-term research into performance trends, and they're not really to be used to make comparisons between root server operators – not to say one is better than the other in some way.

Jeff?

JEFF OSBORN: It's noted – sorry about my voice – just above Section 2.1 or 2.2 that this is a starting point for discussions about the financial work that the

PMMF is going to do. I've been following along and showing up at all these meetings, hoping to find a place to get a toe grip on where do we begin work on the financial function. What we get is a statement that this is a starting point. So, while we're not addressing that, at some point that's another path we've got to go down at some point.

BRAD VERD: Can we—

DUANE WESSELS: Can you – sorry.

BRAD VERD: I was going to say, staff, can we add that to the work queue? It sounds like we will volunteer to chair the working group.

DUANE WESSELS: Right. Jeff, I was just going to clarify a little bit your thoughts. You're not saying that this work party needs to address those items. Because I hope not. You're thinking it would be a follow-on work [party]?

JEFF OSBORN: I'm spending the last few months dreading and looking forward to saying what I said I would do. I think the financial function should probably be different from this because it's going to involve walking around a bunch of people with sparklers and gasoline-soaked clothing. Not everybody wants to do that.

DUANE WESSELS: Okay, thanks. There is also in the background and scope a subsection that mentions prior work. It's pretty straightforward. It mentions mostly other RSSAC documents and things that RSSAC has worked on.

Then there is a terminology section. If you've seen an older version of this, the terminology section used to be a little bit longer because some definitions had been copied from RSSAC026 into this document. But now those definitions that appear in both are in synch and it just says, "Go see 026 for these other ones."

There's really only three new terms currently defined in this document, which are on the next slide, Ozan, please. Things defined specifically in this document are measurement, metric and threshold. The definition for measurement is along the lines of, "This is an individual, one-time measurement, like a query response, from what we're calling a vantage point to a server." The individual measurements are aggregated together into metrics. The thresholds are applied to those metrics. So those are the only three things really defined here.

We struggled a little bit with this at the start. I think maybe not everyone was perfectly happy with these definitions, but at least we've stopped arguing on that and settled on this. So I think it's pretty much done at this point.

Next, please. We talk about advantage points. In earlier versions, we use the word "probes" a lot. I believe that almost all, if not all, of the references to probes have been stricken, and we're sticking with vantage points. This is a recommendation, I believe, from John

Heidemann because, in his opinion, probes was ambiguous or meant something a little bit different.

There was a specific recommendation in this section on the number of vantage points, or at least the minimum number of vantage points that there should be. At this point, the work party is recommending approximately 20 vantage points that would be used to perform these measurements and metrics. There is a section which talks about where vantage points should be located – again, another recommendation – and the text is something like, "Distributed evenly among five geographic regions." Then it actually lists those regions.

There's a follow-on recommendation, which is to maybe try to do better in the future, to not be so focused on geographic distribution but on topologic distribution instead. But that is a future work item.

[BRAD VERD]: It is definitely.

DUANE WESSELS: I'm sorry. Definitely a future work item. Then there is a section on connectivity requirements for vantage points. It says that they should be located at data centers with good power and good connectivity and that vantage points in the same geographic region should use different connectivity providers. You don't want to, obviously, have all the data points using the same ISP or service provider. That could skew the results.

So those are the recommendations in that section. Any questions about that before we move on?

BRAD VERD: My only comment is, while I understand the location comment or the location bullet that says "evenly," I'm not so sure that "evenly" is really … My fear is, given how strict interpretation of words have happened in work parties and everything else, that that will be strictly interpreted in the future. I'm not sure that's what you want, given that the user base is not evenly distributed in the geographic regions.

DUANE WESSELS: Yeah. My thoughts on that are that, given that, at this point, we're recommending, say, 20 total vantage points and [there are] five regions – that's about four vantage points per region – that's pretty coarse. That's where we have to start. I could see that becoming more of an issue as the number of vantage points increases. If anyone has suggested text, we'd welcome words to make this better. At this point, it's super detailed. We're being a little bit vague on purpose, I guess.

PAUL HOFFMAN: It actually says, "Evenly among the five following," meaning, if its 24 per, it doesn't indicate at all "evenly within each region." I think that's especially because of the follow-on, which says you really want to have these on different networks. And you might have to put two of them in a single data center connected to different switches in order to get that.

BRAD VERD: I know. I don't think we need to spend a lot of time on this. I'm just more concerned about how, the next time this goes around and somebody sees the word "evenly," and you get 1,000 … I could see where this becomes a problem.

DUANE WESSELS: Yeah. To me, this is a very hard thing to get right: describing how—

[BRAD VERD]: There is no right. [inaudible]

DUANE WESSELS: Okay. All right. Next slide, please, Ozan. This section – general – is a section that applies generally to all of the measurements and metrics. It talks about time stamps and scheduling. I think it says, for example, that time stamps will always be in the UTC time zone. Regarding scheduling, it says that … So all these measurements are done on five-minute intervals, and I think there's text that says something like – what did we settle on? "Sleep for some random amount of time between 0 and 60 and then do your measurements." So the idea is that the measurements always occur at the start-ish and middle-ish of the measurement interval. But there's a little bit of a random distribution so that not all vantage points are doing their measurements at exactly the same time.

The section on elapsed time and timeouts. I think the elapsed time one is specific to TCP, if I recall correctly. For example, when measurements are done over TCP, the clock starts when you initiate the connection, and the clock stops when you get the response back, essentially.

There's a little bit about connection errors, how those are handled and reported. With regards to spoofing, it says that the vantage point or the software should take all the possibly precautions to prevent spoofing. That means using the standard techniques that we're aware of: good query ID randomization. Responses are to be matched based on all the protocol parameters: the source address, the destination address, the port numbers, and so on.

Some of the metrics use a type of query that we call [randnxd]. This is a randomly generated name that is designed to elicit an [NX] domain name response. It describes how that query name is formed. The query names generated under this method are prefixed with the word "test" so that they could be picked out in data collections, like DITL and other sorts of things if desired.

There's a section about Anycast, which says that the purpose of these metrics are not to intentionally try to hit all or different Anycast sites of a single root server. It I know X-root has a site in Chicago, I'm not sending a query specifically to that site. I'm only sending a query to the root server, to the X-root, and letting the rooting system decided where that query is delivered.

This ends with a section on measurement reuse. As we'll see, a lot of the measurements are really derived from the single query and a response. So one query in response can be used to determine a number of different metrics – for example, latency and availability and so on. This section now I think is very much more explicit about that and says that the implementation of these metrics should reuse the same queries for multiple measurements.

PAUL HOFFMAN: There are some questions and comments sprinkled throughout this. Before we get into the more meaty fun part, do we have on the agenda to go over these sections at a time before or after we hit the metrics themselves?

DUANE WESSELS: We have time on the agenda dedicated to the metrics, but we don't have specific time generated to this general section. So if you wanted to—

PAUL HOFFMAN: I have a concern on some of those. For example, on the [randnxd], I actually have a different proposal for what we would do instead of that, which would then apply later. So I don't when in the agenda over the next couple of days that would be appropriate.

But also, because once my finger starts moving I can't stop editing, there are some other comments on some of these in the document for before. I don't know if we're going to try to hit that during these three days or not. Because some of them will affect our discussions later.

DUANE WESSELS: Okay. Why don't we see if there's time at the end of this session to talk specifically about that?

SURESH KRISHNASWAMY: Another general question about the assumptions that we're using in this document. I read the section about what we expect in terms of each of the vantage points, in terms of the connectivity, but there's also some variance in the way that each vantage point might contribute to a reading. For example, there might be a local issue that might produce some sort of latency issue that might factor into the measurements.

Are we taking into account any of those in the way we construct the measurements? Is there any attempt given to try and reduce the effects of local problems? Are we just assuming that the vantage points are well-connected and will not suffer any of these issues that would make the metrics go bad?

DUANE WESSELS: Thanks. It's more the latter. It's more that we are assuming that they are well-connected and that any variations and thing like latency will be countered by the fact that we're looking at median values and things like that.

There is a little bit of description in the document about how care should be taken to only use measurements from vantage points which are known to be connected. So the document doesn't go into detail about how you deal that, at least not at this point, but certainly we should exclude data points or measurements from vantage points who are having connectivity problems.

UNIDENTIFIED MALE: [inaudible]

DUANE WESSELS: Yeah. So it's kind of addressed, but kind of not.

DANIEL MIGAULT: If we exclude the quality of the connectivity, it may impact, I think, the measurements because, if you have no connectivity, then you have a [lousy] connectivity. I don't know how can exclude those measurements, but we need to be able to detect those or to find a way to be able to exclude those.

DUANE WESSELS: Well, the way it's phrased in the document right now is that it's a requirement that the vantage point be located somewhere with good connectivity. So, if it's not, then it's not a good vantage point and it needs to be taken out and repurposed or something like that. If it's a temporary problem, where it was bad for a day, then either we have to live with the fact that it was bad for a day, or we use our ability as software developers to find a way to eliminate those data points. But, in general, if the vantage point is not well-connected, then it's not a good vantage point. It needs to be moved of taken out entirely.

PAUL VIXIE: I have two concerns there. The first is that we are trying to serve the whole Internet, and we should try to find out how well we're doing, even the somewhat vulnerable, less well-funded parts of the Internet. We need to know what it's like to try and reach this service from those places, even if it does not necessarily direct our tactics – its input for

transparency, among other purposes. We ought to be measuring from places that don't have good connectivity.

The second is, no matter have good the connectivity may have been or may usually be, every place on the Internet is subject to other people's networks and can have a bad day through possibly no fault of its own. That would call for a second parallel baseline measurements of something that is not us which they can do in parallel with probing us so that we can find out when something that is not us is also unreachable that we can maybe buy us our treatment of that metric as probably due to something on their end.

Those are my two concerns. Thank you.

KEN RENARD: I think we had a lot of this discussion in the near versus far probes. I think a lot of it is captured in Section 4.8 (Unexpected Results), where you can look into this. "There's something weird happening. Let's look into it." So take a look at 4.8.

PAUL HOFFMAN: I think, based both on Vixie's and Daniel's concern, 4.8 needs to be more explicit. Right now, 4.8 says, basically, if you're a root server operator and you've been failed for some reason, we can report to you why and such like that. I think it needs to be more explicit where, whoever's running the system and can look and say, "Vantage Point 12 is obviously offline or being heavily affected. Therefore, we're not going to include it in the metrics until it is either better or not." We don't say

that in 4.8. So I think 4.8 needs to have that added to it to give whoever the measurer is or the pass/failer is leeway for that.

Now, obviously, that would have to be reported and such like that, but, in the case that it's down, that's pretty obvious. In the case where it's impacted, I still think that that is something where the measure should be able to say, "We're just not going to count this for pass/fails right now."

DANIEL MIGAULT: If I had to program those, one way I would have done it to connect the Internet connectivity would have been maybe to have had a few connections with reference websites to evaluate how the connectivity is in one location. So that's the idea I have behind that.

I can add text if it's helpful.

DUANE WESSELS: Yeah. What's most helpful is specific descriptions of how that – put something in the document on specifically how you would propose to do that.

DANIEL MIGAULT: Okay.

RUSS MUNDY: One of the things we've talked about some in the work party is that the definition that's specified in this document is really focused at the

metrics activity and not at trouble-shooting. So, as we get to the edge point of identifying if there's a problem somewhere, whether it's a [low]-impact problem or an availability of a vantage point being up and down, I think we need to be at least cautious about what and how we say, that we don't move into being a trouble-shooting specification versus a metrics spec.

DUANE WESSELS: Yeah. Thanks, Russ. As Ken brought up, there's actually another section under General, which is unexpected results, which I guess I missed on the slides here. But we've talked about that no, so I think you're aware of that. There's actually quite a lot of text there, but, as Paul said, there's maybe something missing. So we'll add to that in the future as well.

All right. Any last chances for questions on this before we dive into the actual metrics themselves?

Okay. Next slide. These are the five metrics defined for root server operators. We have availability, response latency, DNSSEC correctness, correctness by matching, and publication latency. To the point about measurement reuse, #1, #2, and #5 are all measured with the same set of queries and responses. Then the correctness ones are on their own. They stand alone.

All right, next one. Next slide. Root server operator availability is relatively straightforward. The proposal is to actually report separate metric for the different transports. There's a v4 UDP, a v4 TCP, a v6 UDP, and a v6 TPC metric, or there would be essentially four values.

Yes, Paul?

PAUL: Can we be careful about the term "report"? Because I thought we agreed that the pass/fail was going to be on the combination of them. But we would be able to report to an RSO if they were failing which ones. Am I wrong about that? So will this generate four red/green boxes on the public website, or just one, in your mind?

DUANE WESSELS: In my mind it's four.

PAUL: Ah, okay. I thought we had gone just to one, to a combination of them. That's fine. I'm just saying—

DUANE WESSELS: If were to do that, we would – there's currently no definition or proposed way of how to combine them.

PAUL: I thought that actually the aggregation did combine them. So I misread the aggregation statement. Okay, great.

DUANE WESSELS: So, as proposed right now – again, we're all making the sausage right now – in the reporting, there would be four separate lines or sections—

PAUL: In the public reporting.

DUANE WESSELS: In the public reporting, yeah.

PAUL: Great.

DUANE WESSELS: The different transports would be reported separately. So the way that you do this is you send some number of q queries in the time interval and you count how many responses are you received. Then availability is just the ratio of those two. It's how many responses you got over how many queries you sent.

As mentioned before, this and all the other metrics are done on five-minute query intervals. So that's 288 per day and there's, say, 20-ish probes. So that gives you 5,760, I think, possibly q. In a given day, q should be 5,760. So we're talking about aggregating approximately 6,000 measurements over the course of a day into a single availability metric.

[SURESH KRISHNASWAMY]: Just as a point of clarification again, all of these measurements are again modulated by what we say are queries and responses in the sense that a single query can return multiple responses if there's some other

activity is happening – someone is replaying responses. So we are assuming that these are valid responses, so we've done everything in terms of matching query IDs and all of those things. So these are legitimate responses to queries that we've issued? Okay.

DUANE WESSELS: Yes, that's right. The measurement software needs to do everything possible to ensure that spoof responses are ignored. Under normal operating conditions, the expectation is one query/one response. It's possible to do two timeouts/[inaudible]/whatever. You may get multiple responses, I suppose, but the software should be able to handle that.

In this metric, timed-out queries are not retried. So you try once and wait. I believe that the time of value is four seconds. If you get a response within the four seconds, then you count the response. If not, it's no response and that lowers your availability for that time period.

RYAN STEPHENSON: Hi. It says period of one day. What's the definition of one day? Is it just Greenwich Mean Time? Are we going to be using that as a clock, or – yes. Okay.

DUANE WESSELS: Yeah. All times will be in UTC time. So that's also the boundaries for the days.

RYAN STEPHENSON: A follow-up on that. Okay, so that's going to be the boundaries of the days. It's not going to be rolling day. Say, five minutes after, five minutes after, for – okay, great. Thanks.

ANAND BUDDHDEV: I just have a clarification question. A response could also be something like FORMERR. I'm assuming that we considered that to also be a response and assume that the server is available. Whether it's a correct response or not is I think for a later section. Am I correct in assuming this?

DUANE WESSELS: That is how it is currently written, yes. I go back and forth over that's the right thing to do, but that's how it's currently described. If anyone wants to argue for one way over the other, I'd welcome that. But, at this point, even if you get a refused or a FORMERR, it's still counted as a response.

ANAND BUDDHDEV: Thanks, Duane. I would prefer that as well because that's different from correctness. This is just availability, and a form error that refused all of that should be considered as available.

HIRO HOTTA: I had the same question. I'd like to know whether the content does matter or not.

DUANE WESSELS: For this particular metric, the idea is that the content does not matter. My slight hesitation with that is that it's much easier to spoof maybe a refused or FORMEER than it is to spoof a legitimate response. So we may be a little bit susceptible to some—

UNIDENTIFIED MALE: [inaudible]

DUANE WESSELS: Yeah, right. It may be more likely to have a false positive in this metric than some of the other metrics because of that, but that's the way it's currently described.

Any other questions about availability? One thing you can do – we don't have the document up on the screen – as you're going through this, if you do have the Google Doc open, you can scroll down to the example section and get a sense of how this might look like when they're being reported formally.

All right. Let's go on to the next slide, Ozan. This is root server response latency. In the previous one, there are four separate measurements for all the transport combinations. In fact, it uses the same query response from the availability metrics. In this case, we're just looking at how did it take to get the response. I said before, over the course of the day, that there are potentially 5,760 such measurements per RSO, assuming 20 vantage points. All those would be aggregated together. The median value would be calculated from that aggregation.

[FRED BAKER]: Four?

DUANE WESSELS: Four separate medians. For example, again, as currently proposed – when we get to the threshold discussion, it's possibly that a root sever operator could be below the thresholds for UDP v4 and above a threshold for UPB v6.

Fred?

FRED BAKER: Dumb question. If we're only using the Anycast address of the RSO, how do we address all these probes?

DUANE WESSELS: How do we address all of the …

FRED BAKER: The document wants to calculate the median RTT latency from all RSO probes for that RSO. If we're only using the Anycast address, how do we get to all of the probes?

DUANE WESSELS: The probes are separate from the RSO. There would be 20 probes, say, operated by ICANN or a third party. It's perhaps possible that all of those 20 probes would reach the same root server instance. We hope

that doesn't happen. That's not what's supposed to happen, but it's possible. But each probe will measure all of the RSOs. So, 20 probes measuring, today, 13 RSOs.

FRED BAKER: Earlier, we changed from probes to vantage points, right?

DUANE WESSELS: Fred tricked me.

FRED BAKER: It's your slide.

DUANE WESSELS: Well, somebody didn't read these slides clearly. I'm sorry. Yes, vantage points.

Ozan, can you quickly fix that? No, I'm just kidding.

DANIEL MIGAULT: I see the median as a way to remove the outliers. When we're doing the median over time, it means we assume that there is a way to have a meaningful value over time. When we take the median among the different probes, we're implicitly mentioning that there is one value that represents the [dys]connectivity around the world. So I just want to make sure. It kind of works if we except to have the same connectivity between all the vantage points in the RSOs. Otherwise, it might be

tricking because we can have the median in one place instead of the other, depending on how many measurements we're going to have. So they might be clustered at different places. So it's just something I'd like that we clearly understand what it means.

DUANE WESSELS: I think this is an important point: everyone understands how the aggregation is proposed to work. For this, it's a straightforward median of all the vantage points and all the measurements over the course of a day for a single root server. We can consider something more complicated if there's a reason to, but that's what's currently there.

DANIEL MIGAULT: Just a complement. I think, if we have 20 probes located at the network exchange point[s], it might make sense. Where it won't make sense, probably, I would say, is if we have one of those vantage points into an island very far away with local activity. That might [inaudible] the measurements. So that's the …

MATT: Considering that we only have vantage points, I would argue that we want to do specifically to places that are well-connected. To Paul's point, Paul said that we should be able to take measurements from possibly less well-connected areas throughout the world. At some point, we just have to draw a line. If it's going to be 20, it's probably one of the most connected areas. If we expand the number of vantage points, I would say we then go down that road.

DUANE WESSELS: Also, I want to just remind about something that we've talked about in the work party in the past. Initially, we struggled a lot with the idea that root server operators would be held accountable to service level expectations for parts of the network that are out of their control. That's one of the reasons that we proposed that the vantage points be located at well-connected data centers: to eliminate that concern. If the vantage points were, as you say, at these far-away places or even in people's homes, then, to the extent that we're building service-level expectation metrics here, we're holding operators accountable for things that they can't control.

So this was the balance that we struck. Vantage points aren't right next to the servers. They're a little bit far away, but they're not really far away. So that's the balance.

Liman, you had a comment that I interrupted? Sorry?

LARS-JOHAN LIMAN: I was just thinking that the fact that we may use 20 vantage points for the accountability thing doesn't preclude doing the same measurements from other vantage points to investigate expansion of the network for future [inaudible]. So there's no direct contradiction here, I think.

DUANE WESSELS: Yeah, that's a good point. My intention in what the work party is producing here is to write something that anyone could implement if they wanted to. I'm assuming that there will be an official metrics

system operated by someone to be determined in the future. But you could have a separate system running for similar purposes for whatever, and you could use the same techniques. Or not. But this should be generally applicable to anyone that wants to take advantage of it.

All right. Should we go onto the next one? By the way, these are in the same order as they are in the document. The third RSO metric is called correctness by DNS validation. Now, unlike the previous two, there are not separate metrics or it's not reported separately for each transport. Instead, the idea here is that the transport for a measurement in a given interval is chosen at random. There's a little table there on the right, which describes the nature of the query. This also is chosen at random with non-uniform probabilities, as you can see here.

So the idea is that, half the time, the query is for a top-level domain's DS record. For 20% of the time, it's an SOA query for the roots; for the 20%, DNSKEY query for the root, and with 10% probability, it's one of these random NX domains of type NS. So the idea there is to get good coverage of the content of the zone and the types of responses that are commonly seen.

This metric or this measurement uses a trust anchor. It also requires having the root DNSKEY to validate response signatures, and -- I know, Paul.

PAUL: [inaudible]

DUANE WESSELS: And the metric is defined – over the course of the day, you would count how many responses that are validated as secure and divide by how many total responses you got. So it's a straight percentage.

One requirement of doing this is that the vantage points would need to keep the trust anchor up to date. So, I believe, as currently proposed in these measurements, the validation is done on the vantage points. I'm guessing that … Is that – yeah. Paul was raising his hand because he and I were discussing that maybe it makes better sense to have some sort of centralized doing the actual validation. Then you don't have to have the vantage points keeping the trust anchor up to date and doing the validation.

Go ahead, Paul. Why don't you state your case?

PAUL: I assume we're going to discuss this later today, right? As we go through—

DUANE WESSELS: Yeah.

PAUL: One of the thing important though is that, even if we decide to have all the validation for this and the matching in the next one being done by the central point, who's going to be doing the reporting anyway, the vantage points need to have an up-to-date list of the TLDs for them to do the fourth one. So it's not that the vantage points can have no

current state. They still need some current state in order to pick good TLDs. But they don't necessarily have to do anything other than use a TLD list. At least from the proof-of-concept one I'm working on, that works okay. So you push a TLD list to each vantage point so that it can make this. But you don't have to do anything else.

DUANE WESSELS: Right. That's a good point. I would also add that, even if your TLD list is slightly out of date, it still works, right? You validate the response you get, not necessarily the response you expected.

PAUL: Right. Even if your TLD list is a month out of date, you still would only have a 1 in 1,000 chance of having picked a new one anyways.

DANIEL MIGAULT: I'm just wondering what is the purpose of these measurements? Do we intend to evaluate that it's the right zone that is being published? Or do we assume that the RSO or something in the middle is not altering the zone?

DUANE WESSELS: In the document, in each section of the metrics, it has a little sentence that says what the purpose of each metric is. For this one, it says the purpose of this metric is to characterize whether or not a single root server servers correct responses. So, in this case, it's not necessarily about being an up-to-date zone. Given the way the root zone is

currently designed and signed, you can still get a response that validates even up to a week after the zone was no longer updated. So you can get stale-ish data that's still correct from the DNSSEC point of view. That's what this metric is about.

Does that answer the question?

DANIEL MIGAULT: Basically, we are not trying to evaluate that one actor is modifying the responses, not providing and response from, even if it's an old zone? Let's suppose one actor is redirecting a .com to another server. That's not something we're trying to check?

DUANE WESSELS: I'm not sure what you mean by "actor." Do you mean root server operator or some third party?

DANIEL MIGAULT: Anyone.

DUANE WESSELS: Theirs is a built-in assumption here that a response from a root server flowing through the network to the vantage point has not been tampered with. So, in my mind, this is really designed to prove that the server sent the right data, not necessarily that the network delivered the right data to you.

DANIEL MIGAULT: Okay.

DUANE WESSELS: The reason this is important is because we've gone through an exercise of asking people to begin to suggest thresholds. I think every threshold that's been suggested for correctness is 100%. 100% correctness all the time. So, if there is an on-path attacker who's able to insert a fake response that does not validate, that could affect this metric.

DANIEL MIGAULT: Mm-hmm. But my point is – go on, Brad.

BRAD VERD: This is not a man-in-the-middle detection system. This is a correctness: "Am I getting the correct answer from whom I ask?"

DANIEL MIGAULT: Yeah.

BRAD VERD: That answers your question, I think. So this is not a detection system to find a man-in-the-middle attack.

DUANE WESSELS: Now, I will say that there is text in this section that says, "In cases where correctness fails, those particular responses are, of course, logged and

kept for forensics," so that someone can go later and try to figure out what actually happened.

DANIEL MIGAULT: Yeah. But the threat I have in mind is you have an RSO and they're not serving some TLDs, which means, when it receives the query, it's not sending back the correct response. I don't think we have a chance to capture that if we don't take all the TLDs or if we don't request any – throughout the day by some way, if we … Because this kind of threat is not going to be randomly distributed into the root zone. It's going to be focused most probably on some TLDs.

DUANE WESSELS: That is the threat that this is designed to detect. Now, to your points about whether or not you would actually detect it, that's a tricky question because I think, if someone … First of all, due to our small number of vantage points, we can't hit all of the instances. The measurements are infrequent. The source addresses of the vantage points are probably known, so you can give the right answer to the vantage points or the wrong answer. So there's lots of ways you could cheat this system.

I think we're doing as good as we can with this metric. I don't know if we want to be super, super aggressive and try to catch all those cases or not.

DANIEL MIGAULT: I think, in the document, we should clearly identify the limits of the measurements. That's my …

PAUL: A question. Why, for the random NX domain solicitation, are you choosing Q-type NS [inaudible]. An NS query of a non-existent name is something that no actual resolver is going to do other than a fairly recent one that is doing q-name minimization. Wouldn't you like to an A or Quad-A instead?

DUANE WESSELS: Happy to change it. In our document, it said TBD for about six months. Before this meeting, I said, "I can't go to this meeting with TBD," so I put NS.

PAUL: My recommendation is to try not to look like a tester for at least some of your queries. www.randomsecondlevel.random.toplevel should get you a DNSSEC-signed NX domain. But this will be the most common case in which you won't, whereas, with an NS record, you will often hear did air because the middle box doesn't see a reason why you need to know that. Or you will say that they will say, "Since this is an NS, it's not an opportunity to insert the address of an advertising server." They'll let you through, even though they would not have let you through had you been asking for address.

So, in this case, I would suggest looking for a Quad-A or an A or maybe choosing one of those two address types in random – in any case, doing something that some real recursive would do. Thank you.

DUANE WESSELS: Paul, I have a proposal on this. One of the reasons why I asked Duane earlier, "Are we going to talk about the earlier sections," is that I have a proposal basically saying, not with the same solution you have, "We should look at this and come up with a different way of doing this." So, once we get to the section later in the agenda, I would really like to bash on this.

BRAD VERD: Liman has his hand up.

LARS-JOHAN LIMAN: Going back to what Daniel talked about, I agreed that we should limit ourselves here to something that we can implement today. I think it might be a good idea to mention somewhere in the latter part of the document or in – what's it called? – an appendix or something that we have realized that there are some issues with security that we hope will be addressed in the future but that this is a first implementation to get something going so that future generations will look at it and say, "Why didn't they do this?" Because we put a limitation in here.

RUSS MUNDY: I was going to add or comment on what Daniel was saying earlier, and that is that at least it's my belief that our document currently doesn't have an explicit requirement for checking the totality of all the TLDs. I think the concern of having just some of the TLDs be changed and not others is one to be thought about. But, in doing it in a way that's doable at this point in time, we probably do need to limit to what is feasible at this point. I like Liman's suggestion of having some kind of notes somewhere that said, "Perhaps at a later point, stronger enhancements would be added."

Th other comment I was going to make is that, even though this is pointed at identifying what's coming from the RSO operator for content, there's no way that this testing system can guarantee that, if a change occurred, it wasn't done by some man-in-the-middle attack. So that's probably something worth noting in the document, too.

NAELA SARRAS: I think I'm hearing that we have a concern that we don't always have a list of [inaudible] system – the root zone – at any given time. No?

UNIDENTIFIED MALE: No.

NAELA SARRAS: No? Okay.

UNIDENTIFIED MALE: We are testing against them.

NAELA SARRAS: Oh, we are testing against them. Oh, we're not testing against all the TLDs, right?

[PAUL]: Right. I think Daniel's concern was, especially if we have 2,000 TLDs, is that we're only testing five minutes and there's only a 50% chance we're not going to be hitting them all the time. But there's no concern about getting the whole list of TLDs, either from you or just pulling down a root zone and saying, "What looks like a TLD in this root zone?"

NAELA SARRAS: Okay, good. Thank you. I just wanted to make sure. Thank you.

DANIEL MIGAULT: Just a last comment. If we put some limitation, I think it's good to keep that measurement just to avoid giving the false impression that we're doing the full test. If we go to the garage, it's not testing everything in your car. It's doing some checkpoints. That's what we want to provide. I'm happy to provide some text on that.

DUANE WESSELS: Thank you, Daniel. Hiro?

HIRO HOTTA: Sorry if I've missed your explanation, but about the probability, it is intuitive numbers or are these from some statistics?

DUANE WESSELS: Intuitive, not from statistics. Just made up based on what seems like a reasonable distribution.

All right. Let's move on to the next metric. This is very similar in the sense that it's also checking correctness.

BRAD VERD: Sorry. I hesitate to say let's bring this up, but I'm going to do it anyways because I emphasized doing that in the beginning. A couple things that we've touched on here and we've accepted as okay: not validating every TLD and not seeing every Anycast instance. These are all things that we're saying are okay. But then, in the end, we're like, "Well, there's a possibility that we're missing something." Then we should document it, saying that these should be maybe reviewed later or expanded or whatnot.

I guess the thing that runs through my head – don't throw things yet – is I got back to this self-reporting that I've talked about in the past. I'm jumping way ahead now. I'm going to speak out of turn. If and when we ever get to some SLAs, those SLAs could be self-reported. Every Anycast instance could be self-reported with whatever tools that would need to be developed. Every instance could walk the zone and do the correctness and do a self-reporting of these metrics.

So, to me, I think, when we talk about adding something to the document, that's what she be added to the document: something like that.

DUANE WESSELS: So you would like to see the document have something about future work for self-reporting to get more coverage and to get –

BRAD VERD: Well, if we're going to call out that these are gaps, we should say that, because of the design of the Anycast and the way it works, no matter how many probes you have, you're never going to hit every instance. The only way to guarantee that we're getting an accurate picture is self-reporting.

[PAUL]: Just to lob a hand grenade into that discussion, the other way to that is to have every RSO report a Unicast address of every instance that does answer DNS queries

UNIDENTIFIED MALE: ["Grenade" in quotes.]

DUANE WESSELS: I'm not sure Brad heard your question.

BRAD VERD: You guys both said the same thing.

[PAUL]: Exactly.

BRAD VERD: While I think that is technically feasible, I'm not sure that that is technically reasonable.

UNIDENTIFIED MALE: Mm-hmm

[PAUL]: I was just objecting to your characterization of the only way of being self-reporting.

[DUANE WESSELS]: So noted.

BRAD VERD: So let's put of those in the document.

UNIDENTIFIED MALE: Mm-hmm.

UNIDENTIFIED FEMALE: [inaudible]

DUANE WESSELS: All right. I heard someone on remote, but I think they're not trying to ask a question. So let's move on – oh, and I'm sorry. We didn't get past this one yet. So this is another metric for correctness.

UNIDENTIFIED MALE: [inaudible] break? [inaudible]

DUANE WESSELS: 10:30.

UNIDENTIFIED MALE: Oh, I'm sorry. [inaudible]

DUANE WESSELS: So one problem with doing correctness by DNSSEC is that some parts of some responses aren't signed with DNSSEC. So glue records are not generally signed. If you wanted to do correctness checks on those, you need something like what's proposed here. The idea here is that you send a query for data in the zone and you compare the data that you get to the zone itself, to some authoritative copy of the zone or whatever. This does not rely on DNSSEC validation, although you could consider using this technique to also check signatures and so on.

As currently proposed in the text, I believe, this metric references what we're calling a central processing system. That is where the actual determination is made. For example, a vantage point would issue a query, receive a response, and then it would have to transmit those full query and responses back to this central processing system, which would do this matching detection.

In order to do this, we believe you need to keep some history of zone files because you don't want to get in a situation where a root server has a different version of the zone than the central processing system. So it's currently proposed as two days. The central system keeps two days' worth of zone file history. When a response comes back, you would have to look in all of those zones to see if one of them matches the response that you got back.

Again, as proposed, this is a little bit different than DNSSEC because here you can get a valid match-up to two days in the past with correctness. With DNSSEC, you could probably get a valid matchup to seven days in the past.

Now, one big open question here is that you might consider this correctness by matching as just a superset of DNSSEC. Does it make sense to have both or not? Any correctness that you can do with DNSSEC validation you can probably do equally as well, in theory, by correctness matching.

Personally, I like the idea of DNSSEC validation, but I can't argue that there are things that you can do with DNSSEC that you can't do with correctness matching.

So they're both here for now. If people think that we don't need both, then that's certainly discussion that we'd like to have, probably not right now in this session but later on in the meeting here.

This was your baby, Paul. Is there something that I missed about this?

PAUL: I think the biggest one is the fact that you can actually test for NS. For DNSSEC, the reason why you do the DS records is the fact that the NS is not going to come back in the answer. It's going to come back in the additional. Here, if you're looking at everything, you're really looking at everything. At least my logic here was that this goes to the root of why are we testing correctness at all, and, again, not thinking about the man-in-the-middle but either a root server operator or an instance operator going rogue. By the way, going rogue could be due to a government demand. An instance in a particular country is not allowed to say anything about X, so they say something different. They may so it not just an NS record but in some tricky way. We want to know if the resolvers are getting 100% correct. So this allows us to check everything in the answer, in the additional, and in the authoritative sections.

DUANE WESSELS: Other than, the aggregation is the same. It's just straight percentage: number of matches divided by number of responses over the course of the day. Again, for this, since it's one – yeah, it's the same number of measurements as the others, I think. It's 5,760, right?

Daniel?

DANIEL MIGAULT: I think it's really down to testing NS. It's not about the correctness of the zone—

PAUL HOFFMAN: Actually, let me just jump in and pretend I'm Paul Vixie for the moment. No, it's not because we could actually have queries for www.example.com and see whether a rogue instance is – for an A-record and see whether a rogue instance is responding authoritatively or unauthoritatively with the A-record. So this really is matching against any of the contents in the root zone. So, if the one of the tests was for www.example.com A or Quad-A, we could easily see if there was a positive answer. Well, that's not in our root zone, but these queries could really be anything. And, as we've all seen, anything gets sent to the root servers. We want to make sure that the root servers are answering from the root zone.

LARS-JOHAN LIMAN: Liman with a nefarious comment. That's already being tested by Google Chrome a gazillion times a day.

[PAUL]: But not very effectively.

DUANE WESSELS: Yeah, Daniel So, if you haven't seen it, in the document there is some rules for you do the matching. For example, it says, if the [R code] is zero, then the answer session should be empty and so on. So there is kind of that [inaudible].

RYAN STEPHENSON: Quick question about the query setup. Is it going to be – I'm sorry; I'm only using BIND in my head – but would it be just with a no-recurse option? So that way it's just, "Hey, just give me what's in the zone. That's it. Done"?

DUANE WESSELS: Yes. The idea is that an implementation of these metrics – they wouldn't necessary use the BIND software. They would be issuing queries themselves with the appropriate flags. So just receiving the answer that a recursive name server would also receive.

RYAN STEPHENSON: Thank you. And, yeah, I just wasn't specifying just BIND. I was just thinking that plus no-recurse. That's where that came from. Thank you.

DUANE WESSELS: All right. We might be getting a little short on time, so let's try to keep going here. The last metric defined for an individual RSO is the publication latency. This one is one of my favorites because I think it's really, really complicated. This one also reuses the SOA queries from the availability metrics. Every five minutes, there's an SOA query. It's great

because we have all these SOA responses, which could contain a serial number. If you envision shuffling all those responses back this central processing system, then you can detect when new root zones put out, either by looking at those serial numbers or maybe you have another way of knowing when a new serial number gets published, when any root zone gets published. You can perhaps receive notify messages or something like that. So that part is not strictly specified how the central processing system knows that a new root zone is published. There's a couple of options, but there's no strict requirement there.

With this metric, the idea is that – for example, on a typical day, there are two new root zones published. Each time there's a new zone serial, that's one set of measurements. So, when the system detect that a new zone is published, it starts a clock or starts looking and calculates how long until all of the probes saw that new version. On a typical day, that would happen twice or maybe three times, so there would be something like 40 or 60 measurements from all the vantage points.

[PAUL]: No, there's measurements every five minutes. There are two or three metrics that come out during the day.

DUANE WESSELS: I disagree.

[PAUL]: We should have this discussion later then.

DUANE WESSELS: Because, in my mind, in all of the metrics, there's just one metric per day that you would compare against thresholds. For example, if there's a day with two root zones published, you would aggregate those two into a single metric. If there's a day in which three root zones were published, you would aggregate those three into a single number for the day.

[PAUL]: Then we're agreeing.

DUANE WESSELS: Okay, yeah. So here's another case where we're using medians. The publication latency is defined as the median of all the vantage point measurements for a single day. Also note that, since these queries are only made every five minutes, that's your resolution for how long a zone takes to pop out. So it could be five minutes, ten minutes, for a single probe root server instance. Again, because we're going be taking median values, it's going to be one of those. It's going to be, like I see, a zero, a five, a ten, or some multiple of five.

This is a little bit complicated. You might want to read through the actual text of the document. I think it goes into a little bit more detail about how the central processing system would do this, when the timer starts, and so on. Also, I think it's important to note that, whereas in this metric we're proposing to use a median value as what you compare the threshold against, in RSSAC002 what get's reported in the 95th

percentile of a similar measurement. The RSSAC002 is self-reported – how long the root server takes to publish the zone to all of its instances – and there it reports 95th percentile. But here we're proposing a 50th percentile median.

By the way, this is probably a good time for me to mention one of my ideas. If you don't know, Paul has been prototyping all these measurements and collecting data. We haven't done this yet, but I really would like to take Paul's data and put into a GitHub repo for people to look at and then, for some of us, take that raw data and calculate these metrics from it and make sure we all understand how this should be calculated and get the same answers out. I think that's maybe future work for this work party, or at least for Paul and me. I think that would be important: to have a reference data set that we can all understand and get the same numbers from before we start doing this in production.

PAUL: One of the reasons why I think that's actually valuable was, when we started the work party and we were talking about publication latency and there was the question of, "Five minutes is too long," and such, once I started doing the numbers, we found multiple RSOs over days often taking 20 minutes. So, when people were saying, "No, we need much less than five minutes," that would have excluded the last instance. Only checking out of eight, a bunch of them actually didn't update for almost 20 minutes. So it was good to actually see real data on that.

DUANE WESSELS: All right. Let's plug ahead. The next section is about RSS metrics. Here there are four, essentially the same as the RSO metrics. But the one that's missing is the last one we just talked about. There is no root zone publication latency for the RSS as a whole.

Next slide, please, Ozan. RSS availability. I think now is not the time to discuss this because we're going to have a discussion later. Currently in the document I think the RSS availability is incorrectly specified or not well-specified. So I'm going to propose changing this. We'll have that discussion later.

Let's go on to the next one. RSS response latency. I know this one is a little bit controversial as well. Maybe we'll end up changing this. As currently defined in the document, this is a single aggregation of all the measurements from all the RSOs. Whereas, for an individual root server operator over the course of day we would have about 5,760 measurements, in this one, when calculating response latency for the whole system, we have almost 75,000 measurements. They're all just bundled together. You take the median and that's what you use as the response latency for the root server system.

I think, Daniel, your suggestion was to use other than median – or was it Hiro? I'm getting confused. One of you two had a proposal for using something other than median in this case, I think.

UNIDENTIFIED MALE: Actually, it was Shinta.

DUANE WESSELS: Oh, Shinta. That's right. He's not here. Okay, thank you. So we can have a discussion about whether median is the right approach. Or maybe we need something more sophisticated here. I don't know.

UNIDENTIFIED MALE: Shinta is online.

DUANE WESSELS: Oh, Shinta is online. Can we hear him? Does he want to say anything?

SHINTA SATO: I'm here. I suggested that, [just because] we're getting the measurements for all these measurements, that does not show the meaning [from] the resolvers. So I asked to change this one because the resolvers will choose one of the fastest responses from the 13 or 26 root servers. The performance of the response latency would be showing that kind of measurement. That's much better understood from other parties. So that's why I suggested that in this kind of measurement.

DUANE WESSELS: Thank you, Shinta. I don't have super strong opinions about this, I guess. One reason that I can imagine keeping it as median is because, earlier in the work party, we talked a lot about the purpose of the metrics and what are we measuring. Shinta I believe is correct in that a typical

recursive name server would probably select a lower latency server for the majority of its queries. That's the way we expect most recursive name servers to work. But if we start to design our metric around that, then we're making assumptions about how all recursive name servers work and we're starting to design the metrics based on recursive name servers rather than root server expectations.

BRAD VERD: Yeah. I think we're not measuring recursive behavior. The goal here, I think, the metric, is to measure the latency of the distribution of the zone.

DUANE WESSELS: Right. That's the struggle we have. I'm a little bit sympathetic to both points of view and I don't know what's the right way. We don't have to solve this right now. We can discuss this later on in the session.

Quickly.

[DANIEL MIGAULT]: Yeah, very quickly. When you buy a car, usually the car can speed up 200 kilometers per hour. It doesn't mean you drive 200 kilometers per hour all the time. So I don't know what's the answer for that. As long as we understand what we clearly mentioned, I think that's – because we will never be able to have the right answer, I think, at the end.

DUANE WESSELS: Okay. Let's go to the next slide, Ozan, please. Again, in the metric of root server system correctness by DNSSEC validation, the document is proposing, again, a simple aggregation of all the measurements from all the server operators. The question in my mind is that, if – this is maybe a big if at this point because we haven't talked about thresholds. The direction we're heading is that we're proposing a correctness threshold out of 100% for root sever operators. Then what's the purpose of this metric? If all operators are going to be held to 100%, then of course for the root server system it would also have to be held to 100%. It's sort of obvious, right? That's another thing we can discuss later on in the week.

Can you go to the next one, please? This is the same. This is just applying the same technique to the correctness by matching metric. It's a simple aggregation of all the measurements from all the vantage points and all the RSOs.

[PAUL]: Since you're about to go to the next slide, which doesn't have what I was hoping, why, Duane, did you not have an RSS publication latency aggregation? Because, going back to what Shinta said, we're going to have a hard time describing what the RSS metrics mean anyways. I think we'll have no harder of a time describing that one. But, since you didn't even put it in, what was your reason?

DUANE WESSELS: Well, it was in at one point and it got taken out. It's been so long ago that I don't remember exactly the reason. I think, if I remember correctly, we were sitting around a room in person and struggling to

come with the idea: what does it mean for the root server system as a whole to have publication latency? We just had a hard time.

[PAUL]: Again, I will repeat. I think it's just as hard to describe that as some of the other ones. So, if we are going to struggle with the other ones, we might as well struggle with it as well because someone reading the table of contents is going to see a mismatch and have the question of why is that not there.

DUANE WESSELS: Yeah.

WES HARDAKER: Well, I think there's another reason, which is that, if you are – if everything is 100%, yes, you don't need it because they're going to align. But the reality is that, if there is a slippage at any particular point, then you want to know what's the percentage accuracy of the whole system? So having that metric and knowing, well, even though this operator is maybe [50%] bad, what's the effect on the whole system? That's going to be a different number. It's not going to be 50%. It's going to be based on how many instances are hosted by how many different places. It gives you a feeling for the difference between one operator versus the whole system when something doesn't work.

DUANE WESSELS: Wes, yes, but also keep in mind that – we haven't really talked about this – we're proposing, when the reporting is done, that it's only a pass/fail reporting.

UNIDENTIFIED MALE: No, that's for thresholds. We're not talking thresholds here.

DUANE WESSELS: Okay.

WES HARDAKER: Somebody was too quiet to hear.

SURESH KIRSHNASWAMY: In general when we come up with any numerical estimates of something, it's very difficult to just state a number and say that's a metric or a measure. You generally need to have a measure and some notion of spread to see how much it's varying. Sometime the spread itself becomes a metric in terms of when, when we're talking about a system-wide property of "This is the latency," we're not so much interested in what is a specific number. It is how variance there is amongst that system that defines what is latency. So sometimes just treating that variance or standard deviation as a metric might be useful.

UNIDENTIFIED MALE: [RSS section].

DUANE WESSELS: Okay. Next slide, please. At the end of the document, there's a section on recommendations. Some of these we've talked about already. It talks about the number of vantage points, their locations, and their connectivity requirements. There is a recommendation that an implementation of this or the official implementation of this should be open-source software so that everyone can see it and have confidence in its operation. Thresholds are going to be a big part of the recommendations. We hope to have that discussion in the next few days about thresholds.

As we were just chatting about a minute ago, when results get published, the idea is that you publish only whether or not – it's only sort of a pass/fail – the threshold was exceeded or not exceeded. As we already talked about as well, there's a recommendation to investigate ways to improve the distribution of vantage points in future work.

Let's go to the last slide. There's a long section of example results. This is just one of them. This is the RSO availability threshold. This is designed a little bit after the IANA report that Naela shared at the last [workshop, I think] or one of the previous meetings. For example, we've got the RSO availability metric for the four different transports. The green and red indicate whether or not the threshold was met or not. This example is designed to be a little bit ridiculous so that we don't get too caught up in the specific of the numbers. Here the threshold is 15% availability for UDP and 12% for TCP. Then, as a way of expressing precision, there's also a column that says how measurements were used in deriving this particular metric.

In these document, these exist for, I think, most of the metrics, just as an example of how they might be reported. We don't have to report them this way, but this is just to give you a sense of what the final output might look like.

That is the last slide in this deck, so we've got five minutes for questions if we want it. Do you want to spend some time talking about your alternate to the randomnxd stuff?

[PAUL]: Well, do we have a slot for that?

DUANE WESSELS: We do not.

[PAU]: Five minutes is not sufficient.

DUANE WESSELS: Okay. All right. So we'll have to invent a slot for it then.

DANIEL MIGAULT: One of the questions you asked is, do we need RSS metrics if we have the equivalent for RSO?

DUANE WESSELS: You're talking specifically about the correctness one? Whether 100%?

DANIEL MIGAULT: Yeah. And basically all the metrics. I think, if we have the measurements or metrics at the RSS level, it's where we most care about [it] than the RSO ones, especially if we want to show that the system is working well and how the impact of one RSO is impact the full system. So I think it's important that we get exactly the system one level below. They may have some different thresholds, even in the future.

PAUL: Daniel, I would completely disagree because we are only publishing thresholds. We are not publishing the metrics. We already agreed on that early on. If we publish thresholds for each RSO, you're proposing we actually publish metrics for the RSS. Unless we're going to come up with thresholds, which is an argument that I had later of, what's a threshold? It's not like any is going to say, "My God. The root server system is not meeting its threshold. We are going to throw out that root server system and create a new one." Since we know that's not going to happen, I believe we probably won't have thresholds for the root server system. So, with what you're proposing, that would actually be a separate publication completely. We can do that. That's not hard to. It's just numbers. But that should be completely different in this document. If we're having RSS metrics for publication, that's really, really different than RSO metrics, which are not for publication. They're on for the threshold values.

DUANE WESSELS: Yeah. This is getting into the topic that we've reserved for the next whole session, so I suggest we save it for then because we're going to spend 90 minutes talking about it.

I guess, if there's no other quick questions, we can take a break.

BRAD VERD: Carlos, anything? Are we okay breaking three minutes early?

All right. Let's do a break. We're back here at 10:45. To those of you who send me an e-mail, I'll go work it. Thanks.

AUTOMATED VOICE: This meeting is being recorded.

DUANE WESSELS: All right. We're back with the second session to continue talking about the metrics. In this session, we wanted to cover two overarching meta questions, which we touched on a little bit in the previous session – one of them. The two big questions are, should the work party make recommendations for RSS thresholds? There's general agreement that the work party will make recommendations for RSO thresholds, but there is not yet consensus on whether we should make recommendations for RSS thresholds. We can spend about half the time in this session talking about that.

The other big question is, do we need two sets of thresholds? The document pretty clearly talks about the need for specifying thresholds

for minimum performance levels. But also, when we talk about metrics, sometimes we like to talk about or think about what I would, for lack of a better word, call good performance levels. So, rather than just the bare minimum, what should be aspire to? What levels of performance should we try to achieve as wanting to provide good service? So the second one is a little bit more nebulous, and at this point, we don't really have a better way to describe it than to. "Say good performance levels are what good looks like." But we need to settle this and decide if this work party should have both types of thresholds.

So that's this session. I'm going to suggest that we just go in order of these two questions. So first, I think we should talk about whether or not to have recommendations for thresholds for the RSS. I'll open the floor.

BRAD VERD: I'll go first. My answer is yes, emphatically, for availability and latency. The other ones I'm not so sure about. If we all say correctness is 100% for an RSO, then maybe we should say the document is that, just by course of math, the RSS is 100%, so we're not going to measure that. 100%.

Does anybody disagree with that?

Paul?

PAUL: I think I may agree, but you would have to tighten the language. If what you mean is that all questions that need answers –so not DDoS packets

but all legitimate cache misses from recursive servers get an answer from some root name server so that no transaction fails anywhere in the world through the lack of ability to reach one of us or to be able to reach at least one of us, then I'm with you.

If you mean that everything has to work all the time, whether it would have a service affecting impact if it didn't or not, then we have a longer conversation.

BRAD VERD: Well, I think we're in violent agreement on the first part, which is that every legitimate packet needs to have an answer. That's what my statement was about: that the root server system should be available. So, if you start with that premise, it's 100% for availability. Then maybe in the document you put in some caveats, which is that, under DDoS, under certain conditions, there is the exception. This 100% might come into question. But I hope that list is small and concise.

PAUL: Well, again, I want to make sure we're using language with enough detail that we know whether we disagree. I don't care about retries and I don't care about having to try another server. I would only are about transactions which fail. There must be none of those. But as to how many times they have to ask us a question over a course of seconds or how many of us they have to direct that question to before they finally get an answer, I wouldn't care about any of that because that wouldn't change to me that it was 100% available. So, yes, we can add various caveats about things that might cause certain transactions to take

longer, but all valid transactions getting an answer is my definition of 100% available.

BRAD VERD: Again, for clarification of words, you said it didn't matter how long. That means it could take me 24 hours to get an answer and that's still available?

PAUL: No. It has to be in a few seconds before the underlying transaction that caused the DNS cache miss itself [to] time out and treats the unreachability of us during that period as the root cause of that failure. So I don't want to be the root cause of any failure anywhere, but that means that the intermediate server may have to retry a couple of times. It may have to try different servers because there may be something going on. But, as long as we give an answer in a single-digit number of seconds because the recursive server tries hard enough, then I don't really care what happened in that handful of seconds because ultimately the underlying transactions succeeded, sometimes with a little bit of delay and complexity, but did not fail. As long as no transaction dependent on us fails because it can't get an answer from us, I treat that as available.

BRAD VERD: Again, we're getting into a bit of the math of it, but, since we're here, we'll just keep on going. The monitoring intervals are five minutes. Is

that interval okay for your single digit? Or is sub-ten second your single digit?

PAUL: I think we can't put a monitoring load into every interval. I think that the quantum of availability is the wrong measurement window, even if we space it out so that we're always asking some Anycast node some question so that, at any given time, we were testing something. There would still be long periods of time where the first server some people were going to use was, at that instant, not reachable and we didn't know it. To fix that, we would have to burden the system with so much extra traffic that we'd only do our own cause more harm than good. So, no, I'm fine with five minutes.

[BRAD VERD]: [inaudible]

DUANE WESSELS: Yeah. As Brad alluded to, we're already trying to design the measurements here. I just want to keep us focused on the question of should the work party make recommendations for RSS thresholds or not.

[FRED BAKER]: Well, I was going to ask what an RSS threshold means. I don't think we should try to specify things that we're not real clear on what they mean. If I can't get to F-root at all, I kind of know what threshold means. But, if

I'm trying to get to the entire RSS, no server among the thousand of them or however many there are … So what does that RSS threshold actually mean? How do I interpret the results?

DUANE WESSELS: Let's assume that we have a defined metric, a defined way of calculating RSS availability. Let's fantasize for a second and say that we can all agree on that.

[BRAD VERD]: Hypothetically.

DUANE WESSELS: Hypothetically, this group agreed on the definition of a metric for RSS availability. So we can do measurements and we can get a number out at the end of the day. Now, the question before us is, should this work party propose thresholds on that number and report them as, for example, pass/fail?

[FRED BAKER]: Okay. So suppose that it fails. It took eleven seconds to get Paul's answer. What would be the interpretation? What do we do?

DUANE WESSELS: I think we all agree that there will never be SLAs for the RSS as a whole. There's no one that you could hold accountable for that. So, in a way,

it's just informational only. We would have the knowledge that, on this day, this threshold was not met. But it's not necessarily actionable.

I know Paul's been very patient. I know he has a lot to say on this.

[BRAD VERD]: Actually, Rob has his hand up.

DUANE WESSELS: Okay. Go ahead, Rob.

ROB STORY: Well, I think this gets back to defining whether or not this means that all of the RSOs or whether or not the [an of m].

DUANE WESSELS: Let's assume that we've solved that problem.

BRAD VERD: Can we turn your mic off when you're not talking? Because it only …

DUANE WESSELS: We'll have that discussion, but let's pick one that's easier. Maybe it's latency or something like that. So we've got the metric defined. The question is, do we want to make recommendations on the threshold for the RSS?

UNIDENTIFIED MALE: Who was next?

UNIDENTIFIED MALE: Liman was.

LARS-JOHAN LIMAN: Okay. A couple of things. The first one is that I think that, yes, we should specify such things because, if this work party doesn't, who will? Would you have to be pushing this in front of us, trying to give the ball/hot potato to someone else? I would rather see that we have discussed and tried to either reach the conclusion that, yes, we will and this is the number, or, no, we won't.

I think having that information might be useful because, if it does take eleven seconds for the system to deliver this answer, then we have a systemic failure, a system problem, that the strategic part that we've been talking about in the RSSAC037 context should be notified and possibly do something about the entire system.

DUANE WESSELS: Yeah. Either we have a systemic problem or we have a failure to define it correctly. But, right.

UNIDENTIFIED MALE: Fair enough.

PAUL: I don't think we should actually specify a threshold because a threshold is used by somebody looking at the pass/fail of the threshold to so something. As you just said, and I think everyone here agrees, nobody is going to say, "Look, they failed to meet the threshold. We have to throw out the root server system." I think it's fine to publish metrics. At this point, given that your imagined world where we agree on how to collect those for the doesn't-exist-yet, I think it's fine for us even that world to publish metrics and maybe later come up with a threshold where we said, "We agree with the world that, if we didn't meet this threshold, we failed, even though that failure doesn't mean anything." But I don't think that coming up with thresholds now is at all meaningful to the world. Publishing metrics is meaningful.

As a separate comment—

DUANE WESSELS: Can I ask you something? When you say publish metrics, do you mean public the numbers, the percentages, or the formulas? You're talking the numbers the values that come out?

PAUL: The values that come out, yeah. As a separate one, so far we have only been discussing availability. Any latency discussion we have immediately bashed its head against our RSSAC042, which allows any root server operator to run their root server system exactly how they feel. Given that was a very definitive thing that everyone said that came

out, us saying that the root server system is failing a metric for response latency when that failure is a direct cause of RSSAC042 being implemented and everyone saying, "This is how I want to do this" I think undermines RSSAC042. I think we might be able to eventually come up with a threshold for availability but not for response latency.

BRAD VERD: I had my hand up. Again, I disagree. Maybe the term isn't "threshold." Maybe it's "expectation." I don't know what it is, but I think it's 100%. And I disagree with the fact that it's not actionable. If we failed the 100%, it would come back to this group or to the work server operators and action would be taken. We'd be like, "Look, guys, we all said it was 100%. Right now, our tools say it's no." So it's something new to us. We're figuring this out. We don't know what happened. We need to figure out what happened. And action would be needed to be taken. That's how I interpret what would happen.

Going to what you said, I agree with you that it would go to the strategic function once 037, if it happen. But if it doesn't happen, it would come here. So I don't want to get hung up to do this for 037. The expectation in my eyes would be 100% availability – we can talk about the other ones – for the root server system and, should we fail it, it's this group that should take action. No, nobody would get fired, but this group would take action to fix it.

[FRED BAKER]: I don't think it is accurate to interpret RSSAC042 as saying every root operator can do just exactly what they darn well please. If that were the

proper interpretation, then a root server operator could turn off all of his equipment and say, "That's how I chose to run it." We might not go along with that. I think the point of RSSAC042 is that an external body cannot tell an RSO what to do, which is a whole different ballgame.

PAUL: Fred, let me just dive in on that. Do you believe that 042 allows a root server operator to place a lot of instances in places that are going to have horrible latency? That's really where I was going to on this. I believe that the no one in this room would think that 042 allows you to turn everyone off. But I believe that 042 would allow RSO to, for example, turn all of them off except for one for a bit while they're doing something.

FRED BAKER: Probably. As far as putting servers someplace where latency is going to be horrible, I would have to believe that that particular RSO was trying to optimize something for some reason.

UNIDENTIFIED MALE: [inaudible]?

FRED BAKER: Yeah. If he's doing that, that's probably reasonable for him to do. But that doesn't allow him to destroy things.

PAUL: I agree, but I think that comes back and affects some of the thresholds that we are discussing here.

[DANIEL MIGAULT]: One reason I would like to have thresholds on the RSS is that the RSS is what we care about it more than each RSOs. That should be the main trigger for any investigation on each RSO. Suppose, for example, that the RSS is not available for an hour. I'm sure that's going to trigger something like, "What are you doing? Even though you [inaudible], all the RSOs have been out for this hour, and it was acceptable, given the document, we will have to change the document."

[FRED BAKER]: Yes.

[DANIEL MIGAULT]: So I think we should not forget, Suzanne or Paul.

SUZANNE WOOLF: As far as setting thresholds for the system to meet, first of all, I think looking at 042 is a little bit of a red herring because it says specific operators do have their independence. But, at the same time, they can expect to be held accountable if they make decision that impact the performance of the overall system. The interplay makes [sense]. I don't think those ideas are contradictory. I don't think one is normative and the other isn't or anything like that.

But more to what I think is the point here, we live in a world in which people are going to interpret these numbers because they exist. Setting thresholds from this room, this group of people, the process we're talking about, just means that the people most knowledgeable about the system also are being to say what they think the thresholds ought to be. Somebody is going to impose thresholds because that's just the way the world works. Having what we think are good ones out there means that we're part of that conversation as opposed to being silent and leaving it entirely up to the rest of the world to interpret whether a given set of numbers is good enough.

DUANE WESSELS: Any other input from folks? Maybe folks who haven't spoken yet.

UNIDENTIFIED MALE: Do you think …

UNIDENTIFIED MALE: … answer your question. I'm not sure—

DUANE WESSELS: Well, yeah.

UNIDENTIFIED MALE: [inaudible] know.

DUANE WESSELS: I hear strong opinions on both sides, so I don't know how to get resolution on this. It's tough.

UNIDENTIFIED MALE: [inaudible]

[PAUL]: Can I suggest a way forward? When Brad said, "Well, this isn't just about RSSAC037," I'm sort of surprised because I read that in 2.1, not that it said it's all about our 037. But the paragraph before the two bullet points, to me, really sounds like this work was for 037. If we want to have the metrics work also inform RSSAC, then thresholds are completely reasonable. I would just say we just need to reword the limitations of this document to not be only about 037 because we can always tell ourselves whatever we want. I fully agree with Suzanne that, if a threshold is going to come out sometimes, it should come out of here. We are the appropriate people for that.

Again, I wasn't part of the 037 creation, and I looked at this and I said, "Oh, this is just about 037." And it goes back to what I said to Daniel earlier, which was that anything about the RSS, if we're not having a threshold – I mean, we can now, if the group wants, have a threshold, but publishing the RSS metrics – we're not going to publish the RSO metrics – I think could be something that this group wants to do, at which points throwing in a threshold will actually give those metrics some context.

DUANE WESSELS: Okay, thanks, Paul. A couple of things I wanted to respond to in your points. You mentioned some bullets points in the document. I think you're talking about Section 2.1, which is purpose of metrics and thresholds. I think you're pointing out a valid thing, which is that someone could read this and say, "This is all about RSSAC037 and it's not about these other things, so we need to do a better job of describing that in this section." Certainly, in my mind, this document was always about more than just 037. So that's probably my fault for not capturing that well enough in the purpose. So we can work on that text. I think that is a good way forward.

The other idea of publishing the actual metrics values for the RSS but not doing that for the RSOs I think is a very interesting idea. I could support that. I'd like to get other people's thoughts on that as well, if not now maybe later on in the week. I think that's a very interesting idea.

[FRED BAKER]: I'm going to be brave here and say I think we may have reached agreement that we should have thresholds for the RSS.

DUANE WESSELS: If we redefine the purpose a little bit or clarify the purpose.

[FRED BAKER]: Redefine the purpose. Obviously, though, one of the huge challenges will be getting an accurate description of what they mean and how they're obtained.

SURESH KRISHNASWAMY: Just one point of clarification. Are we treating thresholds to be synonymous with SLEs?

UNIDENTIFIED MALE: No.

SURESH KRISHNASWAMY: Okay.

DUANE WESSELS: I wouldn't say synonymous, no, because SLEs is a tricky term . I think it maybe means different things to different people. I guess I also depends on the extent to which you would say that someone could be held accountable for not meeting an SLE or something like that, right? Because, in this case, there is not entity to be held accountable for the RSS.

SURESH KRISHNASWAMY: Okay.

[FRED BAKER]: And SLE might be that the RSS meets the thresholds. But that would be the use of the threshold and the expectations that result. They're two different things.

DUANE WESSELS: Any more on this particular question, or should we move onto the other one?

BRAD VERD: [inaudible] couple questions.

DUANE WESSELS: We have one more question, 2, yeah. I'm not talking about the next section. I'm talking about the question about having maybe two sets of thresholds or just one. Are we ready to move on?

Okay. Again, the question here is – I think we already have a good agreement that the work party should recommend minimum performance level thresholds. These are the thresholds at which, if you do not meet threshold, then you get called into the principal's office or the PMMF gets invoked and thing start happening. You're evaluated in some way.

Now, in addition to that, should we also have recommended thresholds for good performance levels? Please, somebody, come up with a better description than just "good performance levels."

Fred, was your hand up first?

FRED BAKER: Well, I'm going to ask the same question I asked a little while ago.

BRAD VERD: [inaudible]

UNIDENTIFIED FEMALE: It was me.

FRED BAKER: What in the world does this mean? To me, the minimum performance threshold – if I'm doing the minimum … and maybe that's good. If I'm doing the minimum and that's not good, then what is it? That's the question I would ask.

DUANE WESSELS: I think it's a reasonable question. I don't have a great answer for it. I keep thinking of the movie Office Space, where you have to have a certain number of pieces of flair when you're working there. I don't know if you've seen it, but that's what comes to mind to me.

You could also maybe think about this. The work party did consider briefly, instead of a pass/fail approach, a traffic light approach. We had red, yellow, and green. So you're in the red zone if you're below the minimum. You're in the yellow zone maybe if you're slightly above. And you're in the green zone if you're way above. So that would be another way to think about this, but the work party did not end up taking that approach.

UNIDENTIFIED SPEAKERS: [inaudible] Go ahead, Liman.

LARS-JOHAN LIMAN: Thank you. I am with Fred. I think that we should not publish – what do you call it? – Level 4 for good because what we want to avoid is bad. And if we've said, "It's bad below this," then the rest is good or good enough. You run into the traffic light leading to the rainbow effect here and "Please get more and more complicated." I think we have enough problems in front of us to get this going, so we don't had to add more complexity to the system that we're deciding on at this point. This could be a future enhancement of the system if we find that the current design that we're working with here doesn't work or doesn't fill the needs or doesn't fill the future needs that may appear. So one step at a time. Let's not do the [inaudible] thing at this time.

DUANE WESSELS: Daniel, go ahead. Then I'll go next.

DANIEL MIGAULT: I think I agree with Fred and Liman that we should not try to define what is good. When I had the Excel sheets providing the level of what we think is good, I interpreted this as a way to define what we believe is not acceptable. So it just was another highlight of the same question.

The threshold, in my opinion, I think should be only about defining the minimal and the limit of what is not acceptable. I think that's all we should talk about because other things are going to be much more complex and maybe not so useful. So that's my personal opinion.

JEFF OSBORN: This might be an unpopular statement. There's a can we've been kicking down the road for as long as I've been around here, which is we are not defining this from scratch. We call come to this with this thing in place. So we are extremely reticent to come up with standards other than that which all of us can pass on a bad day. So, if we can simply recognize that we're trying to make the least performant thing on a bad way work, then I think at some point we need to recognize that because you wouldn't design it that way. But we're choosing to measure it that way. So I just wanted to speak the unspoken.

BRAD VERD: I'll add to that. I've been withholding and letting everybody get their opinions out. Maybe this is my fault. I think maybe I'm the one who has created the idea of two different numbers. This is where I will muddy the waters with 037 a bit, and that is a agree with Jeff. I think that what's being defined as the minimum is the least common denominator today. That'll be the minimum and life is good. I feel that's now what you would do if you were building the system today. The number would be entirely different. Things would look different. So what should good look like? Which is what I've been saying for four years now.

If I can't get that out of the minimum, at least what I think the minimum is – that's my opinion, and I get that – then there's a second number, and the second number is, what does good look like. I've been beat up by a bunch of people in this room when I've talked one on one with you about what does good look like, and where I ended up with was, down the road, let's assume – because we're all talking about. We're talking about the number that we don't want to talk about it, either 100% or

something else. We're talking about these numbers. We're just not going to document them as this is what good looks like or this is what we feel. So what I feel may be the way forward or the way to think about it is, in the document – I talked to Duane about this – maybe the work party states something like, "In the future, after the implementation of 037, and with the potential of SLAs (which we all knows means money), there's a different number."

[PAUL]: There's a number for good.

BRAD VERD: For good that comes with money, meaning, if you get paid for the service, the minimum is not the number you're meeting. It's something else. So there's the grenade.

UNIDENTIFIED MALE: [inaudible]

[PAUL]: I want to strongly disagree with something you said, Jeff, which was that the thresholds that we are proposing are based on letting everyone pass. I don't think that you read the mailing list then because a couple of us – I put out a set of rationale for what I considered to be the minimums, and that had nothing to do with letting everyone pass. The numbers that I proposed in the spreadsheet that was sent on were based on things that, quite frankly, when I did, I didn't know if everyone

would pass. I think, if we have a reasonable rationale for minimums, that rationale does not have to be "Everyone here passed" for a couple reasons. One is people might get worse with 042. You might say, "I'm going to let me latency go to hell (average latency go to hell) because I can." I think that that's a perfectly reasonable thing to do.

I very much agree with Brad that, for good, we don't need to do it now. We should allow it to happen in the future. And good might come out of measurements that are not the five that we have on the table now. There are other measurements that can happen that we don't feel are needed for minimum that in fact are different ones that would be for good. I don't think we have any idea of those now. So that's why I think that defining good now is vastly premature optimization. But knowing that it might come in the future.

BRAD VERD: Can I reply just really quickly?

[PAUL]: Yeah.

BRAD VERD: So we are in violent agreement. What I'm asking for is the document to state exactly what you just said, which is maybe this changes – we don't know – but, when you talk money, when you talk SLAs, it's different than what the minimum is.

[PAUL]: Absolutely.

PAUL: I agree with Daniel and Liman. Brad, I do agree with you about having something in the document about, if there is an SLA, then define what good would look like. But, going back to the purpose of this document, it was again to set the minimum number so that, this way, we could find out if there was any issues with the RSS that are happening. Maybe, for example, six of the RSOs are only hitting the number. Is something going on with the RSS that we need to look at as a warning sign?

But as far as they discussion about good, I think it would be good to have in the document. Have it defined as something not as numbers or thresholds but, as you were saying, something like, "Hey, if you're receiving an SLA, then you may be subject to something that's good, and good may be subjective to individual RSOs, depending upon how much money they're financing or how much money they're taking out."

So, instead of setting up good threshold limits, it would probably good to just have – well, using the term "good" a lot. It may be best to have, if an RSO [inaudible] accept money or some type of financing from another party, then they may be subject to just whatever the terms of agreements for that SLA/SLE would be. That would be subjective to, again, each individual RSO as to what good is.

UNIDENTIFIED MALE: Paul, I'm sorry. I must have been missed that because I've been looking for something higher than this for years. So my apologies, and good on

you. I'm just wondering whether we don't want to very explicitly state at this point that this is not the requirement we're looking at down the road in which there may be an SLA for some higher level of thing. This is the first pass low bar.

I'm also wondering whether we don't want to call out specifically that we intend to have a follow-on or any additional or a different work group that puts dollar numbers and SLA figures together because, absent mentioning it, I fear people will say, "This is it. Just staple an a to this and that's the way it goes." And that's not the intention.

LARS-JOHAN LIMAN: I kind of agree with both, actually – Brad and Jeff here – about a second set of numbers if you go for an SLA or an SLE. So I'm fine with that. I can also accept a level that specifies good for a different purpose than the SPF or net performance enmeshment thing in the model that we've designed.

[PAUL]: Such as?

LARS-JOHAN LIMAN: Such as internal use for the root server operators noticing that the system as a whole is degrading. We collectively need to something about it. So I see a difference between passing the bar and maintaining a good system. If that higher number, the better number, is used for a different purpose than passing the bar, then I'm fine. But I still think that we should first get these pass-the-bar numbers out the door before

we start to address the [inaudible]. That's just a matter of working resources and stuff. So take it one step of a time. Get this document out first. But I can agree with what you've said and taking further steps beyond that. Thanks.

BRAD VERD: Again, I'll say a couple things again but I'll try to put a different context on it. I agree with getting a document out the door, but I am also really worried about what Jeff said, saying that this becomes what the SLA numbers are. I think that is going to be a big failure on our part if that happens.

I've had some private conversations here. I'm going to put them together and share the sentiment that I got from people, which is we're defining the minimum today, which to me is, if you engineer to the minimum, you're engineering for – I mean, do we ever engineer for the minimum? Meaning, for anything you do … I don't know. Maybe that's a broad statement. But I don't think you engineer to the minimum because we're constantly growing. You define the minimum, which I feel like we're doing this round. I feel like, if we needed to find a second set of numbers and maybe that's tied to SLAs and money, we need that second bar.

Then the other aspect I've heard is, "Well, everybody is going to negotiate their own contracts and their own SLAs and whatnot." I think that's fine, but I think those negotiations should be anything above that second bar. Does that make sense? I'm just using numbers here. Let's say the minimum is 10% and the SLA number is 50% and then somebody

negotiates, “Well, I’m going to do 90%.” So that allows for those different things that I’ve heard. If I got that wrong, Jeff, tell me. But those are some of the conversations I’ve had with people.

KEVIN WRIGHT: This morning, we talked about the scope of the document and, in particular, what’s not in the scope of the document. There’s an item that says “Making comparisons between RSOs is not in scope of the document.” It seems like this second threshold might be skirting on that idea.

[KEFF OSBORN]: I just wanted to answer Brad, where you said the 10% is the baseline and then the 50% is what we’re internally choosing to target ourselves to and then go into a 90% when you get paid. I would argue that possibly being held to the 50% commercially with teeth is enough to call that a contractual thing, rather than having an actual higher level.

BRAD VERD: I don’t disagree with you, but what I’m saying is that some people came to me and said, “Well, everybody contract is going to be different. I’m going to negotiate my own SLA.” So, if you as a root server operator wanted to offer an 100% SLA but it’s going to cost whoever is giving you money this much, that’s up to you. But you can’t say, “I’m at 10% and 50% is the bar and I’m going to negotiate to 40%.” The bar is 50%. You can’t negotiate lower if you’re getting money. See what I’m saying?

JEFF OSBORN: Sure.

BRAD VERD: That's all I'm saying. You create a new minimum if money is involved.

JEFF OSBORN: [inaudible]

DUANE WESSELS: Okay. So ration them. Ryan, you wanted to say something? And then Paul.

RYAN STEPHENSON: Yeah—

DUANE WESSELS: No. Go ahead, Suresh.

SURESH KRISHNASWAY: Just quick. So we're talking about the RSS thresholds now. It's not the RSO thresholds, where there might be SLAs or SLEs in place. So—

UNIDENTIFIED SPEAKERS: No. We've moved on. This is the second bullet which doesn't say RSO.

SURESH KRISHNASWAMY: Got it. I misread that. Thank you.

DUANE WESSELS: Ryan, was your hand up, or was Paul next?

RYAN STEPHENSON: Go, Paul, because [inaudible].

PAUL: From the point of view of the motivations of the community, or I guess ICANN as the representative of the community here, what they're really hoping is that there's a really good system and that, when there is a bottle neck, it's never us. So the pressures that push people toward the negotiating table around that often end up influencing operational procedures. If you are maybe getting paid, maybe not – but if you have some motive towards meeting some minimum service level, that's going to lead to a bunch of operational practices. For example, if you're being measured on your availability, you want to make sure you answer all the questions that get to you. You're probably going to end up with an internal rule, like, before you bring that name server down, make sure that you've withdrawn the route so that it doesn't attract packets that aren't going to be answered – questions. You might even have another one, which is that, if it is down, the first thing you should do is make sure the route is withdrawn. But again, that's up to you: how you fulfill the requirements that are being negotiated. But the person who's negotiating them with you wants to make sure that you have some

motive toward creating operation practices that will then create the measurable outward circumstances that they're demanding from you.

The culture of how we all got here was, just make it work as well as we can and hope that that's good enough. That's the corner we're turning. But I certainly think that, until there's money on the table, nobody is in a position to make any demands. I agree with those who have said we shouldn't merely document current practice and say, "Don't be worse than that," because it may be that some of us are not meeting a minimum standard that we would feel bad about if we knew what it was and if others here knew also. Unfortunately, we don't know what we don't know, so we end up talking in generalities, like we're going to answer all the questions that get to us, unless they're part of a DDoS, in which case we have an exemption. But in terms of how long it's going to be before we answer it or whether some other person's network or so-called OPM is going to mean that, even though we answered it, the answer didn't get through, or some OPM is going to mean that, even though a query was sent, we never got it? We cannot answer for those things, and yet those are the things that any component negotiator representing the community would ask us to stand in for and make premises about.

So, in general, I just want to say, when there is some consideration, then the community will be in a position to say, "I will pay you more if you do better," or, "I will pay you at all if you promise to do certain things." Right now, almost anything we could choose, however arbitrary it was, would be better than the nothing burger that we have coming into this meeting. So the perfect being the enemy of the good can't be allowed to continue. We've got to just come up with something.

Ultimately, we're going to come up with something and, no matter what it is, somebody somewhere is going to say it's the wrong thing. Let's absorb that now. Let's get ready for rejection, despite all the horrible things people are going to say after we take a stand, and then take some stand. Thank you.

DUANE WESSELS: Paul, we are coming up with something. The question is, do we come up with one something or two somethings?

PAUL: One.

DUANE WESSELS: One. Okay.

RYAN STEPHENSON: In trying to channel what Kevin Wright was saying – Kevin, please, if I jangle this up, please let me know – about comparisons between RSOs, I'm glad it states that in the document upfront and before anything else. I agree with Paul: let's come up with the one thing.

What I'm nervous about is, if we do come up with … And it sounds like we're in consensus of coming up with the minimum – that's great – and what the minimum should be. But, by having the good, I think it would, again, start allowing people that read this document, interpret this document, looking at the numbers, to start doing comparisons. Even

though this document says it's not intended to compare RSOs, people that take this document and absorb it are going to start trying to compare RSOs. That's where I think it becomes publicly dangerous to start comparing RSOs, saying one RSO is better than another.

So that's why I'm like – I hope I'm understanding that correct, Kevin. If I'm not, just let me know. So I think that's what he meant by, with a comparison of RSOs, having a sudden type of good for the RSS or even for an individual RSO publicly. That can then draw comparisons between two RSOs. I'm hoping I'm getting that out right.

BRAD VERD: I don't know how to say this any different. Anybody can stand up a monitoring system and monitor the roots today, and anybody can provide that data publicly, and anybody could then compare the RSOs. So think it is naïve to sit back here in this room and say, "We want to make sure that we create a document and a system that you can't compare root server operators with."

[MATT]: I'd just like to second what Brad said. We've had the RIPE monitors for years that, for lack of any other authoritative source, have become the de facto authoritative source. How many of us haven't at one time or another gone to the RIPE webpage and looked or orange and red? It's just a fact of life. The elephant in the room is that people compare operators. Even people in this room, I'm sure.

UNIDENTIFIED MALE: [inaudible]

DUANE WESSELS: Okay. Go ahead, but I guess, based on this discussion, I want to ask the question of do we need to keep that bullet in the document that says the purpose of this document is not to compare operators? Go ahead, Liman.

LARS-JOHAN LIMAN: A couple of comments. Of course, Brad and Matt are quite correct in that anyone can compare them today. The question is which data is used by the Performance and Metrics Subgroup to make decisions and recommendations. That can be a different thing from what the general public use and compare.

The second I wanted to make is, when it comes to money and levels, I feel very awkward when someone says, "You're not allowed to negotiate any level you'd like," with whatever organization you're [giving] money to, with one exception, and that is if there is a central function within the model that we're creating. If that will provide money, then you can put requirement on what's allowed to be negotiated. But you should put requirements not on the root server operator but on the financial model for what it is allowed to negotiate. Thank you.

KEN RENARD: In the RSSAC037 context, the SAPF is going to have to make a decision – is the RSS big enough or whatever – so they will need some metric of

good, whether we do it in this document or not. So, if we decide to do it in this document or now or offline or as just an appendix, it could be for entertainment purposes only. Just, while we're here, should we take a stab at it if we're putting numbers down on paper? It doesn't necessarily have to go in this document.

The other—

DUANE WESSELS: But if we have one set of thresholds, that still satisfies that need, right, for the RSS. The question is, one set of thresholds or two sets of thresholds?

You need to think about it?

KEN RENARD: Yeah. The other point here is that, if we publish these minimum-level metrics, I'm sure somebody in the ICANN community is going to scoff at that and say, "Well, that's way too low. That's not good enough. And that's not what they're for." I'm sure we're going to fight that battle.

BRAD VERD: The only piece I'll add, Liman, to the PMMAF – I'm going [inaudible] pull that one out – is that the intent, I think, when we wrote 037 was that all that was transparent. So all that data was going to be available to the different groups. So, again, I go back to the comparison. We've had this big discussion here about metrics versus thresholds versus what we're publishing and not publishing. That's fine now, but I expected, when

and if 037 goes in, that discussion will be hashed out because the intent when we wrote this document was full transparency.

PAUL: I think I was the one who put in the not-comparison stuff because there was a whole lot of contention on one of the early work party calls about "I don't want to be compared against somebody else, if I'm compared somebody else." Of course, as you all have said, you can be compared by people outside of this, but these numbers will have some weight, and it caused a bunch of people concern that they were going to be on the wrong end of comparisons. So we put that in. We made it so that the only things published were pass/fail.

I'm hearing a reversal now from some people. I think we should by careful to differentiate what is coming out in the set of numbers that are being collected here and published here versus what is going to happen in the real world.

One thing – again, I wasn't part of 037, so I don't know how the money changed from this to this to this and where we ended up – that I've heard from a number of you folks separately, as, again, I wasn't part of the 037 discussion, was that you're also getting money from other people who want to know how well you're doing. So you're coming up with your own thresholds, or they're coming up with thresholds for you. Nothing in this document should look like it is the definitive set of thresholds, either minimum or good. I think, with people saying they don't want to do good now, that's good for the PMMF or whatever comes out of 037/038.

I think it would be great if this document talked about that and said numbers for good will come later. And many root server operators have their own internal ones, maybe that even compare them against other root server operators. That's all fine and it's totally out of scope for these measurements. I think that that would be good to admit just so that, if a reader goes, "They didn't do this right," well, this document did it right. But that doesn't mean individual root server operators aren't doing exactly what you want.

BRAD VERD: Sorry to say [inaudible], but I'll just say it one more time. I'm fine with not publishing a second set of numbers. I will be fine with it on one condition, and that condition I that, in the document, we state somewhere that, should money be involved or SLAs – because what's going to happen – this is real simple. Play it out. 037 goes in PMMF or SAPF is formed. The question comes up: "Hey, whoever wants to sign a contract with one of the root server operators. What should the SLAs be? Who do we ask? Oh, let's come right back here and ask. What should the SLAs be?" That's what would happen, I believe. What is the starting points for the SLAs?

So this conversation that we keep pushing off and not wanting to have is going to come back here anyways. As long as we put in the document that we realize that that conversation is probably going to happen, then fine. What I don't want to happen is that what we define here is the minimum for SLAs going forward because I think that would be a failure.

LARS-JOHAN LIMAN: I'm comfortable with that, with one exception, and that is the financial central function. With that dialogue, you're spot on.

BRAD VERD: [inaudible]

LARS-JOHAN LIMAN: Yeah, but as long as you're talking about money as if that was a single thing. For me, the system as it is today and as I would like it to be continue to be is that it's driven by money from a plethora of sources with a plethora of agreements and contracts and whatnot that all make up the stability of the system. I see no reason for a negotiation between an individual root server operator and a new entity that wants to provide money to the root server system to force them to use a specific starting point for the negotiation. That's a non-starter for me.

If it is with the central function that we have defined, then it is fine. But I want to be able to talk to anyone about money for the root server system. I want to start with a blank slate.

[FRED BAKER]: I would agree with … It seems to me that any such negotiation is going to sound something like this. "Gee, I would like to give you money. If only you would give me something." The root server operator then says, "Well, for this amount, I can give you X, and for that amount, I can give you Y." They talk about it a while and they wind up with Z, whatever that might be. It seems like that's characteristic of business discussions. We need to allow the root server operators to be businesses, I think.

DUANE WESSELS: Ryan, go ahead.

RYAN STEPHENSON: This goes back to Paul Hoffman's question. Are we just going to be publishing pass/fail? Or we actually going to be publishing real numbers?

DUANE WESSELS: What the document currently proposed is, in all cases, to publish pass/fail. During the discussion today, Paul forth the idea that, for the RSS, maybe we publish numbers. I quite like that idea, but that's not reflected in the document yet.

BRAD VERD: Again, going back to SLAs, I disagree very strongly with Liman and Fred on this. This is not a normal business. This is a global service that so far and we've stated as a group that we are good stewards of the Internet and we will keep it running. I think it would be irresponsible for me to have a negotiation with a company and agree to take money and say that I will operate my root server instances that you pay for at 10% availability. I think that is irresponsible. In the scenario you just proposed, that's perfectly reasonable.

LARS-JOHAN LIMAN: I wouldn't say perfectly reasonable, but what if someone came to me and said, "I wouldn't give you money to operate root servers if they're all painted red"? That's a different [issue. I would say,] "Okay, fine." And that would add money to the system so I could rate the level of performance in general for the entire system of I-root instances. It's a totally different thing. We shouldn't limit the root server operators to what they put in the contract that eventually lets money flow into the system. I think that the limitation here should be the opposite. It should be put on a central function, the financial function, if there is such a function coming out of this long work with [inaudible]. That's a perfectly fine place to put such a limitation, and that would probably make an example that other bodies that want to contribute to the system through other channels will follow. If you have a central function that is clearly displayed and is being used and it function well, then people tend to copy that. So I think it will happen, but I don't want to mandate it.

PAUL: I'd like to bring up something different again. Not only was I not part of making 037 but I also wasn't part of making 042. Someone earlier said 042 was really about outside organizations not being able to tell root server operators. That's not what 042 says. So, if that's what the intention of what 042 was, you need to make an 042 Version 2 because the words at the beginning is, "RSO must remain independent from each other, as well as from any overarching organization." I read that as, "You all can't tell each other how to run your internal business." That's, like, right up there in bold. So, if that' not what you meant, then I think you're going to have a whole bunch of problems coming up, like the

difference between Brad and Liman and Fred and stuff like that. Again, I'm confused because I wasn't part of this, but I read that – by the way, I like the words here. I like the words here better than what I'm hearing around.

FRED BAKER: Let me put that in very close context. Jeff and I are from ISC. Brad at Verisign is not allowed to tell us what to do. That is what 042 says. To us, Brad is an outside entity. Now, if I'm looking at the entire RSS, ICANN is an outside entity and we've got the root ops and so on and so forth. But, from this particular RSO's perspective, Verisign is an outside entity. That's what I was getting at when I said 042 was saying that an outside entity can't tell u what to do.

PAUL: But I thought you were saying that RSSAC could, which I don't – RSSAC is made up of RSOs.

FRED BAKER: Well, I don't saying anything of the kind.

PAUL: Okay. Then—

FRED BAKER: Liman, you want to pitch in on this?

RYAN STEPHENSON: To add to what Fred is saying, between the individual RSOs, they can't tell each other what to do. Overarching organizations is basically the outside. That's the generic catch-all for the outside organizations. But RSSAC does is we just produce advisories. That's really what they are: just advisories. It's a gentlemen's agreement that we all abide by these advisories for our service expectations. That's it.

LARS-JOHAN LIMAN: I agree with you, and I would like to support that statement by saying that, with the work in RSSAC, Netnod in a discussion eventually agrees to something. Netnod is willing to agree to certain requirements on the service that we specify, that I take part in specifying in this group. I'm not willing to let RSSAC have a say about how we drive our business. That's a thing outside the scope of the group, from my perspective. That's why I have the standpoint I have.

With the technical part, Netnod through me can agree to a certain technical level of performance. I'm willing to discuss having a high level that we all strive to meet, but I see a difference between a requirement and something we aspire to reach. That's the discussion I would like to have and I see us having. So, so far, so good. Thanks.

DUANE WESSELS: We've got about 15 minutes before it's time to break. I'm going to attempt to summarize what I think we've heard and, I guess, a way

forward for this particular question, which is, should there be one or two sets of thresholds?

I heard almost no one arguing for having the work party recommend two sets of thresholds. I think everyone agrees that, at this time, that's either too much to take on or it's out of scope. So we won't do that. But the document does need to include some new text about the idea that there may be other requirements placed on root server operators from SLAs that would [accept] different or higher thresholds and that those would be done separately and that money may be involved and that sort of thing.

Is everyone on board with that way forward? Did I capture everything, Russ, or did I miss anything there?

RUSS MUNDY: I think you got it. We have a section in Section 2 that would be, I think, a good place to deal with that because it is already talking about the PMMF in 037 and so forth. So I think you've got it summarized well.

DUANE WESSELS: I think Russ has been making comments in the document as we go along so we don't forget about this important thing. But anybody else, please always go to the Google Doc and add your own comments or contribute text. We'd welcome anyone doing some writing in the document.

I didn't mean to wrap-up so quickly. I guess, if there's no discussion of that – oh. Go ahead, Jeff.

JEFF OSBORN: Well, as I was debating, if we had run out of time, whether to bring it up or not. I hope we're not beating a dead horse. This implies that this is a certain level of performance we're talking about and there's another one further down the road. It was always my belief -- and I think from Fred and Brad I was getting that idea – that we were going to have a different but similar group that's starting with some of this information that was going to put together something like a financial model. If it was as simple as Liman seems to think, where you sit and have a negotiation with somebody, I wouldn't worry so much. But every step of that, just the RSSAC and the root server system, is not a normal group. The ICANN community is not a normal group.

When asked if I could sit down with a decision-maker and do a negotiation, which I asked for thousands of time in my life, I got an answer back I never expected or heard of and I still don't understand, which is, "That's simple. Just create a policy and the community will coalesce around it." I was like, "I'm looking for the guy with the checkbook," and what I was told was, "Create a policy and the community will coalesce around it." I don't know what that means, but it seems like we need to –

UNIDENTIFIED MALE: That was part of the intention.

LARS-JOHAN LIMAN: [inaudible]

JEFF OSBORN: Anyway, at some point, we need an organized group that goes off and does that. I think it has been premature up until now. Maybe it's still premature, but it'd be nice to make the decision to either start it or not start it yet, rather than have it accidentally be, "Oh, no. We have a week to do a month's work," or worse. Thanks.

DUANE WESSELS: Jeff, what I'm hearing is that, for you it may be helpful if this work party made a recommendation that there be a follow-in work party to explore financial models. Is that …

[JEFF OSBORN]: Brad said it better than me. He did this.

DUANE WESSELS: That's why I am …

[JEFF OSBORN]: [inaudible] I concur.

DUANE WESSELS: In all seriousness, is that something everyone agrees on? Would everyone like to see a recommendation like that out of this work party/

BRAD VERD: If I may, it seems like we spent a lot of time talking about it. So it seems that maybe the recommendation should be that, given that this wasn't in the statement of work, as far as, "Let's create this," it consumed a lot of our discussion. So it seems like this work should be done. That's a way to think about it.

FRED BAKER: Jeff, I have a suggestion for your proposed policy after you write it. Find some guy with a checkbook.

JEFF OSBORN: That's what I've been doing all my life.

DUANE WESSELS: So I think we've accomplished, as far as this work party is concerned, what we wanted to accomplish in this section. So thank you very much. I know it's coming up on break time, but I'll ask if there's any last-minute discussions before we break a little bit early.

UNIDENTIFIED SPEAKERS: [inaudible]

DUANE WESSELS: They may not quite be ready for us, but I think they're close.

BRAD VERD: [inaudible]

DUANE WESSELS: Yeah. Well, let me use this time then to say, as you can see from the agenda, we have lots of other sessions where we'll be talking about metrics. I've already heard from a couple people about specific things they would like to talk about while we're here. So, if you have things that you would like to talk about that you think might not already be on the agenda, please find Russ or myself. We will spend some of our break time doing a little agenda bashing and making sure that we get to talk about the things that are important. Or you can throw them out now, since I've got …

FRED BAKER Well, I do think that Paul mentioned a couple things that I'd like to discuss.

DUANE WESSELS: Paul and I have talked already.

[DANIEL MIGAULT]: From the discussion we've had now, do you expect some inputs in the document, or do you expect that we provide text or comments?

DUANE WESSELS: From the discussion that we've had so far today, I think we have a pretty good sense of what we need to do. We always welcome text and comments in the document. So I would say to definitely take a pass

through it and make sure that your concerns are represented. If they're not, then add them in. If you feel like adding words for specific thing, that would be great.

[DANIEL MIGAULT]: Okay.

[PAUL]: Even though you're the primary document author and I seem to be throwing in a lot, this is a Google Document which has a funny feature, which is that you can add text or you can add a comment. As this has been going, then you and Steve try to incorporate such like that.

Adding comments has a couple of really bad features, as it turns out. One is that you can add a comment and I can click Resolve and your comment goes away. I did that by accident, told the person, didn't remember what the comment was, but they were able to bring it back. I have been resolving my own comments today, like the one where I said we're not going to do thresholds for the RSS. Now that we are, I resolved my own. But I propose that people actually add text.

Let's say, Daniel, that you add some text and I don't like it. I can simply add another paragraph underneath, saying, "My alternative is this," and then people can pick and choose. But doing everything in comments and the threaded-ness of the comments is good for some things but not for this, as we've discovered because, quite frankly, quite few of you have added text. By my own count, at least five places where we want to clarify or add text so far … So please add text, even if you believe that

someone else is going to want to rewrite it. Add it as text and let them go from there.

DUANE WESSELS: Yeah, I agree. Keep the comments. For longer things, definitely add the text in the main document.

LARS-JOHAN LIMAN: This is Liman actually asking, doesn't the comment reappear if you roll back the history of the document?

PAUL: Yes, but the way that they do rollbacks is incredibly painful. This is not—

UNIDENTIFIED MALE: You have to roll back the whole document.

PAUL: Yeah. This is not SVN or even [get].

LARS-JOHAN LIMAN: [He sent to the person who uses [inaudible].]

[RUSS MUNDY]: So if I could ask folks to take a look at the agenda for the rest of the meeting that's focused on the metrics discussions, take a look at the

document and see if the things that you would like to discuss are already covered on the agenda. Make little notes to yourself to bring them up there. If they aren't, that's where I really need to hear from individuals about things that we need to discuss on the document. So this might be a good time to try to do that while they're finishing setting up the food. Thanks.

DUANE WESSELS: All right. I think we're done with the session. Thank you, everybody. The lunch should definitely be ready in a few minutes.

BRAD VERD: Lunch is ready. I guess we'll see you back here at 1:30.

[END OF TRANSCRIPTION]