

**NWS-US DEPT OF COMMERCE**

**Moderator: Michael Monroe  
September 15, 2016  
8:12 am CT**

Tommy Wright: And welcome to the Fall 2016 Meeting of the Census, excuse me, of the Bureau of the Census Scientific Advisory Committee. I feel kind of happy today.

There's all kinds of good news coming out of the Census Bureau for the nation. I'm sure you've heard about some of this. I see some people nodding their head at the Census Bureau.

I think there was some historic event in terms of a point estimate of how much the median household income went up. John tells me to not – don't get into the details but in terms of the point...

John Thompson: No, no, no. Look Tom...

Tommy Wright: Yes.

John Thompson: I think it's really important to note that the Census Bureau takes no position on the relation of what happened with the income release. Other people have been interpreting that and making statements.

Tommy Wright: Yes.

John Thompson: The Census Bureau strictly produces high quality scientific data and that's our position on the data. We have no comment on its increase or decrease or anything else.

Tommy Wright: Okay. All right. I'll second that. I also heard that a weatherman used a term this morning. There's magnificent weather for the next two days so I think that was a special order for the Scientific Advisory Committee meeting so we can keep that in mind.

We're really as I said glad to have you join us as members of the Committee and look forward to a productive discussion and sessions for these next two days.

My name is Tommy Wright. I'm the Designated Federal Official of the Census Scientific Advisory Committee and as such I am actually required to preside over the Advisory Committee meetings as specified by the Federal Advisory Committee Act.

Before we begin please note the sheet at your seat outlining the emergency exit and safety procedures. The proceedings are being recorded and transmitted live on Webcast by way of the Census USTREAM channel.

Please be advised that any side conversations will be heard. Any side conversations will be heard. Every time you're ready to speak please turn on the microphone and clearly state your name for the records.

All meeting materials have been posted on the Census Advisory Committee Web site for the public viewing online. Now I'd like to introduce people sitting at the head table.

To my left is the Committee Chair, Barbara Anderson. Next to Barbara is Tim Olson who's the Associate Director for Field Operations. To Tim's left is Bill Bostic, Associate Director for Economic Programs.

Next to Bill is Shirin Ahmed, Assistant Director for the 2020 Census. Next to Shirin is John Abowd, the Associate Director for Research and Methodology.

No one is next to John but someone will be sitting next to John. Joanne Crane, Associate Director for Administrations. To my right is John Thompson, the Director of the Census Bureau.

Next to John will be sitting a little – joining us a little bit later Nancy Potok who is the U.S. Census Bureau Deputy Director and Chief Operating Officer. Next to that place is Harry Lee who is the Assistant Director for Information Technology.

And next to Harry is Enrique Lamas, Associate Director for Demographic Programs and around the corner next to Enrique is Jeannie Shiffer, Associate Director for Communications.

And joining us a little bit later I hope is Ted Johnson, Associate Director for Performance Improvement. Barbara and I will share in facilitating your deliberations today and tomorrow.

Between the two of us we will do our best to keep the discussion moving on time, ensuring that we hear from everyone who has a comment. I'm scanning the audience and this is time to acknowledge people who may be visiting.

We want to welcome all staff who are participating by way of Webcast and the public participating in person. I don't myself see anyone from the Department of Commerce office or Congressional staff members, not that I know of them all but some faces are familiar.

I don't necessarily see regional staff. I should note that for the Committee Roberto Rigobon who's – will be a discussant for one of the sessions. Peter Glynn and Juan Pablo Hourcade – all Committee members will be joining us by way of conference line for parts of the proceedings so they're not here physically but will take place – take part in the meetings.

I understand there is an iPads demo now. (Kyle) – I saw (Kyle) a minute ago. Where is he? Okay (Kyle) will give us demonstration.

(Kyle Wilcoxon): Hi. Good morning everyone. My name is (Kyle Wilcoxon) and I'll be giving you a brief demonstration on how to use the iPads for today and tomorrow's conference.

If you need any assistance throughout the demonstration we do have analysts that will be around to help you. All right. Your iPad should currently be at the Home screen.

If they're not at the Home screen please press the Home button, which is located on the right side of the iPad in the middle. You will see three applications on the bottom of the screen.

The first application is MaaS360. If you open that up this applicant – this application contains all of the presentations for today and tomorrow. So if you select the CSAC folder the list of presentations are located on the left hand side of the screen, and once you select a presentation it'll open it up on the right side.

If you want a wider view of the presentation there is a X on the top middle of the screen, so if you select that it'll give you a wider view and if you want to go back to the original view you will select Docs, which is located on the top left of the screen.

All right, if we go back to the Home screen the next application is Notes, and this application will allow you to create and email your recommendations to the CSAC Chair.

So to start in Notes you will see an icon on the top right side of the screen that looks like a square with a pencil, so you select that and then you can type in your recommendations. So once you have completed writing your recommendations you can then...

Barbara Anderson: I'm sorry. When we went back to the thing there's something that says messages but nothing - when we went back to the screen there's something that says messages but nothing that says Notes.

(Kyle Wilcoxon): It's the Notes application.

Barbara Anderson: Sorry.

(Kyle Wilcoxon): Notes here.

Barbara Anderson: Okay just...

(Kyle Wilcoxon): Okay.

Barbara Anderson: Thank you.

(Kyle Wilcoxon): Okay so once you – once you've completed making your recommendations or typing them up you can email them directly to the Chair from this app. So to do that you will see an icon in the top – on the right hand side of the iPad.

It will look like a square with an arrow pointing up. All right, so when you select that you will see an option that says Mail, and if you select the Mail app it'll bring up the ability to email directly to the Chair.

So the email address for the Chair is [csacchair@gmail.com](mailto:csacchair@gmail.com) and in the subject line if you could just put your names so that the Chairperson will know who is sending the recommendations.

If we go back to the Home screen the last application is Safari and this application will allow you to surf the Web for any information you may need for today and tomorrow.

All right, you have two handouts in front of you detailing the functions of the applications and instructions on how to email from the Notes application. And if - anytime you need assistance throughout today and tomorrow, myself and (Raymond Lee) will be here to help you so thank you and have a great conference.

Tommy Wright: Thank you. Thank you very much (Kyle). Now look at the – today's agenda. Our meeting agenda reflects a broad range of topics and as always it was

developed in response to our need to share and introduce critical research and program developments requiring your attention.

In addition the agenda has topics that you have recommended on critical program areas. Sessions have allotted time for discussants, presentations and Committee member discussions as well as an opportunity for you to jot down any notes as just mentioned that you might have following the presentations.

All presentations, papers, supporting materials and notes section are loaded to your iPad as you also just seen. Before moving to the agenda we want to thank everyone.

Sometimes at the end of the meetings we forget to thank everyone but let's – we will do that again but initially I wanted to thank everyone: the presenters, the CSAC Chair, discussants, working group presenters, the Advisory Committee Coordinator and working group subject matter experts.

A lot goes on to make a meeting like this happen and thanks very much. First on today's agenda will be our Committee Chair, Barbara Anderson, who will bring remarks and introductions of the CSAC members.

Following her John Thompson and Nancy Potok will provide executive remarks on important Census Bureau programs/activities. Bill Bostic and Nick Orsini will present the economic programs updates with Krishna Rao as a discussant.

We'll then at about 11 o'clock have a short break in which the official photographer will come and take a group photo. After the break Carma Hogue will present the big data initiatives with Roberto Rigobon as a discussant.

During the working lunch Shirin Ahmed, Deborah Stempowski, Maryann Chapin, Atri Kalluri and Patty McGuire will present the 2020 Census program updates with Sunshine Hillygus as a discussant followed by Committee discussion.

We'll take a break at 2 o'clock. Following that break Barbara Anderson will lead CSAC in a brainstorming session on potential census data products. John Abowd will talk about following that disclosure avoidance.

Then we will move to the CSAC Committee discussion and formulation of recommendations. To me – today's meeting will end promptly at 5 o'clock. Just to remind you there is – but there is coming up a virtual meeting to discuss the national content test results and this meeting is on October the 6th, 2016.

For the 2017 regular meetings the first one is March 30 and 31 and the fall one is September 14 and 15. These are dates for the in-person meetings. I would like to remind each of you to always state your name and speak directly into the microphone each time you speak.

As a reminder to the audience during any of the question and answer sessions occurring later today only Committee members are permitted to ask questions and/or make comments to Census Bureau panelists.

The public will have an opportunity to speak tomorrow, Friday, at 11:00 am. If anyone wants to make a comment at that particular time please leave your name at the registration desk out front.

A few housekeeping things. Committee members must stop by the registration desk today to retrieve your packet for travel reimbursement. Please remember that the refreshments are for Committee members only.

The bus will leave for the hotel this evening at 5:15. If you need to take a taxi you can stop by the desk. It's really nicely set up there. And the restrooms are behind that particular wall. Now I'll turn it over to Barbara. Please welcome Barbara.

Barbara Anderson: Hi gang. I'm really happy to see all the Committee members and all the Census Bureau staff. I think that we are – am always very happy to hear about the progress toward the 2020 Census, and I'm really happy that we're going to get a lot more information about that in these two days.

Also as you know at the last meeting we made an extensive set of comments and recommendations to the Census Bureau, and I am not surprised but I am delighted that we received extremely detailed and thoughtful responses from the Census Bureau, which I'm sure that we all really appreciate and in that also we had a lot of things to say about ACS.

We had very good responses including a response from the Census Bureau that they would do some research to get at a broader, deeper understanding of the nature of respondent burden.

As all of you know and as I emailed you the rules for developing our recommendations have changed. As I understand it it wasn't that we'd done anything wrong but it was related to advice, instruction, whatever you call it from the legal people about how this needs to proceed.

And as you know it is strongly recommended that we be able to present our final recommendations in a session at the end of tomorrow, although I understand it is technically allowed for us to have a further public conference call.

That is to finalize things. That is not desirable on anyone's part. The – I think we will try very hard to continue to give reflective and constructive comments to the Census Bureau.

I think this will be a challenge in this new arrangement but what it means is the very helpful things said from the IT people are especially important. So this means gang it's really important that you all when you've finished your – that you all actually write down notes about what you think about the various presentations, but that also by the end of today that you actually email me as it was explained to you those notes.

And so then we are having a – the end of today but tonight I will try to put this together in something as coherent as possible given we're – added that time so that we'll be in a pretty good position to try to do everything we're supposed to tomorrow.

So we have to – and I also would like to thank the Census Bureau staff, (Sara) and the IT people for doing a lot of work in order to make this conceivably actually possible.

I know this was not easy. We said, “All right, we're going to need...” I said, “And we're going to need this to make this work,” and I think there's – it's actually likely it will work because they've invested enormous amounts of energy into - I know this is tough - into making this happen.

I also will and I think (Sara) will say more about this at some point and (Sara) is always – has been absolutely fabulous. We all know that. That – there is an optional dinner tonight and as I understand Director Thompson will join us.

I regret that Associate Director Potok cannot because of a professional prior obligation and – but it's been great when they both came. This is going to be at the restaurant at the hotel, and I will say I had dinner there last night with a former student and the food was really good.

I had salmon. He had haddock. We both just loved it so I think that we'll enjoy it and we'll also enjoy the food. So I look forward to a very – these meetings are always congenial and productive but another congenial and productive meeting where we continue to give constructive and critical advice to the Census Bureau.

And I'm sure they will continue to act the way they do and to take it in the constructive spirit in which it's offered and thanks a lot folks.

Tommy Wright: Thank you very much Barbara and next we'll hear from John Thompson and followed by Nancy Potok. John?

John Thompson: Thank you. Good morning. Let me talk a little bit about some of the developments – I'm sorry. Let me talk a little bit about some of the developments here at the Census Bureau.

So first we do have a CIO, Kevin Smith. Kevin would be here today but Kevin is a brand new father so he's got a lot of things to deal with in his life right now, and we're all happy that mother/father/baby are all doing really well so that's good.

I had a hearing in June. I testified, our CIO testified and Harry Lee testified and so did GAO. It was a fairly extensive hearing. It went a little over three hours so that was enough said there.

I do have another hearing on the books right now scheduled for June – I'm sorry, September 29 assuming that the Congress is still in session. The topics are going to be a readiness for the census for 2020 and security, our testing this fall and particularly our address canvassing, which is in Saint Louis and Buncombe County.

I hope I got that pretty close to right. Jeannie's nodding. That's good. So I am going to testify, Kevin Smith, our new CEO, is going to testify and two areas of the GAO are going to testify so stay tuned.

And I'm – it – it'll be on C-SPAN so it'll be – or it'll be live stream, one or the other so if you're interested in looking at it feel free. A couple of other things that are – that have been going on.

We have completed our first round of consultations with the American Indian tribes and we are starting on a second round. We're going to be consulting with the Navajo Nation next Monday and we have just last – about two weeks ago we went to North Carolina and had consultations with a number of state-recognized tribes.

Those consultations are going very well. I think the big thing we're learning this time around is – well a couple of things. We're learning that the technology – actually there are – there is technology on some American Indian reservations.

There is the ability to use cell phones and Facebook and various communications. On others there's not. The big issue that has arisen deals with tribal enrollment and that is kind of – it's rather different this decade.

So for 2010 we held consultations and we got a strong message that we should collect tribal enrollment data and that's different from – we collect tribal affiliation right now as part of our race question but – and enrollment has very specific criterias.

It varies from tribe to tribe and it can be blood quantum. It cannot be blood quantum. It can be a variety of things but there are specific requirements for enrollment.

So, you know, learning from the past we started too late last decade with the consultations to do anything about it and so we just said, "Well sorry it's too late."

So this year we started consultations two years early leaving enough time to do this, and even thinking that we were going to get the same recommendation we started - and we had gotten one request from a federal agency for this data as well.

We started a process to do cognitive testing to prepare for a test of collecting tribal enrollment, which is very challenging in a self-response census. We're not actually sure that in a self-response setting we can collect tribal enrollment accurately but what - we're planning to find out this fall.

However the world has changed with respect to the American Indian community's view on tribal enrollment. So there's very strong comment from the Native Alaskans that we will not be able to collect this accurately in

Alaska because of the complexities of the - Alaska's varieties of situations and that as a result Alaska would suffer if this was used to allocate funds, so they're very much against us collecting it.

The other thing that came up in the Southwest and actually in a number of other settings was the feeling that the government - the United States government has no business collecting tribal enrollment.

That's a function for the tribes to do and we've been hearing that very, very strongly. And in fact the National Conference - Congress of American Indians passed a resolution saying we should not collect tribal enrollment.

So where we are on that right now is we are going to do a test to look at - to - because there's no information. There's no scientific data right now in the United States on the ability to collect tribal enrollment in the self-response setting.

There's just none that we've been able to find so we want to produce some scientific data on that. However we're - before we would move forward to collect tribal enrollment as part of the 2020 Census would involve some very serious consultations with the tribes to do that because we don't want to lose the - we need their trust.

We need their cooperation to conduct a census on the reservations and get them to participate - the - with the off reservation. So it doesn't mean that we're - just because we're testing it doesn't mean that we're going to put it in the 2020 Census, but we do feel it's important to gather some scientific data on this.

We put out our residence criteria for comment – proposed residence criteria for comment. It went out in late June and the comment period ended I believe September 1 Jeannie. Yes September 1.

I think we set an all-time Census Bureau record at least in my history with the Census Bureau in terms of the number of comments received. We got over 77,000 comments on the residence criteria.

All but probably about 100 to 150 dealt with where we would count incarcerated people, and most of those have various themes about counting incarcerated people at their pre-arrest location so that's – those are the comments that we see.

We're in the process right now of going through the comments to categorize them and prepare responses, but it's going to take us a little while to do that because we didn't anticipate we would get quite that many comments.

And as we mentioned we are in the process now of finishing up our national content test analysis, and we're in a position where we can brief our advisory committees on this.

And we're setting up – I don't know if they're set up yet but we're in the process of setting up two virtual meetings where we can brief the National Advisory Committee and brief the Census Scientific Advisory Committee on our findings from that test. So yes with that I think that I will turn it over to Nancy Potok to continue.

Nancy Potok: Oh they're not. Good morning everyone. Welcome. Thank you for coming. As always we really appreciate you being here, and as Barbara mentioned earlier we do take your recommendations and advice very seriously and find

them to be extremely helpful so we really appreciate you giving your time to move this forward.

I think just about everybody knows that my part of the opening remarks always revolves around the budget so I will get to that. I do want to mention something really exciting that's going on that we're able to at least assist a little bit and that's the Evidence-Based Policymaking Commission.

I hope all of you are aware and paying attention to this Commission, which was created under law. They had 18 months to come up with recommendations on how to advance program evaluation and evidence-based policymaking.

I think, you know, looking at the – at federal programs but also state and local policies as well it's a very illustrious Commission. They've kicked off. They've had some meetings.

The reason why we're very interested in it - not only because we're very interested in the outcome and especially some of the things that they're looking at in regards to reducing barriers to record sharing among federal agencies and how to do more to really advance research, but also because in the legislation the Census Bureau was named as the entity that would provide administrative support and other assistance as needed and actually pay for the Commission through its appropriation.

So we're really excited about that and I bring it to your attention because they do have a very short lifespan to get this report out. They're working at a very fast pace and if you – they're planning to get a lot of input from the research community and from people who are doing program evaluation.

And so I just bring it to your attention so that you follow along with what the Commission is doing, because I think if you're interested in what the Census Bureau is doing you'll definitely be interested in what the Commission is doing and the kind of recommendations that they will come out with on their report.

So I just wanted to bring that to your attention and there are other agencies that are helping, you know, get information to the Commission and things like that.

But if there are opportunities and you're interested – anyway so they have a Web site and they had their first meeting in June. So on to the budget. We're at a fairly critical juncture I think in terms of our budget.

Fiscal Year '16 is coming to an end rapidly - September 30. We're closely watching Congress as we always do at this time of year. Our assumption based on all of the public information that we have which is really all we have is that there will be a continuing resolution this year as we've had in the past.

The information that's kind of out there – I can't really vouch for what ultimately will happen because that's completely up to the Congress in terms of how they want to proceed.

But our expectation is that there will probably be a continuing resolution at least through the early part of December and then, you know, Congress is expected to come back for a lame duck session.

And so at that point they will either, you know, put in - a budget in place for the entire rest of the year or do probably another quarterly continuing resolution and let the new session of Congress pick up the budget.

For us what's really critical is that particularly for the 2020 Census we have big cyclical increases in 2017, because in the beginning of 2017 we need to really be focused on all of the work that we have to do to get ready for the 2018 end-to-end test that we're doing.

Our eye is really focused on that ball for the 2018 end-to-end test. That is probably our biggest milestone, biggest, I mean, it's a series of operations and events but that is what we're really focused on.

And so that causes us to really have to step up our funding, and particularly in the second quarter of the year it's very critical that we have more money to spend.

So what happens with the continuing resolution of course is of great interest to us because there's sort of three funding levels out there. There's the House level, there's the Senate level and there's the continuing resolution level.

And the continuing resolution level because neither the House or the Senate has actually passed the bill would be our limit so we'd be spending at the 2016 level.

That will get us through the first quarter of the year but if we hit the second quarter of the year and we're still at that flat level it is – we are just going to have to make some very undesirable choices in terms of how we – how we're able to pay for the 2018 test, which is something we just have to do.

So a lot of other things that are very important would go by the wayside at that point if we don't have more money by the second quarter. There's a couple of

ways you can get more money by the second quarter if Congress hasn't passed either the House or the Senate bill.

I think some of you have heard the term anomaly which is, you know, you get permission to do something other than what applies to the whole government when there's a continuing resolution.

Sometimes it means you get extra money so you have a little line in there that says, "Okay everybody has 2016 but this particular agency gets a bump up." That's always great but very difficult to pull that one off and it hardly ever happens.

The other one is that you don't get more money but you're allowed to spend faster in anticipation of getting more money by the end of the year so you don't end up in the red at the end of the year and that's something that I think we'll really, really be pushing to get.

If we can't get more money we need to spend it faster than the rate would normally allow us to spend it so we can be ready for that test as we get towards the end of the year when it launches.

We have again, you know, these three numbers out there - the 2016 level, which will not get us to where we need to be. We have the House funding level, which is quite a bit below what we asked for.

It's about \$121 million short for the 2020 Census and then we have the Senate level, which is also not everything that we asked for. It's about \$73 million below but I think we're - we feel like if we can't even at least get the Senate level we're in dire straits.

There's a lot of sacrifices we're going to have to make even at the Senate level, but if we can't even get to there – and, you know, what's challenging is that a lot of times when you have a difference between the House funding and the Senate funding there's a tendency to kind of split the difference.

And that kind of – what we see as a common way to reconcile between the House and the Senate level is just simply not going to work for us so we're following this very closely.

We're getting, you know, really close to kind of some critical dates for us in terms of making decisions on what we have to do, and how we weigh kind of what's the likelihood of getting different funding levels versus when we have to make critical decisions about operations that are rolling out during '17 and how we get to where we need to be on the 2018 test.

So I think, you know, the next couple of weeks will tell us a lot in terms of those decisions, and of course as soon as we know more and we're able to make those decisions we'll be letting everybody know where we are.

So a lot of uncertainty and risk management going on right now and that's about it.

Daniel Atkins: Dan Atkins. Could you say again who appointed that Commission on Evidence-Based Policymaking?

Nancy Potok: Yes. The Commission was established in law so it was – it's also for short called the Ryan-Murray Act because it was Paul Ryan and Patty Murray who really co-authored the legislation establishing the Commission.

The commissioners are appointed so they're appointed by the Congress and by the White House and it's split so that there's – on - the Senate appoints some and the House appoints some and the majority and the minority - so Republicans and Democrats in both the House and the Senate got an allocation of like three or four I think each to appoint, and then the White House got three people to appoint and that was in the law how that was going to happen.

Yes and the White House got to name – there's two co-chairs and the White House got to name one of them.

Tommy Wright: Nancy, Tommy Wright. Just a point of information. There is a short briefing on the agenda tomorrow morning just to remind you on this Commission. It will be given actually by the Executive Director of that Commission.

Irma Elo: This is Irma Elo. Just one thing about the budget. Do you know - when can you be specific about what those cuts would be if you don't get the money, because I know this always comes up when people go and try to help the Bureau on the Hill?

They need to know what those specific cuts would be so when would you anticipate that you would know those specifics?

Nancy Potok: I think we'll definitely know, you know, in the next couple of weeks. We pretty much have to because even if there's a CR, you know, a continuing resolution I think we need to – we'll need to be very clear on what our appropriators in Congress know - what kinds of decisions we have to make even under a continuing resolution.

So, you know, we don't – I think we're looking at what we think we need to do right now. We don't live in a vacuum. We live in an administration so we have to, you know, work things out with the Department of Commerce, with OMB and all that to make sure that everyone's in agreement with our decisions.

We're working through that process now so I think soon. It's going – it's – and obviously it's critical that we let people know what our plans are very soon.

John just reminded me there's – there is – he was talking a little bit about some of our internal staffing but there is something else that I think you all would be very interested in if you don't already know about it.

So it's, you know, it's sad news and it's good news. The sad news is that this January we will be losing Bill Bostic to his – probably to his happiness because he's retiring so it's our loss.

Maybe his gain is he gets to spend – devote a lot more time to his private endeavors so we're very sorry to see him go obviously. He's just been such a major figure and contributor to the Census Bureau over the course of his career, which I think is what, at least 40 years?

John Thompson: I think it's 41.

Nancy Potok: Forty-one years which is just phenomenal and, you know, his – Bill has just contributed so much. The good news is that Ron Jarmin who is currently the Assistant Director for Research and Methodology has agreed to move over to Bill's position and be the Associate Director for the Economic Programs so we're very excited about that.

And of course I know that you all know Ron pretty well and are excited with us so that will take place October 1. And another thing that we're doing here is as we're ramping up for the 2020 Census a lot of the – what we call the mission enabling functions that will let the census occur such as knowing that we have to hire a lot of people, get a lot of space, do a lot of those kinds of functions as well as continue to do some very critically important things that we would be doing anyways like our succession planning, really looking at our strategic workforce planning and what kind of skills we need here going into the future.

Some very critical work going on in that area, and Bill has very kindly agreed to spend his last couple of months heading up the new function that we're establishing here to oversee all of those activities while we're looking for a permanent head to be our – what we call the Chief Administrative Officer who will be overseeing a lot of those key strategic initiatives that are mission enabling.

So come October 1 Bill will be moving over to act in that new position that we're setting up until he leaves us at the beginning of January. So he'll still be here to, you know, have a guiding hand on Ron's shoulder until he rides off into the sunset.

Tommy Wright: Tommy Wright. We still have time for questions and comments. I think I saw Willie's hand.

Guillermina Jasso: Yes Willie Jasso. Very quickly is it possible for us to have copies of that report, the Income, Poverty and Health Insurance Report?

Nancy Potok: Yes absolutely. We can get those to you...

Guillermina Jasso: Thank you so much.

Nancy Potok: ...in a few minutes probably.

Tommy Wright: John? That is – my name is Tommy Wright. I'm calling on John. Here we go.

John Thompson: Yes. So the Evidence-Based Policymaking Commission has a public call for comments that went active yesterday and the comments are due November 14, and we'll make sure you all have the URL for that before the end of the day. Thanks.

Tommy Wright: Tommy Wright. Noel?

Noel Cressie: Noel Cressie. I'd just like to congratulate both Bill and Ron. It's a pleasure working with Bill over the years and certainly in the Big Data Working Group, and I don't think we'll lose Ron to the Big Data Working Group. I'm just very happy for both of you. Congratulations both. Thank you.

Tommy Wright: Tommy Wright. We actually – are there any other questions or comments? We're actually a little bit ahead of schedule but I'm going to turn it over to Barbara who has one task.

Barbara Anderson: Folks I never claimed to be perfect and this demonstrates it. I'm – well this part is perfect. I also wanted to, you know, thank Noel but on behalf of the entire CSAC to thank Bill Bostic, which I told him in private earlier for all the fabulous things he's done on Economic Programs and to wish him all the best in all his future endeavors.

I'm sure we'll hear a lot from him in the future even though he won't officially be with the Census Bureau so thanks so much Bill. But the mistake I made was I forgot to have all you other CSAC members introduce yourself.

So I am so sorry but I don't think you lost your identities in the interim. So if we could start with Irma and say your name and your institution, then I think we'll – I will have corrected all my mistakes today at least.

Irma Elo: Okay. I'm Irma Elo. I'm from the University of Pennsylvania.

Douglas Massey: Doug Massey from Princeton University.

Jack Levis: Jack Levis from UPS.

Kenneth Simonson: Ken Simonson, Associated General Contractors of America.

Krishna Rao: Krishna Rao from Zillow.

Robert Hummer: Bob Hummer, University of North Carolina Chapel Hill.

Allison Plyer: Allison Plyer from The Data Center in New Orleans.

Daniel Atkins: Dan Atkins, the University of Michigan.

Noel Cressie: Noel Cressie, University of Wollongong and they've asked us to say University of Wollongong comma Australia. Thank you.

Sunshine Hillygus: Sunshine Hillygus, Duke University.

Jeff Lower: Jeff Lower from IIC Technologies.

Andrew Samwick: Andrew Samwick, Dartmouth College.

Guillermina Jasso: Willie Jasso, New York University.

Barbara Anderson: And I realize the other mistake I made. I forgot to say I'm also from the University of Michigan.

Tommy Wright: Tommy Wright. Thank you very much everyone. We're a little bit ahead of schedule so let's just continue on with Bill Bostic and Nick Orsini who will give Economic Program updates: the 2017 Economic Census, Census of Governments and Improving Economic Statistics. Yes.

William Bostic: Good morning. Certainly thank you for the kind remarks. Before we get started want to play a video. I actually stayed two years longer so we did a major reorganization in September of 2014.

I wanted to ensure that we were conducting business differently but it really helps with a nice segue. And when we talk about the Economic Census, the Census of Governments and some of the efforts we're taking to modernize economic statistics this is kind of the basis of where we've been, where we are and kind of where we're going so a video for your delight.

((VIDEO))

William Bostic: We've been doing a lot of transformation and so we want to take a look at where we've been, where we are and where we're going. Econ transformation is really talking about the directorate being more agile, be more efficient because we are profiling and evolving economies.

We have been in the same – for 20 years and in that process we evaluated so we could make a data driven decision. There were a lot of redundancies. We lacked ownership for some of our major programs so we did a structure that put us in a matrix environment.

We're working across boundaries, we're working more efficiently and we're making decisions at the directorate level. We are migrating functions like data collection or even performing new functions like developing a strategy for the future of dissemination based on a concept of operations of how work flows between the divisions.

We're completing econ-wide initiatives to help improve efficiencies, developing resource-loaded schedules to manage our survey work, expanding the account manager programs to improve the quality and timelier response and implementing recommendations from the Reimbursable Sponsors Survey.

The culture is really about being more innovative, listening to ideas, crossing boundaries, moving away from a silo approach and so we want to respect and value our employees.

We want more engagement between managers and our employees. We've stood up some boards, the General Workforce Board and Advisory Board so that we can have some feedback from those respective organizations and staff and address common themes, introducing big data, new methodology, creative ways, leveraging alternative data sources in the private sector but all that has to align.

We are going to move to CEDCaP and CEDSCI. All this work will help us towards the ultimate vision. We're calling this the Econ Hub. It's our plan to ensure we remain the authoritative source for nations' economic statistics.

So it's a huge transformation effort but we tie and link everything to the Hub and the economic directorate. It's been fun. It's been challenging but certainly the staff has been heavily involved.

We couldn't have done it without our number of volunteers, with the general workforce as well as with management so more to come. We're still looking to even engage employees and provide different opportunities as we move forward.

((END VIDEO))

William Bostic: So this is a video that we – I'm – that I did for the staff to keep it – the reorganization, the reasons, the objectives of why we did it, to keep a focus on there, keep a focus on our progress, how well are we doing, where we need to tweak and change so that's what I focused on the last two years.

That kept me here for the last two years. I certainly didn't think it was fair to turn the directorate upside down and walk out. So we're going to get started with the presentation.

I'm going to cover the Economic Census. Nick Orsini, Assistant Director – he will cover the Census of Governments and talk about our efforts in improving census statistics.

So I'm going to give a overview of the Economic Census. It's – got it. Something ain't right.

Nick Orsini: Okay.

William Bostic: So the census is the most comprehensive data collection of the business community. Certainly it provides the foundation for the national income and products accounts of BEA and it's the benchmark for the GDP.

It certainly serves as our frame for our samples that we draw from the Business Register, so very important and we conduct it under Title 13, Section 131.

So the – I'm still on this slide. Okay, thank you. So the Economic Census – our core programs – we cover 18 sectors from the North American Industry Classification System.

The key economic data items are establishments, payroll, employment and sales/revenue information. We cover the 50 states and the District of Columbia.

We also conduct an Economic Census for the U.S. Territories with the same key data items covering Puerto Rico, American Samoa, the Commonwealth of the Northern Mariana Islands, Guam and the U.S. Virgin Islands.

You're going backwards. Next slide please. Need you to go to the next slide please. There we go. Thank you. We conduct periodic surveys on a five-year basis - the Commodity Flow Survey, which is the only survey that we actually collect either in the years ending in 2 or 7 so for - the 2017 Commodity Flow Survey we will actually begin mailing out in January.

It – we collect sample shipments on a quarterly basis for calendar year 2017. The business expenses supplement – we actually add the expense information to our Annual Survey of Wholesale, Retail and the Services sectors.

The Survey of Business Owners – we conduct that sending out to almost two million companies. It's a sample survey and it's comprised of employers and non-employer companies.

Next slide please. So the video gave some background of when we reorganized. We really have like four pillars of the Economic Census. We did a reengineering effort and we knew it would take two censuses to conduct all of the findings to reengineer the census much more efficiently/effectively to accelerate the data products.

So we focused on four elements for the 2017 Economic Census. The other objectives we've deferred to 2022. Certainly we leverage the current surveys – our annual surveys to kind of test a paperless collection for the 2017 Economic Census core programs.

We actually have cross-trained our analysts, so in our economy-wide statistics division where the Census of Government and Economic Census in our annual surveys are housed, the analysts work on the same industries for the census and for the annual surveys.

And we prioritize kind of the objectives for 2017 to ensure we stayed on schedule. We're really doing a lot. We are leveraging project management tools such as Project Server, having a work breakdown schedule where all who are working on the Economic Census on a daily/weekly basis – they put their time in on the tasks to ensure that we are following the schedule, making adjustments where necessary.

Next slide please. So the four pillars for an efficient 2017 Economic Census – so for the core programs and this is not including the U.S. Territories but the

18 sectors that we're coving (sic) – covering, we're moving to 100% electronic reporting over the Internet.

We're looking to reduce the burden for businesses so we are looking to use administrative records. For the retail sector we want to actually conduct some research to see if we can leverage third party data for the retailers.

We are doing some research of – for some retailers of getting direct feeds from them rather than filling out our collection instrument. With the census now in the economy-wide division we're kind of automating our operations and the processes from the survey lifecycle.

We tended in prior censuses to standardize a lot of the collection efforts but on the back end post-collection we classified differently by sectors. The look and feel of our industry reports/geographic reports were different.

And it was really just a lot of inconsistencies on the back end so now we will process the census like a program for the entire economy, and with these changes we are looking to accelerate data products of the evolving economy.

Next slide please. So in 2012 we received 53% responses electronically. It was really one of the main objectives in getting through sequestration. In 2013, which was the year of data collection for the 2012 Economic Census, we took a \$28 million cut.

The electronic reporting really got us through. We got the data in faster and we know that it's cheaper than paper forms, so for the 2017 Economic Census made the decision we were going 100% Internet.

So we think we will get more self-responses. We think we are getting cleaner data and we'll be able to improve the coverage and data quality. So with the 2012 Economic Census we actually went all electronic for the Survey of Business Owners and that – again that sample was of two million companies for employers and non-employers.

And we conduct that survey in two waves and for the employers we got – we actually received 90% electronically. They could request paper by demand and we finished like 91% electronically for the non-employer portion, so it was a good sign that we could go all electronic for the Economic Census.

So when we look at some of the strategy that we wanted to try to test in advance for 2017 Economic Census we leveraged the annual surveys. So we did some tests with the Wholesale, Retail and Services Annual Surveys for 2016, conducting tests with the Annual Survey of Manufacturers and our Company Organization Survey.

And in January of '17 we plan to do what we call our Refile Survey, but it's the Economic Census Industry Classification report to ensure that we have the establishments within a – within companies we have them in the right industry.

So I'm going to talk a little bit about some of the results of the tests that we actually executed. So in the way of the unit response rate, and that's where we are receiving a response back from the company, for two of the three surveys when we do the 2013 and 2014 comparison where we only sent out a letter and not the paper form, we got the data in faster and the response rates improved with the exception of the Retail Survey, which is the – on the left column where it was a slight decrease.

Next slide please. However when we look at what they will receive checking in from electronic reporting/Internet reporting we almost got close to 100% in all three surveys when we do the comparison.

You could see where we were with the 2013 survey, and so with the annual surveys we are just about 100% electronic reporting by actually deploying some of the strategies that we are looking to do for the 2017 Economic Census.

Next slide please. So reducing the burden for businesses. As I mentioned we're looking to leverage more administrative information, so for the construction sector for the Economic Census in the past it was actually a sample survey.

The methodology was slightly different in how we did our services and manufacturing sectors, so we are actually now using administrative data in the same way that we do for the other sectors of the economy.

Third data – third party data sources. So we're conducting research with NPD. They have seeked (sic) permission from the retailers that provide data streams to them, and we are trying to compare the data that they send to NPD to see if we can leverage and - how close is the data to what they report typically in our surveys – our annual survey and our Economic Census.

So we are researching four large companies. We will look at their 2012 Economic Census reports and our annual surveys - 2013, 2014 and 2015 annual surveys and see how close - the data that they reported to the data that they provide to NPD.

If we are successful where the data really seems to be really comparable, we plan to try to leverage this approach in the 2017 Economic Census for some of the large retailers.

And we're looking to – for the 2017 Economic Census as part of the whole reengineering effort to try some different approaches for the companies to try to reduce their burden so that we could look to replicate more of these approaches in the 2022 Economic Census.

Also we are – we've had a number of focus groups with companies talking about the electronic reporting, and some of what we heard was the need to actually mail out for the census a little later than what we do.

So in past censuses we've mailed out usually in the third week of December, so for 2012 Economic Census it was the third week in December of 2012 before the holiday and we had a return date of February the 12th.

So I think we did for ourselves mail out just before the holidays to get it out of the way, but certainly we know that the companies did not look at the request and it'd wait until usually after the holiday.

So for the 2017 Economic Census we're going to mail out the first week of January and request a due date of March 15. Next slide please. So part of the strategy for some of the annual surveys – we actually did due date reminder letters.

It increased the timeliness of the response and it reduced the need for more costly follow-up so that was really helpful. If you look at this particular slide/this visual you could see where we used the reminder letters and the results as the information came in, which caused us to do less follow-up on

the back end, which is certainly more expensive to conduct than sending out letters and reminders to respondents.

So automating our operations through increase in efficiency – certainly we’re taking a economy-wide viewpoint. We’re looking to leverage the Census Bureau data collection and dissemination solutions.

You’ve heard a lot about CEDCaP and also CEDSCI so we are looking to leverage those systems. We’re going to eliminate duplicative systems and processes.

We’ve streamlined our processes on the back end where we have more standardization/better allocation of our staff resources. So we want to present an economy-wide view of the Economic Census where the data are comparable across the sectors.

So these are just some milestones of operations of where we are - looking to complete some testing with the 2015 Company Organization Survey and the Annual Survey of Manufacturers where we’ll start to actually leverage some of the CEDCaP capabilities in testing these – with these particular programs.

And we have been doing a lot of testing of our electronic collection instrument and that is going to really be the key, and we’ve heard that kind of over and over of having a – an instrument that’s fairly intuitive.

It helps, it’s easy to get through, et cetera so we’ve been doing a lot of testing with that instrument. So we’ve had a aggressive schedule. We are utilizing a new metadata repository, which we call CoMET to load into the repository and the rendering of the electronic instrument, which we are using Centurion.

That's where the rendering of the collection instrument is occurring, so we've been doing a lot of testing with the Centurion instrument. We also will have spreadsheet capabilities for the multi-units.

We used in the past for the census a contractor. That application was called Surveya and we thought that we could build the capability into Centurion. Also one of the big new products that will come out of the Economic Census are statistics based on the North American Product Classification System.

That is the demand-based system that complements NAICS, the North American Industry Classification System, which is the supply side and I'll talk a little bit more about NAICS in a few seconds.

So we are doing a lot of testing on the content. We're trying to reflect some new items based on conversations with associations/with BEA/BLS to try to capture some changes in the economy.

We're looking to disseminate more timely and relevant results and we are looking to accelerate from 2012 and it would be somewhere – most of those spots will be – accelerate somewhere between 6 and 18 months faster than what we did for the 2012 Economic Census.

So we've again conducted a lot of respondent outreach in preparation of the content, and we used a number of venues to try to reach out to key stakeholders in the process through the Federal Register Notice.

We have formal meetings especially with BEA, BLS, USDA, some of the other statistical agencies. We leveraged our state data user conferences. While we are still talking about 2012 results we talk about what we're doing for

2017, and certainly they get to react to the content that we plan to collect in the – in this '17 survey.

Trade shows, company visits – we get a lot of feedback through those venues as well. So the new questions - in fact we have - we typically try – we'll test these questions looking at the OMB guidelines as a framework to operate under.

So the U.S. Geological Service, our own Center for Economic Studies – people want to know about water use for the manufacturing industries. We actually collected that information in the 1987 Census.

We didn't get very good results and we took it off of the census so we are – we're testing that question again to see if we get better results. A business cooperative question that came from USDA - and so we've been working with a group to see if we can capture good information about that particular industry segment.

Restaurant feeding question for food and accommodation industries – that came from Retail Federation – are looking at just how the business model is changing.

I think some – a few years ago we talked about factoryless goods providers. We had some questions that we researched in the 2012 Economic Census. We evaluated those questions and we did not get the results that we had hoped to be able to identify factoryless good providers.

So there was a Federal Register Notice issued by OMB and we will continue to research with some new questions for the 2017 Economic Census. We've been working closely with BEA and BLS as they too are conducting research

on can we identify the – an establishment base set of questions that we can identify or is this more an - enterprise level decisions and results where they can answer that question at a higher aggregation?

And we've looked to drop some questions. Either it was unreliable, unpublished historical data, the respondents didn't understand, not economically relevant, very burdensome.

And we knew we were moving to NAPCS and so any questions that were – we deemed really questionable we looked to move them off rather than to continue to keep them on so we could reduce the burden for companies.

So we used some criteria here. These are just some points that we considered; whether these new questions fit the criteria that we determined in working with some of our stakeholders.

If it could meet the majority of these objectives then we would not include the new questions on the Economic Census. You know, we might use another instrument of the annual survey because of - most of our surveys – it's not an establishment base.

It's more at the company level or enterprise level that we collect that information. So we have done a lot of cognitive testing/research of the existing content.

We did an analysis of the existing products that we issued from the 2012 Economic Census. If we didn't get a lot of hits on those products we are looking to reduce or eliminate those products.

Again I mentioned a lot of cognitive and usability testing for the electronic instrument and we've conducted just numerous respondent debriefings. As we received feedback on electronic instrument we've made modifications.

We go back to these companies and continue to work with them and they have really been very cooperative in this process. So the North American Product Classification System - a trilateral effort that was developed by the U.S., Canada and Mexico statistical agencies.

It's a two level system for the Economic Census so we will use about 3000 broad line product or service categories that cut across the entire economy, and then we have somewhere like 4500 product and service lines for specific industries within sectors.

So for an example if you think about a flu shot you can get a flu shot in the pharmacy. You can get a flu shot in the supermarket, clinics, doctors' offices, hospitals.

So you'll actually be able to see for the first time services and products wherever produced or provided across the economy. So you'll be able to see how the economy is changing the business model, the leverage of technology where products are actually now emerging in certain industries that wasn't there before and you'll see the decline of products in other industries.

A perfect example - if you think about the impact of mobile devices and the ability to take pictures, well the camera stores are hurting and - because of that technology.

Publication wise we're looking to release again the data sooner so we'll be able to show product lines by NAICS industry. That's the legacy approach

where you'll see products under or services under a given industry, and those same products or services wherever they are produced across the economy.

So we're looking to again present the same look and feel economy-wide view of – from the Economic Census so that's going – be I think a huge change. I think the data users based on a lot of the conversations we've had are looking forward to these results.

So again as I've indicated the bottom line is that based on the efficiencies, the electronic reporting, getting data in cleaner, consistent and standardization on the back end of the processing for the census we will be putting out our data products anywhere from 6 months to 18 months faster than we did for 2012 and that will be a huge improvement.

Next slide. So we're going to finalize our content in a few months in November. The mail out Classification Survey, which I've called the Refile Survey – we've moved that particular date from November to actually January.

We know we're going to be on a continuing resolution. We actually evaluated and assessed movement to January will not impact our schedule in any way.

The electronic instrument will be finalized in January of 2017. We are going to OMB in March and looking to receive OMB approval by June, and the mail out of the letters to the companies will start in January of 2018. So I'm going to turn it over to Nick for the Census of Governments.

Nick Orsini: Good morning everyone. For those that don't know me my name is Nick Orsini and I'm the Assistant Director for Economic Programs. I guess one of the people who has the most apprehension about Bill's departure is me.

My background is mainly in the international trade area and in the current programs which deal with the economic indicators and some quarterly programs.

For those that don't know we had four economic indicators released this morning – actually five. Census Bureau has two of those that are going out so we released retail trade this morning at 8:30 and we will release inventories at 10 o'clock.

So to go from a program that's on a monthly basis and releasing relatively current data to programs that release data every five years is definitely a different speed for me but I will do my best.

So let's talk a little bit about the Census of Governments. So the Census of Governments is the only comprehensive source of statistics for the economic activity of state and local governments.

Many of the themes that you'll see in these next slides are similar to what you've seen from Bill in the slides for the Economic Census. I kind of think of it just as a different survey unit so instead of surveying businesses we are surveying governments.

So that's the way I keep them straight in my mind and I think of the Census of Governments similar to the Economic Census comparison with wholesale, manufacturing, retail and services sectors.

So I guess the Census of Governments – the way we look at it today began back in 1957 but there are still origins of some of the stats back to 1902. So next slide please.

Okay some of the key concepts are organization and some of the stats are employment and finance. So organization – this is the equivalent of businesses so there are about 90,000 government organizations comprised of about 50 states, 3000 counties, 36,000 cities and towns, about 13,000 school districts and about 38,000 special districts.

Now these special districts are very interesting to me in that the boundaries are relatively unique, and they cover things like metro services and water districts so they're very specialized.

These are also some of the hardest districts - some of the hardest to get information from. So we collect things like revenues, taxes, charges and fees, various functions of these units like police, fire, utilities, parks and those types of things.

We collect expenditures like payroll and contracts and we also collect capital outlays like buildings and equipment. Please. Okay some of the goals for the 2017 Census of Governments is that we're moving toward 100% electronic collection; not quite the same as the Economic Census where we're mandating or kind of eliminating paper but we're moving very closely to that direction.

We also want to reduce the respondent burden for state and local governments, and you'll see in the next slide that we have typically done that with the Census of Governments in that we use a lot of administrative records similar to what we do in the Economic Census.

So we're automating some of our activities and we're going to attempt to improve the data products, and that's the products themselves as well as the dissemination.

Next. Okay so talk a little bit about the 100% Internet collection. In 2000 – the 2015 – so there are also annual programs associated with public sector statistics.

So the 2015 Local Government Finance, State Government Finance, Public Employment and Payroll and Pension System Surveys all eliminated the traditional paper forms, and instead pushed respondents toward the Internet.

There are a handful of remaining customized paper forms being distributed with plans to eliminate those as feasible as it is for the respondents to move toward Internet collection, so we are pushing people in that direction.

Reducing the burden on governments, similar to reducing the burden on businesses, we're relying on administrative records. Currently almost all the state government finance data and half the local government finance data is collected through administrative means, whether that's scraping data from a Web site – well not quite scraping – obtaining data from their Web site or using other sources so in a sense we have been practicing big data activities with Census of Governments for some time.

Next. Automating some operations to increase the efficiency, we're developing a new processing system for the public sector programs. That should be in production for the processing of the 2017 Census of Governments data.

We're also looking into options for replacing other legacy system tools for the processing of the Census of Governments, which starts next year. We're attempting to leverage enterprise solutions like CEDCaP and I hope we are able to leverage enterprise dissemination solutions like CEDSCI.

Go ahead. Thanks. Improve data products to reflect our ever-changing economy but that shouldn't be economy. It should be public sector or governments.

So we're hosting a stakeholders' conference in a few weeks with – in cooperation with the Treasury Department. We're going to discuss our content and the data products with some of our key stakeholders and hope to get some positive feedback or some feedback on what we could do to improve.

One of the unique data products for the public sector data is the individual unit file containing all the information we collect. So we're not subject – the public sector is a little different than the Economic Census and some of the other economic statistics in that it's not subject to the disclosure requirements of those, so we could publish everything that we deem of quality to release.

So here are some high level milestones so I'm starting with the stakeholder conference the end of this month. As you could see we've already begun some of the initial mail outs for the organization or the unit update process.

That happened in February 2016. We're going to do 100% - and we're going to test 100% electronic collection for the employment and finance pieces, and our publication preliminary assessment dates are in 2018 and 2019 respectively.

Okay so currently public sector data is only available in American FactFinder and in spreadsheets on the Web pages but we're trying to improve that. One of the ways that most people obtain data now is through APIs, so we're hoping that we could work very closely with our dissemination team to have the public sector data available in the API for the 2017 Census of Governments.

I think that's a common theme for many of the statistical programs within the Census Bureau. Some additional products - we have preliminary data releases for the Bureau of Economic Analysis and Federal Reserve Board.

We're also leveraging data visualizations in Tableau to help us create graphics/visuals for public consumption, which I believe are much more entertaining and thought provoking than data tables in many cases.

So we're trying to leverage infographics and some of the technology - the new technology that we have here at the Census Bureau. We're also using the social media avenues to disseminate the data as well.

Okay. That is the Census of Governments. Now we'll talk about some of the things that we're trying to do to improve the Census Bureau statistics, primarily economic statistics, and this may overlap slightly with the next session that Carma Hogue will be doing on big data but there are some distinctions here so next.

Okay so many of our efforts have involved our sister agency, the Bureau of Economic Analysis. We're collaborating to identify some of the major improvements we could do that will help some of the BEA publications primarily - GDP of course is one of the primary economic statistics that's released by the Commerce Department.

Having more complete data when GDP comes out will hopefully improve the quality of that statistics – that statistic and reduce some of the revisions to GDP, which the Department has been identified as a task that we need to do.

Okay. So of course there are balancing acts on any new improvements that are resource neutral. So we have increasing costs with stable or even declining budgets but the demand for data is increasing and the demand for more timely, more granular, more geographic details, more data that is linkable with other sources to produce new data products so we have to keep our planes in the air as the phrase is of current production.

Like I said we had several products go out this morning with new research and new products that we want to put out. One of the concepts that we are striving toward in the economic directorate and Bill mentioned it as well is this Econ Hub vision.

And I have to thank Shirin over there for beginning this process in the economic directorate when she was here. It was kind of – I am sure she had the concept well defined in her mind but many of us did not, and I think over time we have kind of clarified and crystallized it a little more at least in my mind.

So it deals with content harmonization, data coherence harmonization, process alignment and innovation of methods. Basically it's the concept of touching a business/company once, obtaining information from it and being able to use that information across several surveys within the directorate.

Also our process alignment - this kind of goes along with the enterprise solutions like CEDCaP and CEDSCI. Our processes should align with these

enterprise solutions so we could feed into them or take output from them relatively easy.

So we're moving toward this. This is a vision of the economic directorate. We have very few dedicated resources to – toward it but we have a lot of, you know, staff as they have opportunities working on this effort.

We also have met with our counterparts in Stats Canada who have this vision as well and have moved closer toward it and are further along than we are, so we're getting some lessons learned from them.

So as far as improvements to our statistics retail trade is our first focus. I guess this is the lowest hanging fruit that we could – we can identify and changes in the sector require new ways to measure and collect the data.

Of course we have all sorts of data pools for retail data out there. We have scanner data. We have credit card data. We have credit card processing data so there's many pools of retail data out there.

So one issue with the retail indicator which was released this morning is steadily dropping response rates, although our response rates lately have been relatively good.

But over time the response rates have been declining and we're looking for ways to improve that; also to increase the granularity of the retail data and provide some geographic detail for this on a more frequent basis than every five years.

So our plan to meet this challenge is – and improve and modernize the census retail trades statistics – the – focus on five goals. One is to improve what we

already do so the monthly retail indicator, strengthen that, make sure that that is highly – that that statistic is – has high quality and meets some of the current uses for it.

Then we're trying to improve the timeliness of some of the other retail estimates. As many may know our advance monthly retail trade estimate is one of the most timely indicators out there – comes out about two – less than two weeks after the close of the reference month.

But many times the data is not as complete as we would like so using some of these other sources may be able to shore that up. And then we talked about the granularity of the estimates, providing some data by geographic breakdown or other more detailed product level information in that.

Of course reducing the burden, you know, and providing some benefit back to the respondents - how could we make reporting to a survey – how could we incentivize that to the respondents – maybe providing some of that information back to them somehow.

And then the fifth goal is a – more of a recent goal but e-commerce is a big topic right now and we do produce a quarterly and an annual e-commerce stats report.

But the report does not provide the data in the way that some of the users want to see it, so right now we've actually developed a straw man for taking the same data we collect on e-commerce, rearranging it and producing some new table, not putting any new tables - not putting any additional burden on the respondent.

We've also been looking at some of these third party sources to see what type of e-commerce statistics they have and to see if they could help us in that avenue.

So quality – so some of the large retailers are reluctant to supply their data and there are several reasons for that. You know, it might be a security issue. It might be a burden issue but one of the primary reasons I think is that it's a voluntary survey, and many of these companies' legal entities/legal parts say, "Any voluntary survey we're not going to report to."

So some of the strategies we've identified or improvements - we've collaborated with NABE to contact some of the executives of the larger companies - the larger non-respondent companies and see if we could work through their – act through the data users in the companies to help with the data respondents in the companies.

Many times data users within a company don't know that they're a respondent and they are not responding to a survey so - to hook those groups up.

Then of course the big data research and some of the research we've done dealt – deals with first data through a company called Palantir, NPD like Bill had mentioned and a major credit card company.

We've done research with all those sources of data. Some of it is very promising. So – and one of the last efforts -- we've just been working on this - - is to get the Secretary of Commerce to contact some of the CEOs of the larger companies and talk to them about reporting to our surveys.

Timeliness. Timeliness is critical for incorporation into other statistics like GDP so on July 28 we produced a new Advance Economic Indicators Report.

The report contains advance data for international trade and the inventories indicators, so research was done on those two indicators over about two years to ensure the quality at this point in time was of enough – the quality was substantial enough that we could produce these reports early.

For the international trade indicator I believe about 98% or 99% of the data is available when we produce this report. And for the inventories we looked – we researched that with BEA, and for wholesale and retail I think overall about seven out of the eight quarters showed to improve the GDP estimate prior to actually going into production with it.

So some of the other activities we have regarding the timeliness of the stats is, one, use third party data. One of the credit card processing companies we've dealt with now has weekly estimates for retail, so trying to use some more timely estimates from third party sources.

And we've been collaborating very closely with BEA to produce an advance services report. Our services report comes out on a quarterly basis and it is not available until the last – to incorporate into GDP until the last estimates.

So yesterday I had a discussion with staff at BEA and our staff here, and we're fairly certain that we could accelerate that release to include in the second estimate of GDP, and I believe we will begin doing that the early part of 2017 so we will have services added to our advance indicators report.

So the granularity of the statistics – data users want to know, “What's happening in my area?” so small area estimation, which most of you are very familiar with is one of the ways that we could do that.

Some of the questions we're asking is can we produce some national retail statistics and at what level? You know, will it be at a state level, an MSA or a county level?

I'd be very pleased if we could do something on a more timely basis at the state level. That – that's a goal for me. Anything else would be a plus. And the fourth goal, reducing the burden - so this is a graphic of the burden of a company who responds to the Monthly Retail Trade Survey.

As you could see, you know, they could have many surveys that they have to respond to. This is a huge burden especially on a small or medium size company, so the concept of the Econ Hub and using data across surveys would greatly reduce the burden on these companies.

Let me skip this slide. Yes let's get that one. Okay Goal 6, improving e-commerce measures. So in the short-term we're exploring ways to repackage our current collection of retail and e-commerce data to produce something that's more valuable to our data users. We have one proposal that has been widely accepted by some outside organizations like National Retail Federation.

We're also working with Council of Shopping Centers on it. And last week at May - or at this week this May, we circulated some of this proposal with some of the attendees there and they provided some good feedback and all-in-all they were in support of it.

So, one of the challenges we have with this is the timeliness. So, to do it on an annual basis, I think, will be one of the first steps but then to have something more timely for the industry and for the growing e-commerce sector is important. So, we're going to explore ways to accelerate what we do for e-

commerce. We're going to also research some of these third party data sources.

Right now we have a request for proposal that is scheduled to go out I believe at the end of this month and that proposal has some information on e-commerce. It's looking for data sources in the retail and services sectors but it's also looking for providers of tools to manipulate that data. So, we hope that we get some information from that request for proposal from some outside sources that could help with our e-commerce efforts.

Barbara Anderson: Thank you so much. One thing that we ask for after the last meeting was an update on the economic program activities and that was fabulous. And I will use my chair's prerogative to say I found that all really impressive and I think US economy should be jumping up and down with excitement. I'm not an economist but I think I'm allowed to say that. And now we will have more detailed and more insightful comments from our discussant Krishna Rao and then we'll have a general discussion.

Also, gang, please write your comments and reflections on your notes stuff, your notes things. This is the discussion of our first subsidy session and I want a lot of inputs so we can have great comments and recommendations to the census bureau. So, Krishna, take it away.

Krishna Rao: Perfect. I guess I want to start by thanking both Nick and Bill for a great update. It, sort of, really outlined the next high level all the different issues you guys are working to sort of what you're pushing for.

So, going to the slide, it sort of - it's clear already that it certainly makes it even more clear sort of the balancing act you guys are in between sort of modernizing and improving a lot of these products while still trying to sort of

maintain sort of control other cost and things like that and it's - go at my first test of the slides, you know, you just come out way really impressed through the level of thought in some of these questions.

So, right in terms of - there's a lot of ground covered to rather to sort of trying to talk through in general terms. But it does rounds the heads sort of just a few specific questions and comments on different areas that were touched on during the presentation.

So, what was around this move towards internet collection both in the sort of that core economic census and also in the census of governance which I think is probably exciting for a lot of different reasons. So, in terms of sort of timeliness and data quality but also in potentially sort of reducing respondent burden and sort of a future down the line where you can imagine sort of increase customization, automation of some of the ways in which the respondents interact a lot of these surveys and census.

I asked a few question around especially when we think about the economic census which I understand is moving towards for about 100% model where sort of electronic is the mode of sort of responding. As I was thinking through it, I was wondering if maybe from the prior economic census if there are any learning that sort of, you know, who is more likely to thought these electronic surveys versus the previous paper versions and what impact that might have on sort of response rate for things like also for in the effect of retail trade where you've seen the client response rate.

Is there some worry that we need to think about the way these transitions towards 100% internet collection might have an impact there? And also broadly, there - I also want to - it also seems on sort of the more positive side are really rich area for understanding some of the challenges of things like

respondent version which I think is often (unintelligible) to sort of put your hands around in a quantitative perspective.

But your ability sort of look at how people flow through the different sections of a particular survey and, you know, collect data on where they're breaking off or where they're spending longer to answer a particular question might give a bunch of sort of leverage and learnings on being what people sort of breaking off some of the things and some measurement of none response.

In response to this, (Berts), I was wondering if you have any thoughts there or learnings from previous versions of say the economic census.

Also, I have a few questions about the North American sort of per classification system and sort of this introduction into the economic census. There's something I'm super excited about from sort of data perspective I think. You guys mentioned there's a lot interest in sort of being able to look at items in sort of a different way. They were able to think about them previously and really understand sort of the changing dynamics of the economy in lots of ways.

I do - maybe this was more smoke than fire though. I do remember at some point there were - I can imagine at some question about, you know, kind of business owners think about sort of mapping the goods and services they provide to the category. It's often not trivial task to try to think about how to properly categorize these things and as you're sort of wrapping this up and pushing out, it'd be - you know, there's some mention the slides about sort of the testing of the functionality of this sort of broad system.

I'm just wondering if you guys can maybe give us a little more insight into what you think is working well and what isn't working well in terms of the

sort of categorization of things about this new system. Probably sort of a theme across the entire presentation is the idea of getting sort of more data quicker and some of the metrics around here were really impressive so I pulled one up from the slides which is the idea - so, for the economic census you're talking about sale statistics available at industry levels.

You know, 18 months earlier than just five years ago which sort of begs this question as you talk through, you know, improvements to both moving to 100% electronic, improvements in the process and sort of how this data is sort of pre-imposed process. Maybe understanding sort of where do these large games come from, right? Is there somewhere to have decompose that a little bit and then think about how you might imagine applying it to other areas of the census bureau.

On the sort of five-year plan, I wanted to focus a little more on (unintelligible) as the first section knowing that a lot of the later sort of goals are probably talked about what they did the section as well and - I think a previous version of slides that I saw made a point I think that we've talked about in the group before that, you know, surveys are remaining the bedrock of sort of the retail sale statistics.

That's sort of the first goal of sort of improving the quality of that monthly retail indicator data probably is sort of the most important in many ways as a goal. So, I wanted to also, you know, the declining participation rate is worrying, right? And it's - I think you guys came up with a - I tried for a while to think through or what - I don't know what can we suggest that might be sort of new ground to hear it, it's tough.

There are a lot of things that you suggested seems like the right things to do like increase outreach whether through NABE or through the Trade of

Commerce all these things. They seem like good directions. I wonder - their ability to move and deal. It's just that - it's a tough problem and I think it can help but certainly coming from, you know, internally as you guys have mentioned, legal things are legal things, right? It's very hard, it's often not a question of what companies want to do but more a question of what certain people or company can get comfortable in doing, right?

It's more of a binary in sort of an issue of sort of desire effort often and it's hard to know what - you know, I tried thinking through what are other models for sort of improving participation in this sort of data things. And a lot of them just don't seem to, you know, there's lots that you sort of give to get models or more clearly defining benefits for respondents. And it's just tough to find the right fit in the space. Though unfortunately, there are no real answers here but just a - a real answer to this. It's a tough problem but a central problem for what's going on here.

Yes, and probably it does seem like this was a little into the slides, a big part of this have to be sort of emphasizing sort of the benefits for these companies of participation and maybe understanding a little bit of - because there are already set of companies that it feels like you guys are having a tremendous amount of engagement with both a high participation rate but also some of these other initiatives around sort of real time data or credit card data.

So, understanding sort of what - how are these some set of companies clearly able to get very comfortable with a very high level of disclosure and what is it about those structures or the legal process that's going on there that allows them to have that flexibility that maybe other institutions are feeling more a little hesitant of. I think those are the last slide I have. So, I'll leave it there. I'll just get your response.

Barbara Anderson: I think it might be good idea to give Bill and Nick a chance to respond to Krishna's comments and question before we went to questions and comments from the committee generally. So, Bill and Nick, whatever you want to say.

Bill Bostic: So, thank you very much for your comments. They were really right for us to really consider some of the things that you - how I pointed out questions that we have really given some thought too.

So, historically, economic sense is we focused a lot on paper and we used 600 report forms. The product - filling out the product information which typically on the forms where long list of products by industry sector was really burdensome for a lot of companies. So, with the electronic instrument, we made it really response driven. And we have some - given that we don't have paper, we have something like 800 plus pass based on the responses for a particular question.

So our design strategy for the electronic instrument was to pre list kind of the common products in the industry, as we do have some historical data, and then have the respondents actually do writings on the remaining products to account for all of their receipts.

Our youth ability testing thus far has shown what the pre listed products they like that they can go in, et cetera but they would actually prefer to do writings than to come up with a search function. And the reason why is because we don't always use their terminology to describe their products or services, though we are used to handling the writings and then identifying particular product category that they should go to

But it also allows us to evaluate the writing and add to our dictionary so that we can kind of lift those in the common industry or service categories that we

have. So, that's what we've done with the instrument in looking at the implementation of the net products and services.

So, for the large companies as well as the medium and small companies when it's response driven and they don't get to see a lot of questions that don't typically pertain to them which takes up our time and sometime frustrate them along the way. So, we have really seen some results, nice results by making the changes based on the feedback and then we go back to these companies to see if we got their feedback correct in implementing into the electronic instrument.

So, that has helped a lot. We have an account manager program for the larger, very large companies. Though some of the large companies cut across the economy in various vectors. They appreciate having one contact to help them for collecting the data, helping them within any definitions, granting extensions, et cetera. So, that has always been well received by the large companies.

So, the 18 months earlier, for some of the data product which the key for being able to accelerate is really the electronic instrument because we're building in edits into the instrument which allow us to get cleaner data. Certainly as an example when we look at our monthly international trade report, we processed like eight million transactions every month.

When before we looked at the evolution of when we had paper, where we were like maybe 50% paper, 50% electronic usually there was at least 50% failure on the paper side which took up a lot of time. As we move to 100% electronic reporting, as Nick indicated for the advance report, we have 99%. The data comes in so clean that the revisions are really minimal and it allow

us become much more efficient and we were able to actually leverage the same resources to be able to produce this advanced trade report.

But the key was the data comes in clean and we on the back end have to do less follow up with the companies. And that's what we're looking forward with this response driven instrument will with built in at it. We know typically sort of the issues that we see on the paper side and one of the biggest errors even electronically that will really produce less observation by analyst.

Some errors we get is because we didn't put like zero fill like the dollars, thousands and millions. And just zero filling that it's saying that we really don't want dollars and cents. It's going to make a huge difference in our ability to edit the data quicker because we won't get much errors.

And this time around, we are really looking to target the observations as they come in and able to shift resources where we see more edit failures for this particular categories or industry to shift our resources to attack that. So, to have some kind of prioritization as we review our data.

In a way of trying to prove response, right. Having the secretary involvement, we do think because of her relationships in the private sector it might help, is it going to be sustainable in a long run? I think that's a big question. But we think the key is really the value added back to the respondents.

When we do have focus group sessions about the economic census and we really talked about the census being integrative with other census bills statistics like the ACS, our company, business patterns. When we showed the census business builder, these companies really would get excited because then they could see how they could use the data and the importance of

providing responses and they knew where their responses what it was contributing to.

Before, from the focus groups sessions and I think of the things that is difficult for the economic census, we only go out and do a really 15% direct collection to the business community. So, it's the largest enterprises in a largest single unit input. The remaining 85% of the 29 million business employees, the non-employees we use administrative records.

And we really try to work without intermediaries in helping us push up the response but also point to the value of the data. So, we've been thinking about doing a better - a much better job of linking the why with the response - which our responses contribute to the output, how this data can help you to - with the response strategy.

As Nick indicated, some of the issues are, we can prove how the data are used and what it contributes to but their counting - or probably he doesn't care. They just see it as a task that they have to fulfill and for annual in census mandatory helps. But for the single large business establishments and the smaller business establishments, I think that's a huge insight for the companies that they could use this information in a way that's beneficial for them. But it's the large companies that contribute most to the estimates.

So, the other approach is really trying to review their burden and since they produced data, using various accounting software because that's another approach that we want to talk to the SAPs and accounting software companies or perhaps I'm generating some standard format that we could leverage that the companies could benefit and get direct data screen. So, we are researching that approach.

So, we're looking just ways to kind of mitigate to the client in response rates surveys. We're going to need to continue to help us with modeling and to bench mark. So, I hope I covered most of your comments and questions.

Barbara Anderson: Thank you so much. Now, can I have other questions and maybe we could get several and then you can all can respond to kind of everything. So, Ken?

Ken Simonson: Ken Simonson. Two quick comments and two questions. I want to echo what Noel said that Bill you've been a terrific resource for the big day to working group and also recruit NABE. For those of you not familiar with that acronym, it's the National Association for Business Economics. Bill and Tom mentioned before him would have been graded interacting with the business economics organization and individual companies or associations that they represent.

I think it's been very helpful for us to know what census is bringing forth and also what challenges you've had and it has enabled us I think to be more effective in communicating to congress the importance or funding and supporting independence of census on the economic programs and look forward to having Ron step in to this role. And I want to thank also Ron, and Nick, and Carma for their work with the big data working group.

Questions. Bill, the budget challenges that you face in the last year, can you say a little bit about what you've had to either defer or give up on trying to do in this is very broad and modernization. And then Nick, it wasn't clear to me if your role has changed or is about to change since you discuss the census of government, at first I didn't know if that was going to be your sole focus or you're going to be involved with the rest of the modernization that you presented.

Barbara Anderson: Other questions, comments? Noel, Allison, (Jeff), Sunshine.

Noel Cressie: Noel Cressie. So, thanks very much to both of you, Bill and Nick. So, as a general view of this that when we have things like - we talk about things like sector, granularity, timeliness, I translate that to motive areas, spatial-temporal, right? So, for the area estimation is often due to just geography but of course the general notion can be carried to motive areas, spatial and temporal.

So, what I'm getting at is that this goal that you have, certainly a great goal but there are - and they are resources that you're looking into that go beyond the actual economic census to get into using credit card information that we've been talking about in the big day to working group for over a year now. And a little less about the timeliness although that's always in the back of their mind this economic census every five years and have an opportunity to do something perhaps in a fixed monthly or yearly level. But don't forget the motive areas.

It's tempting to take one sector and then work with that and come up with summary statistics about one sector. But those sectors correlate and they correlate strongly and the whole secret behind small area estimation it's correlation be at motive areas, geographic, or temporal. And don't forget the timing to see if the temporal aspect because it's dynamical process and so what you see in this current order or this six-month period will have a large influence on what you'll see on the next six months taking into account non seasonalities.

So, what I'm getting at is this huge opportunity here methodologically to introduce motive areas, spatio-temporal or dynamical modelling to get to sort of the answers that you want and I think some of this going on in the

methodology area and John Abowd I think has got some plans in that area. It's very exciting, great to hear about and I think there's a lot of opportunity both methodologically and for an economic census point of view in this area.

Thank you.

Barbara Anderson: Allison?

Allison Plyer: Allison Plyer from the data center in New Orleans. Yes, I'll echo what others have said about obviously the tremendous work you all are doing to get better data, faster and with restricted budgets. And you know being a data user at a local level, the - one thing that we found was for example we want to figure out what - you know, what percentage of our businesses were minority owned for New Orleans MSA.

And I don't know if others have tried to get a data like that, but the 2012 data became - on that became available sort of in the latter half of 2015, right? And so, you're talking about moving it up by eight months which is great. And obviously there's a lot of work involved in that and which is what you're describing.

And from the data user perspective that data becomes not very useful, right? Whether there's that much of a lag and I guess I'm just always been curious I know on the population side that decennial census data comes out, you know, even the small area stuff comes out more quickly. And then I'm sure you have major obstacles and I know that there are budget issues but it would be great to hear more about how come some of those lag so much.

Barbara Anderson: (Jeff)? Andrew?

Andrew Samwick: Andrew Samwick. I just had two questions and comments. The first, because I've never worked with this data but I'm very interested in how large online retailers that are geographically concentrated like Amazon's, you know, distribution warehouses which are, you know, not spread out uniformly but are certainly taking out a lot of things that would otherwise be in retail sales at a geographic level.

I'm interested to know how you're working with them and what they do to the normal presentation of data? I think that's particularly important for people who are tracking employment by sector because they're having a large impact I think.

And the second comment is about burden and follow up. I mean, in all the things you have control over of really of what census does or maybe what the Department of Commerce does. But I think part of when we talked about burden we have to talk about the burden that all information seekers are placing and I would just back in your area that treasury and labor are also prompting for information.

And I suspect some of the follow up that you have to do they also have to do as well. So, it's just my semi-annual plea for more coordination in where possible across government agencies who are essential contacting the same or similar units with similar needs.

Barbara Anderson: Sunshine and then Roberto is on the phone. And since he's great that he's participating on the phone right after Sunshine says, he can say - it'll be wonderful for Roberto to say what he's thinking about.

Roberto Rigobon: I'm here.

Barbara Anderson: Sunshine should - could go first and then Roberto if that's okay?

Roberto Rigobon: Yes, yes.

Sunshine Hilligus: Okay. Sunshine Hilligus. So, first I just have to ask, did you started in the census bureau when you're like 5, like how is it possible you've been here for that long?

Bill Bostic: I was 2.

Sunshine Hilligus: So, my question is actually not about the work that you're doing but in thinking about the fact that you're stepping out of this goal and how the work that you have done in transitioning to an online survey might inform the work that is happening for instance with the 2020? I recognize that a lot of the, you know, exact questions are entirely different and on so.

But particularly given efforts to kind of move into, you know, a survey, move things under, you know, one umbrella. Is your experience with that transition something that is informing and it can inform the other work within the bureau?

Barbara Anderson: So, Roberto on the phone, whatever you're thinking about.

Roberto Rigobon: So first, Barbara, you see, Sunshine is always first before me so, yes. I was looking at the slide 46 again thinking about the burden that we're putting them on the companies and I am wondering what exactly are we asking the companies to do? I mean, sometimes we would like to reproduce electronically the form that they are already filling in the survey.

And that may not necessarily be the best approach. So, I think that in order - when you are contacting the companies I think it would be so important to understand what exactly we want from them. Example in retail sales maybe we just want one number at the end of the month.

And therefore the question will be that we expect exactly what we need and then try to find what is the best way for them to provide us. So, for example, if we want to get an estimate per week, I don't know how many retailers we'll actually compute the weekly data on their total sales, is sometimes asking for this kind of data might be incredibly worrisome for them in so many has kind of market share data outside the company and, you know, they will treat that very carefully.

So, maybe asking for, you know, in a weekly data just impossible for them to provide, they just don't compute that and they don't want to provide this kind of data. So, what we name in about situation and therefore that's where the lawyer will say well, do not participate before taking any risk we prefer not to participate.

So, in that conversation it would be very important to understand that exactly what are their needs from the statistical offices are. Also to have the coordination, someone that spoke just before Sunshine who's very clear the question understands, we contact them and if we coordinate and we ask this is the information that we require and this, you know, this is a public good that we're doing so they might be more willing to participate.

But if we are all contacting this companies in a disorganized manner and then were trying to reproduce the forms that they are already filling I think that, you know, we would find more resistance.

So, my advice to that is we need to have more conversations with them. And I understand that - how also the Secretary of Commerce may be very influential. My view is that that might be influential for the CEO that is currently there. You know, it's the CEO move and then five years down the road then the quality of data is slightly to deteriorate.

So, in that relationship, I think my advice is that we start a conversation where we understand what the needs of this statistical evaluation is and we space that out to them.

Barbara Anderson: You can see by all the comments that we're really interested. And then the six minutes that you have, if you could respond or comment on all these things or any selection you feel like.

Nick Orsini: Okay. So, I'll just answer a few of these things. Try to answer them anyway. My role - I guess my role will be whatever Ron wants my role to be. I think I'll, you know, pretty much be doing similar things to what I'm doing right now. But me and Ron have to work some of these things out I guess. And then Allison, you talked about some of the data at the local level or some of the characteristics of data, like, we produce in our Survey of Business of Owners which adds characteristics of some of the businesses.

So, that was typically a program we did every five years in conjunction with the economic census. I think we just released our Annual Survey of Entrepreneurs which had very similar data to the Survey of Business Owners. And that is now an annual program. This was a joint project between census and the Kaufman Group and also the minority business - agency within the Department of Commerce.

So, that's available now on an annual basis. We hope it continues to be available on an annual basis. So, that - it's a little bit more timely than the five-year program.

Man: And we just released 2014 data.

Nick Orsini: Yes, we just released 2014 data. And we will be releasing 2015 data I believe later this year -- trying to catch up.

Bill Bostic: So, budget initiatives - so for the FY17 budget we had a budget initiative for data modernization and we had a budget initiative for the acceleration of the economic indicators. We did not get either one of those initiatives in - at least from the house of (infinite) markups.

For the acceleration, we have determined that we're going to still move forward for the acceleration of the quality service survey which Nick talked about that we plan to implement next year. For the data modernization, we're continuing with our research. The RFP, we're hoping to purchase some data and/or to help us with our big data efforts.

We're moving slower than we would like to but it's a priority that we continue working with BEA to help with the GDP revisions and looking at innovation and methodological ways of producing the statistics.

In a way of, you know, the opportunities for the modeling in the small area estimation, the points that Noel suggested, we do have a new director for the big data center here that's sitting in the back and we'll introduce him in the next session. But that's where a lot of our research in the modeling kind of how we will play in the sandbox to address a lot of the issues that you raise. So, we're hoping to do that.

In a way of the burden for all the information providers, there are a couple of projects. And Ron Jarmin has actually been taking the lead on working with BLS and some major companies, really approaching them as a joint statistical agency along with DEA to really look at the burden that we place by when we individually ask for data that could we do something collectively that would reduce their burden in that regard.

In a way of what we're doing for the 27 economic census and whether or not that will be information that can feed 2020, in some way it's almost been the other way around because 2020 has conducted a 2015 test, they conducted a 2016 test in looking at electronic reporting, looking at what kind of support you need to support electronic reporting. We've been actually getting feedback from 2020.

And there are some - if we think about group quarters, conceptually group quarters is very similar to the very large businesses that have multi units in looking at the instrument to allow the capabilities for group quarters to provide their information.

So, we've been really looking at more of a corporate function rather than a separate solution for group quarters and a separate solution for the multi units for the large businesses. But we have been actually working together in gaining some information from the various decennial tests as well.

Some of the comments that Roberto made for our monthly retail, we basically ask for two data items and we know they have it. We know they have sales.

In a way of kind of the weekly feed scanner data, we actually are really trying to leverage other data sources. We know that these large retailers provide

information to other third party. And so, what we have been trying to incentivize in the process is can we leverage these other data screens that they provide to other entities in lieu of them having to fill out the economic census which is really a big burden for them to provide the information on our instrument.

So, a number of companies have been interested in hearing what we have to offer. And the key is, does the data from these other data sources - will they really be comparable to what we need because in some instances the data that they provide to NPD is somewhat different than the data that they provide from a financial standpoint to fill out some of our collection instruments and even from the IRS.

We are having a number of discussions about e-commerce in doing a number of company visits because the problem with e-commerce is that companies define e-commerce differently.

For example, that we know that for some organizations, you purchase your product online. But if you have to return it, you return it to the store. But that e-commerce sale still stays in e-commerce. And so, in some instances the brick-and-mortar sales get deflated and e-commerce is a little inflated.

So, we're doing a lot of research so in regard to the whole e-commerce situation and see if there is some comparable theme that we can focus on rather than trying to change every year every time the business model changes. But it's very complex. And we've been talking to a number of companies that they're even struggling within their organizations. And the key for us is how do they keep it in their books. So, I know we're short on time so I'll stop there.

Barbara Anderson: Well, thank you so much. We clearly could have gone on for at least two or three more hours talking about this. But you're going out with a bang, Bill. So, thanks so much.

Bill Bostic: Thank you, all.

Man: Thank you very much, Bill, Nick, and Krishna and everyone. We have a break starting now at 11 o'clock. But the committee members are requested - I think the official photographer is here. (Michelle)?

Woman: (Unintelligible).

Man: And I - well let's see. Within three minutes, you're requested to assemble I think at the sign outside, just outside the door near the blue sign. And we'll come back at about 11:15. Thank you very much.

Woman: I got to go to the restroom.

Man: All right. I'll hold it.

Woman: It's all done.

Man: It looks like it. With you - Barbara's coming. Barbara's coming.

Woman: I was just saying that - I spoke with Bill.

Man: I know. I saw it. Behaving very nicely...

Woman: Thank you so much.

Man: Thank you for saving that six-minute left for me. That really had an effect, a positive effect. Thank you very much for coming back. The next session is a big data initiative. So, presenter Carma Hogue who's a director of Research and Methodology of Economic Programs. Carma?

Carma Hogue: Okay, thank you and good morning to everyone. I think we'll go right past this one and to the next slide. I'll have a clicker thing. Okay, okay. So, some of this you heard earlier in the first session and - but I'm going to go through it quickly because this is mainly picking up big data pieces of it. These are major activities that we're doing in modernizing the economic statistics.

In particular, we want to conduct research on big data sources, methods and procedures. We want to implement the new data sources into retail data products. That's going to be our first focus. And in order to do this, we have to re-engineer the business register infrastructure so that we can absorb and process the alternative source data.

And then another thing that we need to do is to harmonize the items and the unit content so that we can streamline our survey collection and enable passive collection.

Okay, so I'm going to touch on five major projects that we're working on. We do have a very small team but we have a lot going on. And so, some of the things you've heard already today but one of them is we're going to be exploring third party passive collection and then also system-to-system collection and accounting and payroll software collaborations. I will be going into a lot of that -- each one of those individually.

Then I'll talk about continuing our development of a web crawler/scrapper then purchasing more third party data and wrapping up the payment processing

company pilot projects that we had going. And I talked about those at the last meeting. And finally, research and machine learning so that we can improve imputations.

So, first I'm going to go through all of those what we're calling "passive collection." And this is something that actually has kind of taken off this summer. And the first one is system-to-system collection. And what we attempt to do is to plan to explore the possibility of having a company respond, want and let that cover multiple economic surveys that are collecting the same data.

And so, in doing this it could be something like they might be in the monthly or quarterly annual census. And can they just respond to the monthly? And then we would add things up to the other levels or it could be that they're in multiple surveys but they're all asking one common item like sales.

Now, in the latter case it - what we have to make sure that we're doing is harmonizing our definitions and our terms across all of our surveys. And we talked a little bit earlier about the hub and that is one of the things that they will be doing as they're looking across all surveys and trying to - if we say sales in our publications it means one thing. And if we are using a different concept, it will have a different name. We also need to look more as we're doing the system-to-system collection. We need to explore how products would be collected for the econ census.

So far, this summer we have contacted three companies and they are very interested in participating in our research. We have held initial teleconferences with all three of them. And we have started our personal visit. One of them was in August and we have another one in a couple of weeks. And we - it's too early to say that we have any results. But it does - based on that first one,

things are very positive. And they're gung-ho and we need to kind of catch up a little bit with the - they're being patient with us and we're catching up with them on some of these things.

With these three companies, I think we'll learn a little bit because one of them's going through a merger and one of them is going through changes in their system, their IT systems. So, and then the third one is kind of a normal one.

Okay, next thing is third party collection and Bill mentioned this this morning. We've signed a contract with the NPD group and we plan to compare the individual company data with census data -- what we are getting. And in the past, we have looked at their aggregate and that was not as helpful as some of the other data that we've looked at. But we're thinking that perhaps the individual companies that they have reporting to them may be able to also help us.

So, we will get data from up to five companies under this contract that we have. They're a variety of companies. Some of them are national, some of them regional -- different geographies. And in approaching these companies to see if they wanted to participate in this research, NPD was telling them that they would be helping census research-wise to reduce respondent burden. And that's how they are getting the participation with consent.

I find this one to be exciting. And in fact we started this two to three years ago. We have developed a web crawler with machine learning that learns the location of relevant data on public websites. It scrapes unstructured data and organizes that data.

This started at a time when I was working in the public sector. And so, we - in that public sector, you got a little bit of a feel from Nick on that. They - a lot of times the respondent will say, "I have a comprehensive annual financial report or a CAFR. Just go and get your information from that." Well, this has always been a manual process and quite laborious. So, that's when we got the idea of trying to scrape this unstructured data because every CAFR is different.

And in particular, we started by looking at tax data from state and local governments. And so, that is what we've been scraping so far. Of course, all of this is public data. And we used Apache Nutch and Python to build a combination web crawler classifier that will take those PDFs that contained relevant text data.

So, the work on this was presented at the joint statistical meetings in August. The developers of this crawler and classifier are (Brian Dambacher) and (Kevin Capps) and they're the authors of that paper that was presented. They also have a research matters blog out on Census.gov. This thing has been picked up by some of our sponsor newsletters. And also, we've had some interest in the demographic side of the house on this crawler scraper. And we are trying to come up with some kind of name for it.

Since it's kind of taking off, what we're trying to - what we have done is we have a server now in the CAT lab and we're building a prototype that can expand the crawler's usage.

The next phase will take it for public sector's usage. We have to take the data that we've scraped and it has to be cross-walked into their terminology. And so, that will be the next phase. It is not an easy phase.

So, there's another web scraping data collection project that is going on. And this one is researching the availability of online building permit information for jurisdictions that are in the survey of construction sample. And they are focusing on the large - the jurisdictions have large 2015 residential construction. And it's - they're focusing on the large ones because these are the jurisdictions that are more likely to have the online data available. They're looking to complete the top 30% of the sample.

And one of the things that this group is doing is they are looking at cost data so that they can investigate the cost saving on this. The result so far, they do know that there are some sources, data sources available for large jurisdiction. They're available in all kinds of formats. The problem is that there are inconsistencies in the classification.

So, for example the way that these building permit offices classify single family or multi-family or square footage or things like that, they don't quite match what we are using.

They have a high confidence in the timeliness of the data and the validity of the data. It's coming right from the permit offices. But the classification and getting the characteristic details would be something that's causing a little bit of concern.

This next one is certainly in its infancy. And this is to look at partnering with software companies to see if we can possibly have a module or something built with the major financial or payroll software. And they would be able to just push a button and they can get what we need for the census report.

The American Payroll Association has a representative who is extremely interested in wanting to help us pursue this work. And so, we will be - he

actually listened in on one of our meetings this summer. And he is very interested in talking to others and talking to these software companies to, you know, kind of talk on our behalf. And he was also possibly talking to BLS or other group agencies. And with this, we have retailers of all size that will be using the software. So, it could have some bang for the buck there.

So, that is what we've been doing on the passive collection side. And so, now in the big data side, as Bill said earlier we had a request for information out earlier this spring. And so, we got some response back from that and decided to put out a request for proposal which will go out hopefully by the end of the month. We were hoping to purchase something by November and we will be looking at the quality of the data that are coming back in. We're looking mainly for data and then secondarily for tools. And then we'll determine what to purchase after we look at the quality.

So, what have we been doing with the retail third party data? One of them is looking at small area estimation models. We want to improve estimates at evidentially a state by industry level - NAICS code possibly three-digit. But to get to that, we are looking first at nation by industry. And what we're looking at is trying to use aggregates of sales from credit card transaction data.

And then the second thing that we are focusing on is examining trading day weight calculations and holiday adjustments using daily data seasonal models. And where we can, we will work to improve the X13 trading day weights.

So, we have had a collaborative exploratory project with a software company. And Nick told you this morning that it's Palantir and a payment processing company, that's First Data and BEA and census.

And so, in this we are able to get to some consumer spending data. This First Data covers 58 billion transactions annually and that captures about 45% of all point of sale transactions -- it's credit, debit, pre-paid gift cards everything but cash. Now, we got a sample of that. We got five states for this little pilot project.

We were also looking at the tool that was - the software tool that was available. And that tool can update transactions within a day. And we were also able to use our Python through the tool. And so, we were using that to research trading day adjustments and small area estimates.

So, in the small area estimation model, we were using Fay-Herriot model and we were examining the reduction and the variants of the estimates. And that's how we were evaluating the model. We had a few limitations in the part of the small area estimation as a survey data. We were using monthly retail trade survey. We needed it at a state level. And so, we have investigated a little bit about how to improve that. The other - the synthetic part is coming from the First Data.

And what we have found though is that this proof of concept shows that it is possible. It will be. And this First Data are useful even - and what we are using are aggregates as our covariance.

Also, as a part of that pilot, the trading day weight calculations and the holiday adjustments we were working with (Tucker McElroy) and (Brian Montell) of the Research and Methodology Directorate. And they were producing daily data seasonal models and then comparing what we got from that from the daily data to what was coming out of current X13. And they were able to make some improvements to the X13 based on that.

We first looked at the Easter Sunday effect but there are other holidays that just kind of pop out at you when you're looking at the daily data -- that would be Superbowl Sunday, Chinese New Year, Ramadan, Labor Day and Cyber Monday. And these are all things that we will be looking at.

Just to let you see some of it, right there. These are shoe stores. And what you're seeing here are Sundays from October of 2012 up until April of this year. The red on there, those are the Easter Sundays.

And you can tell that it's - you can see quite enough fact from this graphics. There's nothing like visualization. And this is another project that is in its infancy. We kind of touched on it a little bit earlier. This is for the advanced monthly retail trade survey. When the March does not get it - get that data from large certain units, then it's a little tedious to manually impute based on historical trends for the advance.

So, what we're doing is we're looking into several things but one of them is machine learning and another one would be somebody from the RNM director we'll be looking into special temporal work. So, we are looking at machine learning methods that can take advantage of large number of variables and observation from March and from the advance stand from the monthly retail trade survey. We are considering looking at regression trees on their own network. We have already received a server dedicated to this research. We just now need to get started on it.

So, that's some of what has happened. We have been told that we need to do a little bit more outreach. So, here's some of the outreach that has happened this summer. We had a paper - the European conference on quality and official statistics, (Rebecca Hudgitsen) and (Brian Dambacher) presented that. And so

the paper itself is in the proceedings of that conference but it was also picked up by the Spanish statistical journal.

Also this summer, the WebCrawler was presented at the joint statistical meeting so the paper will be in that proceedings. This summer - this is internal to this building but we had a joint census in BE (unintelligible) methodology seminar that was given on big data and it basically covered what we did in that pilot study.

We have proposed a session for the 2017 joint statistical meetings and invited paper session. And then we have submitted for next year -- 2017 conference on statistical practice -- a couple of e-posters submissions. Another thing that came out of that conference in Madrid -- the quality conference -- we are now meeting monthly with Eurostat, just exchanging ideas on big data and monthly video conference and then these other people that we have collaborated with this summer.

Next steps, retail trade data. We will be awarding contracts so that we can get some more third-party data. And then we'll continue looking at the models that we're working on. On the passive data front, we will be working on that commonizing the questionnaires that is crucial before we can go any further on this system-to-system. We'll start looking at that NPD data that we'll have available shortly.

For the WebCrawler, we'll be collaborating with others to work beyond the public sector. And then we're also going to work with public sector though to go into Phase 2 and cross work what we've scraped to what they can use. And as you - so I will be working on more presentation in publication.

Barbara Anderson: Thank you so much. And our discussant is Roberto Rigobon and I really appreciate that he is a discussant and has sent in his PowerPoint and is doing this via telephone. And it'd be hard to think of a better qualified person to be the discussant. So, Roberto, we are eager to hear what you have to say.

Roberto Rigobon: Thank you so much. Let me start right away. I spent a day in the census. I learned about this project and during that day when I was meeting with (Carl Masteen) and then I have to say that I was very, very impressed. I think it's -- if you go to Slide Number 1, the general overview -- I would like to mostly concentrate on two of the projects that I learned the most which were the tax web scraping project and then the retail payment company that they're analyzing for the holidays and other things.

Very, very impressive work especially the WebCrawler and the tax, I think, is a brilliant idea. I think this is going to have a massive payoff for the census. So, that's a very high payoff a project. I congratulate the census for doing that. You know, right now, you ask what are the taxes that were collected, what are the tax revenue collected here on the states, we do that manually.

We have to go through very a difficult process where every single entity that has collected taxes has a report that we have to put them together, we have to organize them together right now in the census because, you know, what if somebody calls sales taxes maybe another person just put that on a different part of the PDS.

And so, it is a (unintelligible) to compute right now and I see this idea to go through the web first to find all the documents that have the word "retail tax" or "sales tax" or "utility taxes" and then classifying those documents as source relevant documents where tax information is there and then downloading that

information into a database and reclassifying, I think, it's a brilliant use of the possibilities.

And this is a very directed process of data collection. So, this is a - what I mean is, this is using some tools of machine learning for the classification of the documents. So, that's a part that - is the machine learning but once that is done then the collection is almost like a survey. So, I think the data quality that will be obtained from this project is very high. I think that they have done all 50 states but, you know, you have to go through all the local governments and all the entities that are collecting taxes so it will take a while to go through the whole United States.

But this will allow us first to understand the granularity to collect, you know, state levels, you know, county levels. And, you know, when I was staying there in Washington - on the census, I was, you know, you get greedy when you see the reports - I have actually looked at several of the reports, several of the states like Alabama for example, has a magnificent report, by the way so because very clear, very simple to understand, very simple to collect and classify the data.

So, in the same way that we're collecting tax revenues, you know, maybe in the future we can also collect data on how the money is being spent, et cetera. So, this is very, very good. The one on the retail sector, I also think is very good. So, the very nice - the big difference between these projects is one is the data is public and, I think by law, these entities have to put the data on a monthly basis. And, therefore, you know, in that sense, there's no issues of privacy.

On the payment system, this is a private company and one of the things - I think what they have done is they mean to see this - specially to use this data

for the holiday adjustment. This is important especially, for example, in some of the holidays we don't necessarily all coincide with the same month. I mean, Labor Day is always in September. So, if Labor Day implies higher sales then that will always be in September and, therefore, in that sense, this September seasonality will pick-up that sales.

But you see, Easter doesn't always - is in the same month. So, therefore, you know, March - we look like a very bad March if Easter gets go into March as opposed to April. And, therefore, because it's switching from one month to the other, understanding that switch might be important for seasonal adjustment at the aggregate level to understand what are we producing.

So, this is very important and we search to understand and - I mean, you saw the graph, the effect of Easter is maybe at least on shoes. So, people don't buy shoes on Easter. We learned that on the summer as well. And so, one important thing is to understand how we adjust the data and this project the way the data is collected might help a lot of that.

But the data we have accessed from this company is not exactly the sales, it's the index that they produce. So, in some sense, still it's very unclear to me how the data was created. So, you know, I like a lot what they have done but, you know, I have a lot of questions about the data by itself. And so - and I believe that the data we're finding with the, for example the holidays, it is in the data. So, I don't have any reasons to doubt any of that.

But, you know, we have indices, we have no idea how much smoothing has happened, how much imputation takes place. I have a lot of questions about that data and maybe, you know, once the engagement changes we would be able to understand a little bit more what is inside the data. But so far, these two are very, very impressive.

And I think it will have the census and the BA -- of course the second one a lot -- for correct - in fact, one of the things that happened after my visit is that I started thinking a lot of about the engagement - a lot of about how the future will look like. And so, let me just raise some questions. And the reason why these are important by now is because we're going to move into some project at the census toward that stage.

I was thinking about one of the incentives on the private sector for doing this. So, first is what do we want to collect from the companies? One is we just want raw data and then we will just compute inside the census the statistics that we require or we just want to compute sufficient statistics and the company just send us sufficient statistics and what we issue what privacy versus public. And I think that depending on the top of the data implies a different form of engagement.

Finally, the important to me is in the case of the companies, you know, how are we going to compensate. So, what are they going to - how do we say it? When I thought what this means, I actually realized that we have a kind of vicious cycle. So, let me explain the two.

First - wait a second. I'm going to the slide on assumptions and I want to make some general assumptions, just in general. I will assume that in general we don't need - if we granulated all the data that is inside the company. We're interested in certain aggregates and geographical statistics because - and those will be important.

I don't think that census requires for example every five second sales from a supermarket for every single product, for every single client - I mean that sometimes we all need the full extent of the granulated data for the things that

we consider important. So, I'm going to make that assumption that we need certain aggregates. We don't have everything.

A lot of that we're interested is that we want to improve on validating data they are already collecting. So, part of what we're going to do -- like the holidays -- is that we're going to use this data, not necessarily to use it directly into the estimation of seasonality but to understand if we can improve the methodologies that we already use to correct for that seasonality for example.

So, in some sense, again, we can learn something from that data but not necessarily going to compel to use that data for the computation of the statistics. So, I think that we all agree in it was used for the presentation. I have not seen the previous presentation but I agree that a lot of what we want to do is to reduce the burden on the firms and also the statistical offices. And also, a privacy and I think the integrity of the data provider is crucial issue that we always pay attention.

Third is we're going to have any access to data -- I hope that is London -- and, you know, that the census I think it runs here for the next 200 years. And so, it's important how are we going to guarantee their reliability in the collection and how can manage that into London.

And finally, something very important to note, everything can be solved through a mandate. And again, the voluntary participation is very important. There are other countries, you know, I come from one that, you know, if you want data we just create a mandate and force people to give the data. That happens in a bad country like mine. That doesn't happen in respectable nations. So, I mean, some things can be mandated but not everything can be solved through a mandate.

So, that means we need collaboration. On those assumptions, let me think about the forms of engagement. First, when the data is public data, like for example the tax revenue project is light, I think that the nice thing about that -- because the data is public -- there are several things that can be done. First, we can collect the data and therefore we have made the data available - I mean, this project - they're using web scraping to collect the data. I think that one nice thing is that because is public.

In terms of the size of the date, we don't necessarily need to bring the data into the census. We could be computing this to the Cloud and you're saving the sufficient statistics. I understand that the way we're doing it right now is that the data enters the census but in principle there's no need to do that.

We only need sufficient statistics from that and that's what we need to be recording. Maybe we want to save some data in case a mistake has been made to go back to data but in general - if the data has become too big, the census doesn't need to totally invest their resources. We can do that in, I think, per place.

However, if something very important - if the data is public, it doesn't mean that time series is public. And this is something that I've been thinking a lot about that. So, if this happens to retailers - if the price is all the retailers are public but the time series is not. And, therefore, the question is when you collect the time series now you present the time series publicly, are you violating some form of privacy or not?

And that, by the way, I don't have the answer. But this is something that it will become an issue. The more we collect public data and organize it - so again, in the case of the taxes, I think it's very, very safe. The taxes are public; the time series is supposed to be public as well. And so, if we organize it and

present it in public, I think it is no issue. So, that's the brilliance of that project.

One very nice thing about the public is at that the burden is mostly a technical issue and in fact, the census has been very aware about that. So, when you connect to the computer, you make a request, for example, to download a particular document. It's very important that we respect the speed which we can connect, that we not decide a company - since if I do that consistently because the computer that is doing this in automatic way might look like a hacker when it's actually not. It's just downloading the data.

So, informing what we are doing and explaining what we are doing, also deciding what time of the day is better like 2:00 am, very early you have request PDF from the State of Alabama at 2:00 am so we can reduce technical issues and in fact, a lot of way pages clarify at the speed which you can call them. So, be respectful of that.

But another thing that is important is that we reduce the fear of how the data will be used if it is public anyway. So, showing some sense -- those aspects -- who has my data and who will use the data is a small - and finally, because the data is public, we are not expecting any financial compensation. I mean, there's no compensation that is expected.

So, when the data is public, I think the biggest aspect here are the technical aspects that we don't crash the server because we made too many requests too fast. And the other thing is, you know, is the question of the time series. Okay, the data is public, is the time series equally public or not. So, it's relatively - this is relatively benign problem.

I think this changes dramatically when we go to the private sector data. So, in that sense, this next slide that we tell payment company - so one is, you know, how to access data. And I think the census is exploring on the system to system is a very clever idea and hopefully it can be done. The question is, are we going to download the whole data or just the sufficient statistics. And that's an important question.

So, are we going to ask the companies and we can provide them a program to the companies and say we just want you to compute this particular statistic in this particular way given the way you are organizing your data internally and therefore that's the only number that we require, you don't even have to talk to us. The program will automatically send that to the census once a month so forget about it. That would be - the burden will be very small but then the question is, you know, will the companies be willing to do that.

I know there are aspects that - so that's where the computation takes place - there's the computation takes place to the company or takes place to the census. And there are advantages and disadvantages. In a second, I will try to do that.

The second one is how we compensate the firms for the data. So, do we pay for the data? And, therefore, we pay for the data. Actually, this would be non-incentive compatible or we actually do it through collaboration? And then - and if we get the raw data then the question is how can we guarantee privacy and make sure that at least the fear isn't there. I mean, we, at the census, can guarantee the privacy. We know we can do that but then the question is do the firms will believe that we can do that?

So, maybe just actually the two - this is the next slide and I used to talk about the two extremes. So, we are going to collect the private data, we are going to

collect all the micro data. So, in the sense we just - the idea is that we will do the computation in the statistical office.

So, we will go to our webpage, we will get all the raw data. We do the computation at the census, for example. One important thing is that the knowledge of how we produce sufficient statistics, how we came to the data and reside - we've seen the statistical office. The burden on how we are going to do the computation will be on the statistical office so it is complicated or not.

Guaranteeing privacy will be a concern for the private sector even though we know - even though we can guarantee it internally, it's not clear that the company will respond favorably to that. And then if there's a compensation required, will you stay for the data? And in some sense, that's the end of the story.

Now from here, what I've done is just to put, you know, what are the advantages and disadvantages in colors. These I want to show - the next slide is when we download sufficient statistics -- we go two slides down -- there would be sufficient statistics. So, this is - instead of us, the raw data, may I ask the companies to do the computations. So, the computation will take place at the firm level. We will collect sufficient statistics.

Now, the knowledge here is actually will be in collaboration. So, in some sense the IT knowledge about how the statistics was computed is joined. This would be a very small burden for the company and in the sense that the sufficient statistics is there. Privacy is better computed but actually this is not incentive-compatible. And the reason is that is -- I think I put that on the next slide, the idea is on the next slide.

So, we might collect the data and the micro data will be the following steps: first, we do the data collected for each firm then we do the aggregation and computation that will take place at the statistical office. Again, as I said, the knowledge will reside in the statistical office and (unintelligible) some economic and social indicator are really public and actually highly valuable.

So, if we produce a better number for GDP, all that would be very valuable. In fact, the financial sector will be saying to that that it would be very valuable to actually produce a better number for (unintelligible), a better number for GDP, a better number for inflation, et cetera.

So, in some sense, we collect the data from the firms, aggregated in the statistical office and produce something that is a very high public value. And in fact, the financial sector also that this is very valuable. Now, the problem is that - because we are collecting - of the privacy at the level of the firm - the next slide - we will have an issue with privacy.

So, we can solve this in two ways: one is to produce a mandate. And the mandate will create a resistance and a burden in the companies. Another thing is to pay. However, if the data that we're producing and the statistics that we're producing is very valuable for the financial sector, the company might want to say, "Well, I would like to produce it myself." And now in that sense, the price in which they are willing to sell their raw data will be so high that we will not be able to pay.

Furthermore, they may want to put a very high price because they want to protect themselves from privacy. So, they are concerned with privacy so I said, "You know what, yes, I'm willing to sell but I will be willing to sell at a price and I know you'll going to say no." And in that sense, the promise that if we produce something that is actually valuable and in some sense it's

(unintelligible) or it has a - the perception that comes right to the privacy of the companies, they might be willing to put a price that is too big.

That's actually worse when we produce a sufficient statistic, reminding that we said, "Okay, instead of actually collecting the private data and the raw data, let's ask the companies don't worry, you produce the aggregate so, you know, you don't have to sell me the raw data." So, the privacy is a little bit more guarantee. So, we develop an indicator in collaboration with the company.

Now the knowledge is shared and the aggregation - what we do is just aggregate all the companies in the statistical office. And again, we'll be able - in the end, produce the same economic indicator that is very valuable. The problem here is like - because the knowledge is shared.

And if the public data is valuable, then the question will be for one of the companies just said, "Well, this is a public good so all the public goods problems will appear. I want all the other companies to share their information and I just don't want to share mine. And mine will be sufficient statistic of what is my topic."

And therefore, again, if we're going to pay for the sufficient statistic, the private sector doesn't have any incentives to share. So, I don't have a solution. So, this is kind of a depressing part. But I have been thinking -- since this might be fit to a census -- I have been thinking how to engage the private sector in a way to contribute to a public good, we have a mandate and we thought the data being public. And I think it would be hard.

I think that this example with Palantir will be a very good on our first day, will be a good step to understand how this engagement again move ahead. I

think whatever the sense the census is my final comment. I think whatever the census - you see, I said actually in the slides, I don't have any answers.

Whatever the census finds, I think will be a very good guidance for the rest of the statistical offices in the United States and by the way maybe around the world. But we have to find a way in which we did the collaboration. And so far, I see that the alternative of paying for the data will be a complicated step.

And for the moment, this is the way we're going to operate, I understand that. But in order for this to be sustainable, I think we have to think a little bit harder. The other alternative our collaboration with these companies is actually valuable for the companies in a way that, you know, we teach them something that they didn't know. Again, the question would be once that learning has taken place, why will they continue engage in with this census?

So, in that sense - this survey, in that sense solve that but I don't know how much we compensate the companies but it's just - we just do that by asking many, many times until they respond. That's -- more or less -- the way, you know, the mother - my wife tells with our kids. She sends them texts until they call. And so, if you send enough texts, they will call. And so, the idea is in spite the - the non-response ratio in my kids is worse that the non-response of the census but my wife has a load of energy.

So, like the census, you get an answer at the end. That thing makes it more challenging to do, when are we going to try to do this electronically. So, let me finish in a positive note. I think these two projects that I learned - I'm just looking all these two because I kind of learned from those two more are extremely, extremely good. And I think as a proof of concept this has a very high payoff for the census and the BA in the very short run.

So, I congratulate the census and the BA for working on those two. And Carma's team I understand is a very small team but in a very short period of time they have done a fantastic job. And again, they are going to engage with these companies a little bit on the medium term. And it will be very interesting that we pay attention to that engagement. And we think one of the ways to improve that relation, to make it long-lasting, in a way that again, it's incentive-compatible.

So, what we don't want is for this engagement to be very fruitful at the beginning and then, you know, two, three years down the road the private company decides is not worth continuing with it or that they find an alternative. So, that would be worse. Okay, let me just stop there. Thank you.

Barbara Anderson: Thank you, Roberto. And Carma, thank you for your excellent report. And you made huge progress. I think we're going to have to proceed to the next part. But I thought - I personally thought that the slides you've made are fantastic and thank you. And I'm sorry we don't have more time. Thank you so much.

Man: Yes, thank you very much Ms. Carma and Roberto. Now, it is time for our lunch. Committee members are invited to over the next ten minutes, grab your lunchboxes here. And I guess the next session, we'll start at about 12:20. Is that all right? That's ten minutes from now? Is that good? Okay. So, it's a working lunch. The next session is a big topic -- the 2020 census updates -- and it has a big cast of presenters: Shirin Ahmed, Deborah Stempowski, Maryann Chapin, Atri Kalluri, and Patty McGuire.

I don't know if everyone's speaking but everyone's here. So, we'll start with. Who will we start with, Shirin? Who's first?

Shirin Ahmed: Yes, I'm going to start.

Man: You? Okay.

Shirin Ahmed: Well, good afternoon everyone. It's great to be here. My name is Shirin Ahmed. I'm the assistant director for Decennial Census Program. And I think I talked in April because (Lisa) wasn't feeling well. I'm happy to report that she's actually on a vacation this time -- a long well-deserved vacation -- and will be back in the next couple of weeks.

So, I thought I would start with some general program updates before we got into the key presentations. And I wanted to start with by sharing some key management changes we've actually made within the decennial census directorate.

So, one of our top priorities at the Census Bureau is succession planning. And part of preparing for the future is giving current executives opportunities to serve in new roles which opens up opportunities for the next generation of leaders.

In May, we did make some changes within our decennial senior leadership team and I just wanted to review those with all of you. So, (Tim Trainer) -- who many of you know -- is now the Census Bureau's chief geospatial scientist known as a worldwide leader in the geospatial community. This new opportunity allows him to focus on serving as an ambassador of the Census Bureau in the geographic field.

(Dierdra Bishop) -- who's sitting right there in front -- is the new geography division chief. She's actually returning to the geography division after providing great direction in moving the 2020 Census Program forward. Her

extensive experience and insights will be highly valued as she furthers the work of the geographic programs and aligns them with 2020.

Deb Stempowski stepped in as the Decennial Census management division chief. Deb is right there. My notes say she brings decades of survey and management experience. She can't be that old. But she does have quite a bit of operational experience and she also brings a census-wide perspective to the Decennial Census management division. Deb was previously chief of our American Community Survey office which brings me to the last change.

And that's (Tori Valcoff). I think she's downtown today. (Tori) has been detailed as the acting American Community Survey office chief. She was the division chief in our demographic directorate. She brings strong technical skills, strong data skills and experience and will really help ACSO if it implements its agility and action plan. So, these folks are fabulous leaders and they've been off to a great start in their new roles within the Decennial Directorate Program.

So, many of you are familiar with this timeline. So, where are we today? This timeline shows our major activities from now through the end of the 2020 census. In fiscal year 2016, we really transitioned from the research and testing phase to the operational and implementational phases for the 2020 census program with our efforts progressing to the 2018 End-to-End Census test as well as other activities.

And I'd like to quickly mention several activities in 2016. First of all, we've made many notable decisions since the release of the operational plan last October some of which are documented in the 2020 census memorandum series.

We've also continued to make key operational and systems decisions throughout the past ten months and will continue to do so in a timely manner as we look at the evidence from the test result and consult with expert opinion.

Since the release of the operational plan, we've actually made about 96 decisions. Two key decisions that were actually announced as part of the 2020 Census Memorandum Series and posted on Census.gov are as follows: first, decennial device as a service. We decided earlier this year to employ the device as a service option in which the census bureau will award a contract to a company that will provide devices and the service contract for enumerators to optimize their cellular coverage nationwide as they do their work.

We made this decision because we determined that this service would be a lower risk relative to the bring your own device strategy that we were testing previously yet avoid the costlier government purchase of all of these devices.

A second key decision was made in May and that was in regard to the Census Enterprise Data Collection and Processing or CEDCAP. Here we decided to use a commercial off-the-shelf platform for core collection capabilities. Based on our rigorous testing and evaluation, we decided to implement the integrated commercial off-the-shelf platform to supply the solutions that we needed for both data collection and for case management as well as a solution that would allow for us to incorporate some of our in-house innovations such as the optimizers that we've developed.

This solution brings the experience of an industry leader with a long history in the market to support us for the 2020 program. And you're going to hear more about the analysis and assessment that we did and the vendor when both Patty McGuire and Atri speak a little bit later.

This decision allows us to stay on-schedule in the development of variety solutions for the 2018 End-to-End Census Test as well as helping us remain cost-effective.

The turning to the address canvassing test in two weeks we'll be starting our upcoming address canvassing test, recall that test as being conducted in Buncombe County, North Carolina -- that's near the Ashville area. And part of the city of St. Louis, Missouri. This test will operationalize our address canvassing method.

Again, the purpose of this test to measure the effectiveness of the in-office address canvassing through in-field address canvassing to measure the processes for in-field address canvassing themselves, to understand the implications of moving from assignment area to basic collection units and to conduct an in-field relisting to collect data to refine future quality control operations.

We'll update about 7,500 blocks using about 300 field enumerators. These staff will be using laptops to accomplish this task using our listing and mapping instruments. This is pre-platform technologies our mobile case management, our optimizer, and our map TIGER.

On the residence criteria and unique situations, I think John gave us an update on that. So let me then talk a little bit about our key contracts.

The census questionnaire assistance contract has two primary functions. The first is to assist respondents by providing information about the questionnaire or general questions that folks have about the census. And the second key function is to actually collect the data over the phone, something that we're actively encouraging in 2020.

This contract was awarded on July 11 to General Dynamics Information Technology or GDIT. And we're often running there.

To support the national head count in 2020, the Census Bureau was planning an integrated communications program to increase awareness and participation in the 2020 census. Effective and strategic communications with many diverse audiences will be crucial to the success of the 2020 census including everything from educating the public to maximizing our response rate.

This contract creates awareness and facilitates participation (unintelligible) racial and ethnic groups across the nation including the U.S. territories. And we're very happy that we awarded that contract on August 24 to Y&R -- Young and Rubicam. And some of the senior officials are there. And we're very excited to have them here with us, and we're often running there as well.

For technical integration, this is a very important contract in terms of our systems. This contract supports all of the design and architecture engineering and integration activities. For example, it's the integration of the System of Systems for 2020.

The contract covers infrastructure planning and design for the data center capability, whether it be physical, cloud, or some combination; the field offices; our data capture centers; and the island areas and any other designated locations that we need. Additionally, the contract provides the design and disaster recovery solutions for the 2020 census System of Systems. And then lastly, the contract provides a specialized expertise for us, for example in fraud detection, in the creation of synthetic data and other items that are critical to the successful conduct of the 2020 program.

We awarded this contract on August 26. It was awarded to T-Rex with many folks that have previously had experience on prior censuses. And T-Rex has partnered with many other vendors including Leidos and Accenture.

The Schedule A Human Resources Payroll System - I can't believe this. My phone's ringing. No one ever calls me. Excuse me.

The Schedule A Human Resources Payroll System - okay, this contract will introduce a lot of efficiencies into the human resources processes by automating all of the recruiting, the hiring, the on-boarding, and the separation. All of that work is very manually done in our regional offices. This automation will also streamline with our payroll capabilities again for all of the work done by the Schedule A hires, which are all the enumerators in the field.

This automation will not only help the decennial program but it will be used for all of the hiring we do for the special censuses as well as our reimbursable and ongoing surveys. The RFP is out for that and we're looking at an award date of mid-November.

And then finally, for decennial devices and service, while we've had a couple of small contracts to understand the requirements, we will be awarding a single contract in February of this year for the 2018 end-to-end test and the 2020 census itself. RFP for that effort will be released next month.

So, another announcement that was made at our recent PMR is that we announced the site locations for the 2018 end-to-end census test. They are Pierce County, Washington; Providence County, Rhode Island; and Bluefield-

Beckley-Oak Hill, West Virginia. The 2018 end-to-end census test has a census day of April 1, 2018.

And the purpose of this test is to make sure all the operations and systems work together so we can conduct a successful census in 2020. We'll be testing and validating the operations, procedures, systems, and infrastructure for both our functional and non-functional requirements.

Additionally, we will be producing a prototype data product as well. And we chose these sites because they provide a great variety in urban and rural areas across the country. And we are very excited about doing that test.

And then lastly, in addition to the contract and operationalizing our procedures and systems, we're making great progress at all the planning that goes on for the 2020 program across our 34 operations. Earlier in the year we chartered and established all the integrated project teams. Those project teams are developing the detailed operational plans.

We've released the first detailed operational plan as a 2020 census memorandum that was released actually last December. And we're releasing nine more detailed operational plans by early October.

And finally, we are making updates to the 2020 census Operational Plan to keep an alignment with all the decisions and changes to the operations and systems. We will be releasing that operational plan very shortly and we'll be talking about it at our October PMR at the end of October.

So those are some very brief highlights of where we are. I'm going to hand it over to Deb Stempowski and Maryann Chapin who will talk about the 2016 test and preliminary findings. Thank you.

Deborah Stempowski Great. Thanks, Shirin.

I just wanted to take a short, quick moment here before I got into the 2016 test. (Unintelligible) the committee members who were so helpful to me during my time in the American Community Survey, helping us make sure that that national treasure got to a good spot.

So I appreciate your support in that effort. I'm sure I can count on you as well in this position.

So, before we get into the details of the presentation on the actual test, I wanted to take a step back and just remind everyone about why we actually conduct the test. As you know, the conduct of a decennial census is a major undertaking with many moving parts, many more than I realized, it seems, from the outside looking in.

As we planned for the operational design for the census along with tried-and-true procedures that we know work, we developed new methods, new procedures, new systems, and new solutions so that we can leverage new technologies and methodologies and take into consideration the world around us both from a social and an economic perspective.

So although we would love it if the test ran smoothly and we could sit back and kick our heels up, we do expect to learn from the test and probably hit some speed bumps in the road. We want to discover what aspects of our solutions work well -- that's always nice. And we want to discover what isn't working well so we can change that.

We want to uncover things. Some things we expect; some things we don't expect. We want to leverage those tests to explore aspects of the operational design that require refinement. And we want to take advantage of being in that test environment to add new dimensions during the test as an opportunity arises.

So as we share our findings - our preliminary findings with you today, you'll hear a lot about what we've learned, and our analysis is still under way. So there's much more to understand and inform the 2020 census based on these test results.

So my job here is to frame and remind everybody about the location of the test, the purpose, the scope, and the schedule. The data collection, as you know, concluded. That ended earlier this summer. And to be consistent with Census Day, of course we used the April 1 - the 2016 test.

We had a site test in two parts of the country -- Harris County, Texas; and Los Angeles County, California -- and combined that with about 225,000 addresses in each location. These sites were carefully selected for a number of reasons including language diversity, demographic diversity, high vacancy rates, varying levels of internet usage, and the multiple locations across different time zones.

Okay. So our purpose in the 2016 test was to refine technologies and methodologies associated with two different operations, self-response and nonresponse follow-up.

So in self-response, we wanted to make sure we were providing non-English support for respondents with limited proficiency in terms of both our contact strategies and the response options available to them. We wanted to form

partnerships and conduct outreach efforts to reach historically hard to count groups. We wanted to refine our non-ID processing, and we also wanted to test languages utilizing non-Roman characters such as Chinese and Korean.

And the nonresponse follow-up objective, we wanted to finalize the strategy for using administrative records in support of that large operation. And we tested multiple ratios of enumerators to supervisors, 30:1 in Harris County and 20:1 in Los Angeles County. We want to implement advanced procedures for conducting interviews at multi-unit structures and be a proxy, which you'll hear more about. And we wanted to continue to refine the use of our technology and automation to reengineer field operations.

And so just in summary before I hand it over to Maryann, highlight these handful of things. We had multiple mailings to encourage self-response. We had partnerships to reach across demographically diverse populations and language support there for limited English proficiency populations. We had that internet response and of course real-time non-ID processing methods.

So over the next few slides, Maryann is going to share with you some of our preliminary findings.

Maryann Chapin: In the upcoming slides, I'm going to present some of the preliminary findings from the 2016 census test. You'll hear about things that went well but you're also going to hear about some of the problems that we encountered.

First, so I'd like to review the self-response contact strategies with you. During the 2016 census test, we used five panels to help determine how best to contact households and solicit their responses. The test will assist us in determining which contact materials we will use to encourage self-response during the 2020 census.

Each of the panels received multiple mailings. For example, Panel 1 received a letter on March 21 followed by a postcard on March 24, et cetera.

The fifth panel involved people receiving a paper questionnaire during the first mailing. This is a decision that we made following our census tests in 2014 and in 2015 where we recognized that certain populations like the elderly really wanted a paper form from the get-go. And so this was a targeted mailing.

The last column shown on the slide, Contact 5, occurred between mid-May and mid-June. Any addresses that we removed from the nonresponse follow-up workload as either Administrative Records Vacant or Administrative Records Occupied received a final postcard that was one last opportunity for them to respond to our test.

With this slide we show some of our preliminary refinings - findings around the response rates as of the end of our data collection. This includes self-responses that we received after the start of the nonresponse follow-up operation. Our overall self-response rate in the Los Angeles test site was 53.4% with a 31.4% internet self-response, a 2.7% telephone self-response, and a 19.3% paper self-response.

In the Harris County test site, our overall self-response rate was 39.6% with self-response rates of 27.4 (Unintelligible) 2% and 10% for internet, telephone, and paper respectively.

Data are being processed and our analysis is under way, but there are some key takeaways that include sending a letter instead of a postcard as the first reminder had a positive impact on improving response rates; providing more

language services such as in a brochure or a frequently-asked-question insert seemed to have a positive impact on response; and as we had seen with our past tests, the 2016 census test reinforced that in some areas of the country we do benefit from sending a questionnaire as part of the first mailing.

Our plans for 2020 include sending for about 20% of the addresses a paper questionnaire in that first mailing. All other areas will be sent a paper questionnaire in the fourth mailing if they have not responded by that point in time.

Next, with partnership and outreach, in the 2016 census test we did work to form strong partnerships and depended on our partners to help encourage response during both the self-response and the nonresponse follow-up phases. In the Harris County site, we had 258 partners and in Los Angeles County 262 partners.

Partner ownership specialists supported recruiting, were engaged in a limited partnership effort in language support, and had a social media presence to help the communities know that we were testing in their areas. We also have a Statistics in Schools presence. In order to assist our partners in generating awareness and to encourage response, partnership toolkits were developed for both the Los Angeles County and the Harris County sites.

I'm not going to cover a lot of detail on this slide. But in summary, the 2016 census test self-response, nonresponse, and other materials provided support in multiple languages.

Shown on this slide are preliminary findings from a self-response perspective. Our test results show, for internet 97.5% of respondents responded in English. Less than 3% of the respondents responded in the three other panel languages,

Spanish, Chinese, and Korean. For telephone, over 80% of respondents responded in English. And less than 20% of the respondents responded in the three other panel languages.

Now moving on to non-ID processing, non-ID processing was again offered to respondents in the 2016 census test. And although we know that the majority of respondents will respond with an ID, non-ID processing creates flexibility and freedom for respondents to respond anytime, anywhere.

In the 2016 census test, we did not sample specifically for non-ID. The non-ID response in the 2016 test was only from those respondents who chose to respond without their census ID.

A sample of the non-ID responses was selected for inclusion in a response validation. Those selected non-ID cases were included in our field data collection workload. Enumerators recontacted those respondents during the nonresponse follow-up operation and conducted the nonresponse follow-up interview.

The data collected during the nonresponse follow-up interview will be compared to the data received or collected during the non-ID response to validate the accuracy of the non-ID responses.

Shown on this slide are our preliminary non-ID processing results. The table indicates the overall self-response and then the specific non-ID response totals. Consistent with previous census tests where the non-ID option was not promoted. The proportion of non-ID responses fell within the 3% to 4% range.

Also, while not indicated on this slide, the proportion of non-ID cases from internet response and census question assistance was consistent with that of previous tests with about a 90-10 split between the two modes.

Our preliminary analysis of the 2016 census test results indicates a cumulative match rate of 95% for the non-ID cases within the two test sites. This was achieved with a 91% matching during the response, an additional 2.5% during subsequent automated processing, and another 1.5% during the manual or clerical matching.

Now we move on to nonresponse follow-up. The 2016 census test did provide us another opportunity to operationalize new methods, a new technology that we could use during our response follow-ups. We had a number of objectives for the 2016 census test related to the nonresponse.

Specifically, one of our goals was to determine the nonresponse follow-up contact strategy that we'll use in the 2020 census. In the 2016 census test our primary contact strategy involves Administrative Records Vacant removal prior to any contact attempts being made, Administrative Records Occupied removal occurring after one unsuccessful attempt to reach a respondent, and a maximum of six contact attempts with cases becoming proxy-eligible after the third contact attempt for all other cases.

We had another objective around the refinement of the field management staffing structure as Deb mentioned. At Harris County we implemented a 30-enumerator-to-one-supervisor, and in Los Angeles County a 20:1 ratio.

We also tested new multi-unit accessibility and contact procedures. And then finally, with the 2016 census test we were able to test the nonresponse follow-up re-interview functionality for the first time.

Critical to our success in utilizing administrative records and third-party data to identify vacant addresses and reducing the nonresponse follow-up workload is our understanding and handling of undeliverable-as-addressed census mail pieces. So, in advance of the 2016 test, the Census Bureau collaborated with the United States Postal Service for the purpose of, one, understanding the postal carrier assessment and assignment of detailed reasons for why a mail piece is undeliverable as addressed; and two, for understanding the Post Office and the mail processing facility handling of those undeliverable-as-addressed census mail pieces.

The collaboration resulted in qualitative information about postal carrier determination or undeliverable-as-addressed mailings. But we viewed this engagement with the United States Postal Service along with any future engagements that we will have with them as opportunities to continue to build on the partnership that exists between the postal service and the Census Bureau. The results from the focus groups that we conducted and observations of the postal workers are being compiled and will be shared with the postal service.

So, as mentioned, we continued our implementation of administrative records and third-party data processing to reduce the nonresponse follow-up workload. We utilized administrative records and third-party data to assess whether a nonresponding case was occupied, vacant, a delete, or whether we could make no determination about that case.

Based on the workload eligible for nonresponse follow-up, the administrative records and third-party data processing (Unintelligible) and occupied determination of 10.5% and 8.9% in Los Angeles and Harris County, respectively; vacant determination of 2% in Los Angeles and 3.9% in Harris;

delete determination of 1.5% in Los Angeles and 1% in Harris; and no determination for 86.1% of the cases in LA and 86% in Harris.

A second phase of administrative records processing based on receipt of IRS data in early June resulted in the identification of an additional 308 cases as AdRec Occupied. To enable the evaluation over administrative of records and third party data processing one and five of each of the cases identified as AdRec occupied vacant or delete remain in the field data collection workload. And were subject to the full Nonresponse Follow- ups contacts strategies.

With this slide, we provide a preliminary overview of the 2016 census test field data collection workload that consisted of the Nonresponse Follow-up production work. The Nonresponse Follow-up interview work and the Non-ID response validation.

You need to copy all the data presented here in this point to one of the problems that we encountered during the test. We are working through some discrepancies that we have seen in our number. So, this table presents our best in understanding to date, however, these numbers will change as we dig deeper into the data.

The original data collection was scheduled from May 12 to June 24. But the actual data collection period was extended one week and concluded on June 30. Our target for the Nonresponse Follow-up workload as a 120,000 cases, 60,000 for site. This workload took into account self-response as well as subsampling of all nonresponding addresses to arrive our target workloads.

We then added to the approximately 120,000 cases. The one and five administrative records sample that we included for the analysis purposes that brought us to an initial NRFU workload of just over 124,000 cases. To that

workload, we've been added 4,444 cases that were selected to enable the validation of the Non-ID responses.

And in addition, as I mentioned earlier the 2016 census test provided as the first opportunity to test are NRFU process. Approximately 10% of the enumerator completed Nonresponse Follow-up cases were selected for NRFU. So, our NRFU workload was just under 7,500 cases.

We leverage the test environment to explore several operational implementation strategies specific to the operational closeout procedures and assignment of max attempt or unresolved cases the highly efficient file staff. In doing this in the Los Angeles types - sorry test - I only - we sent just over 2,900 max attempt cases back into the field.

These efforts were particularly helpful for us as we were beginning to think about how we would refine our closeout procedures and processing of the Nonresponse Follow-up workload. And I'll speak a little bit more about that a little later. So, our final workload was just under 140,000 cases.

From these cases we removed the self-responses, approximately 18,000 that we received after the start of Nonresponse Follow-up. And we also reduce the Nonresponse Follow-up workload through the administrative record and occupied removal by approximately 11,000 cases.

For the remaining workload that required the field enumeration, approximately - or over 80,000 cases were field complete, meaning that we conducted a successful interview either with the respondent or by proxy or the case was resolved as a delete or a vacant.

Over 27,000 cases reach the maximum attempt threshold that I spoke about earlier, six attempts, proxy eligible on the third without completing a successful interview. And just over 4,700 cases or what we recall to refer to us unresolved which includes things like refusals, language barriers, restricted access issues, dangerous addresses and several unique situations that we encountered such as group quarters that we were just not set up to handle as part of the '16 census test.

We have a lot more analysis under way right now that will help us gaining more complete understanding of the numbers and the resolution of the cases in the '16 test.

Now, we'll move on to successes and problems that we uncovered as we implemented the field data collection for the 2016 census test. First, I'll mention staffing ratios. As mentioned earlier, one of the test objective was to evaluate varying ratios of enumerators to supervisors.

From our observations, are they briefing and other information, in general, the local supervisors of operations, the LSOs in both test sites indicated that they do not have any difficulty managing the number of enumerators that they had assigned to them. The LSOs attributed their ability to manage their teams effectively with in part due to the system-generated alerts which identified performance issues and notified the LSOs, such that it enabled them to take quick and corrected action regarding enumerator actions or performance.

We did experience challenges regarding our training. Unfortunately, due to a number of situations such as timing and staff availability, a number of enumerators did not get trained by their supervisors as we had planned. They were trained by other supervisors.

And we did identify a number of topics where the enumerators could have benefited from additional training on specific subjects. We have to continue our progress toward using a blended approach of training through a combination of the online in classroom.

We will explore opportunities to make the online training more interactive and we'll increase the depth and the detail of training to include focuses on such things as multiunits, the handling of proxy interviews, leaving notice of visits and situations that we'll all refer to as not being on the happy path. So, those things that go beyond what the normal, sort of, easy cases are.

The key to our success will be trying to balance the training contents against the critical components of the enumerators job, while also considering both the cost and the schedule for the operation. Next, as I mentioned, one of the objectives for the 2016 census test was to test enhancements to our approach and procedures for the enumeration of nonresponding addresses within multiunit structures.

We implemented a new procedure for contacting building managers that enabled us to resolve many of the vacant addresses right away. And then for the assignment of the attempts to contact the other nonresponding addresses within the multiunit structures. While we observed this working well in many situations, we also observed that it didn't work well in all situations.

In part, we recognize that the layout and addressing of multiunits is not standard across all multiunits. What we experience with high-rise apartments and what we experience from one complex of garden-style apartment is very different from what we may experience with a different set of garden-style apartment. Also, the availability of building managers varied quite

dramatically and was not always something that was considered as sent the enumerators out to conduct the multiunit manager visits.

So, all of this makes coming up with a one-size-fits-all sort of solution a challenge, especially given the automated and optimized environment and which we are attempting to work.

So, as we move forward, we have to give thought to how we can enhance our approach and our procedures that will create some flexibility for the enumerators to assess the unique situations that may present themselves and were necessary to deviate from the optimized ordering of cases, to allow contact attempts with nonresponding addresses and an order that really suits the layout of that multiunit structure.

We also experienced some problems with our proxy interviews, enhancements are needed related to conducting proxy interviews after we've made repeated attempts to reach a household respondent unsuccessfully. In the 2016 census test, if an enumerator was able to reach a proxy, but the proxy was unable to provide the names of the residents at that nonresponding address, the interview concluded and additional attempts were necessary to finding more suitable proxy.

Enhancements to our data collection application need to be considered, such that it will enable the capture of information such as to have new status, whether it's occupied or vacant. And in the case of occupied housing units, to try to get the housing unit population count when no other information can be provided by the proxy.

Further, we want to consider whether it's feasible to collect additional household demographics to the extent that they are known from the proxy

even if the proxy does not know the names of the occupants. Next, we acknowledge that the volume of cases that went to our management review related to such things such as language barriers, dangerous addresses, access within multiunit addresses, et cetera was large. And it was a challenge to review and take appropriate actions in an effective and efficient manner.

We're analyzing what happened in the 2016 census test. And we will take steps moving forward to improve the management review process and the actions that occur as a result of that process.

Then finally, our 2014, 2015, and 2016 census tests have focused primarily on the implementation of these enhanced data collection message. They focused on the control and the management of cases. And they focused on the modified staffing and management structures. Our focus has not been on the realities of a 2020 census production environment in taking all the necessary steps to ensure high levels of completed interviews.

With the increased use of automation in the implementation of business rules that drive decisions after the maximum number of contact attempts have been made, we have seen high rates of cases that have reached that maximum contact attempt without a successful enumeration. So, prior or as throughout the - I'm sorry, prior to the '16 test, we really hadn't done much in the way of developing what I'll say is a full solution for monitoring and closing out the nonresponse follow-up operations.

We did however, as part of the '16 census test, as we were seeing the results come in, begin our efforts to think more about - and to develop our operational close-out procedures and the steps that we are going to have to take to monitor that workload in our rate of successful enumerations, such that

we're minimizing the number of cases that reach that max attempt without a successful enumeration.

Moving forward, we'll focus on defining and developing the capabilities and the procedures to monitor operational progress. We'll look at enumerator performance and use that information in an effort to ensure that we can conduct a nonresponse follow-up operation according to our schedule while still collecting high quality data.

So, in summary as we expected, we have observed, experienced and learned a tremendous amount from the '16 census test. Not everything went well, we did have problems, but we're learning from those problems. And that will inform the improvements that we need to make as we move towards the implementation of the 2018 end-to-end test.

We are continuing to make progress towards a final operational design for the 2020 census. The 2016 census test did help us as we furthered our efforts to understand how to connect with people and encourage them to self-respond. We did continue our efforts to utilize and refine our approach to using administrative records and third party data to reduce the nonresponse follow-up workload.

We did progress along our path to leverage automation across the entire 2020 census design and we continue toward the definition of the 2020 census operational design specifically for nonresponse follow-up.

Our 2016 census test experiences have identified key areas of focus for our near term which include focusing on operational and management reports, looking into the operational control system and optimizer enhancements, thinking further about our approaches to training of the field staff, working on

enhancements related to conducting interviews in the handling of multiunit addresses and working on close out procedures that consider the capabilities to monitor and take steps to minimize the number of max attempt unresolved cases.

We are hopeful with the move to the platform solution that we'll have an opportunity to develop solutions that meet our requirements and address previous challenges in support of the 2020 census. And with that I would like to turn it over to Patty.

Atri Kalluri: Thank you, MaryAnn. Good afternoon ladies and gentlemen. Patty and I are excited to be here to discuss the progress made in relation to the 2020 census systems, the decisions made on whether we want to build or buy solutions for identified CEDCaP capabilities, and our plans that ensure readiness of our systems in support of the upcoming census test and eventually the 2020 census.

We are on schedule to meet our primary goal of conducting a quality Census to count everyone once in the right place. As you heard earlier, we had successfully conducted a larger integrated test this year, the 2016 census test in LA and Houston areas.

In support of the 2016 census test, we successfully deployed and used for the first time in production CEDCaP Components of Multi-Mode Operational Control System, MOCS as it was called; and CaRDS, the Control and Response Data System. These systems communicated and transferred data between each other and other systems using the enterprise service bus, also called as the ESB.

We furnished smartphones for enumeration using the devices or service method. Both Android and iOS devices were used in the test. We saw the successful production use of IT systems in multiple geographic areas and in multiple operations as part of one test for the first time.

We learned so much from the test and importantly, we validated the concepts of multi-mode data collection and adaptive design and furthered strengthen or validated approaches (Unintelligible) optimized automation as part of the field data collection.

We saw the successful use of multiple handheld devices with different operating systems for our field work. And thanks to our engineers and architects, we saw the functioning of many disparate systems, either CEDCaP or non CEDCaP as one integrated System of Systems Including the internet self-response and real-time non ID systems.

In parallel we were conducting the analysis of alternatives for identified CEDCaP capabilities. As part of that AoA, we also conducted the COTS capability analysis and assessment, also called as the CCAA process. We followed a very well defined CCAA process and schedule.

We were guided by the Carnegie Mellon Software Engineering Institute. And as reported through our various memos, announcements and blogs, we made the decision based on the results of the CCAA process to buy an intelligent business process management suite, iBPMS as it's called and make it a part of the recommended high-breed approach of provisioning solutions for the CEDCaP program. We selected an industry leading enterprise application development platform, the Pega 7 platform of Pegasystems Incorporated based in Cambridge, Massachusetts.

At this time, I'll hand over to Patty McGuire, the program manager of CEDCaP program to go through the AoA process and the CCAA process and how the bill-versus-buy decision was made. Patty?

Patty McGuire: Thanks Atri. So since it began, CEDCaP has focused - or the Census Enterprise Data Collection and Processing Program has focused on the implementing strategies following the Census Bureau's IT guiding principles, such as reducing enterprise redundancies through shared services using a standards-based approach and using commercial off-the-shelf solutions over custom development and adapting new technology while maintaining secure systems and information.

The goal has been, since the beginning, to minimize the siloed approach and duplication that exist through the Census Bureau and taking enterprise-wide approach to data collection and processing. So, the capabilities that CEDCaP has to provide have not changed since the buy decision announced in May.

What has changed is that now, six of the capabilities per data collection and processing that CEDCaP provides are integrated into one platform called the Enterprise Censuses and Surveys Enabling platform or ECaSE. And this is where the COTS products comes in. The capability shown in purple include the centralized operational analysis and control, the survey and listing interview operational control, address listing and mapping, questionnaire design and metadata, internet and mobile data collection and the interface to the dashboard for monitoring.

Now, what I'm going to talk to about is how we got to that integrated solution for CEDCaP. About one and a half years ago, the Census Bureau began a rigorous, iterative process to determine the products to provide the enterprise capabilities for data collection and processing. While this process was led by

CEDCaP staff, a large number of stakeholders from our office, Census Bureau, were involved in the process to make the buy decision.

Staff identified the core and key capabilities and conducted market research. They looked at what was available internally and also commercially available. In some cases, within CEDCaP, in-house development teams were leveraged to develop and deliver the technical solutions supporting the defined enterprise business capabilities and the CEDCaP segment architecture.

These teams created proof of concept systems that were tested and used for the Decennial census test. The key aspects of these proof of concept systems was to allow the business area, such as the 2020 program to better understand how technology would impact their operational design and shape their requirements, business rules and their user stories for the technology. The knowledge research testing and solutions provided by these in-house teams were critical to determining an approach to transitioning to an enterprise solution for data collection and processing.

The work of these teams is the foundation on which CEDCaP built to make a well-informed recommendations and decisions on behalf of the enterprise. The internal review of results and lessons learned from the in-house solutions for prior census test was used to develop our requirements and market research parameters. Based on all our research done in late 2014 and in 2015, the Census Bureau decided to move forward with the COTS capability assessment for six of the CEDCaP capabilities, those that I mentioned on the previous slide.

As Atri mentioned, we engaged with Carnegie Mellon University Software Engineering Institute to provide guidance and to assist with the actual CCAA. First, we issued a request for quote. We then awarded the purchase orders to

the interested vendors. And then we completed further physical and engineering assessment, software testing and analysis of the vendors' tools.

Based on the results of this initial effort, we moved (unintelligible) the process of two of the interested vendors and then also included the in-house solutions. This effort then kicked off the actual CCAA in January of 2016. While this was for enterprise solutions around the same time, the CEDCaP Executive Steering Committee directed at the CCAA emphasized the 2020 census program added requirements.

Before the CCAA effort began, Census Bureau staff assessed the enterprise requirements and compared theirs to the 2020 census decennial scope system capabilities and schedules. The team identified the core capabilities and testing needs. And through these efforts we determined initial mission-critical requirements needed to support the 2020 census.

This fed into the CCAA effort which then included a proof of concept and an analysis of alternatives. So, in a few moments Atri will show you what the vendor developed during the 90-day proof of concept phase. But it's important to understand, about the focus on the decennial Census, is that the prototype that you're going to see from Atri, that was developed during this effort, was actually the starting point for the ECaSE capabilities needed for the 2017 census test.

So, I want to talk to you a little bit more about the prototype effort. The 90-day proof of concept prototype phase for the two vendors began in January of 2016. The two COTS vendors installed their systems here at the Census Bureau in our CAT lab allowing the US Census staff to review and test their products. Then the development cycle kicked off.

This consisted of five development sprints with each sprint lasting about 10 days. The vendors were given our requirements. And then there was a session where the Census staff would explain the requirements and allow the vendors to ask questions.

About 10 days later, each vendor would then call me back to the Census Bureau and conduct a demonstration for the CCAA teams to validate that they met the business requirements they committed to meeting. We were impressed with what we saw after each sprint and you'll get to - as I said, you'll get to see some of this functionality in a few moments.

But I mentioned before this proof-of-concept phase was very important to the 2020 census program because the applicable requirements, business rules and corresponding specifications from their 2016 census test were used as the foundation for the development completed by the vendors for this project. And this means that we're able to leverage this prototype developed by the vendors - the chosen vendor as the starting point for development for the 2017 census test rather than starting from scratch.

So, besides the 90-day proof of concept, the CCAA team also conducted an analysis of alternatives that included deep dives of the vendor and the in-house solutions. The AoA ran in parallel to the proof of concept phase. The evaluation included looking at five key criteria to arrive at the CCAA recommendation which included the business functional need, the system design, schedule impact, total cost of ownership and the vendor development team viability.

The analysis and recommendation from the COTS capability assessment and analysis can be found in a final report. As you mentioned, there's been a lot of

information released on it. And that report is also available. And it was made available when the decision was announced.

The CCAA findings and recommendations were presented to the CEDCaP Executive Steering Committee in May 2016. And the Executive Steering Committee confirmed the decision to use a high breed approach to deliver the CEDCaP solution. This is where the chosen COTS platform is integrated with select Census Bureau custom solutions that will ultimately address the short-term goal of successfully deploying the 2020 census while also building the infrastructure transition to the long-term CEDCaP goal to support all censuses and surveys.

Specific recommendations from the CCAA effort include that we will use the Pegasystems platform as the core CEDCaP data collection operations platform. We'll supplement the platform with Census Bureau developed optimizer and routing engine developed by the MOJO team as part of CEDCaP going to deploy the COTS platform as the CEDCaP data collection platform for the 2017 census test including the internet self-response components. But we'll also going to develop a comprehensive performing testing plan and initiate performance testing on the platform to ensure that it can meet the demands of the 2020 census.

So, I've given you an overview of the thorough CEDCaP COTS capability assessment that again has extensive stakeholder participation throughout the Census Bureau and assistance from Carnegie Mellon. The resulting buy decision is helping to reduce the risk of the 2020 census and/or other surveys or Censuses by reducing redundancies and duplication through leveraging shared services, emphasizing standards based solutions, and adapting proven technology to help deliver secure systems.

For the 2017 census test we're able to build upon the requirements, business roles and knowledge developed by the in-house team for the previous census test and leveraged the prototype developed for the CCAA as the starting point for the solution for the 2017 census test. So, Atri is now going to tell you and show you how this solution fits in to the overall solution architecture for the 2020 census.

Atri Kalluri: Thank you, Patty. So, we're calling this as Pega 7-based platform implementation here. As Patty mentioned, that is the enterprise Censuses and surveys enabling platform, ECaSE platform. The development effort using the ECaSE platform stood up as a project in the Decennial Information Technology Division -- that's my division under the CEDCaP program.

I'm happy to report that (Nicole Lee) who may not be here will lead the development effort. (Nicole Lee) is the perfect leader for this effort with multiple years of software development and programming, project management experience especially in developing systems in support of large-scale field operations. It's important to note that the ECaSE platform-based application development effort is not starting from scratch.

As part of the COTS capability and analysis and assessment as mentioned by Patty, CEDCaP components we were assessing. We strategically identified and delivered a subset of key 2020 decennial census capabilities and requirements to the vendors to build and configure related solutions on their respective platforms. That effort of 90 days consisting of six sprints resulted in the creation of applications or questionnaire design and metadata, internet data collection, centralized operation analysis and control, mobile data collection and survey operational control and address listing and mapping.

We identified and delivered a subset of the key 2020 Decennial Census Capabilities and requirements to the vendors. So, now that we made the decision, we are delivering the whole set of needed capabilities and requirements. We have stood up ECaSE platform based application development for five areas: content metadata, all control systems, internet self-response, enumeration and listing and mapping.

We have product donors for each of these work streams and the teams have embedded developers, testers, subject matter experts, and security engineers in them. We adopted Agile Scrum methodology which allows us to identify and prioritize needed application functionality. It also allows us to take a comprehensive look at what is needed by then and create the tail user stories for each of the work streams in parallel.

So, here is the 2020 Census system architecture diagram updated base on the decision to take a hybrid approach to the solutions. It now depicts a System of Systems that integrates the Census Enterprise Data Collection and Processing Components that are custom solutions.

You see the blue-colored had a background in boxes or ECaSE platform based solutions. Red colored header background in boxes, it shows how the enterprise solutions will work for decennial Census. It also shows how the legacy systems are integrated and how all of the systems interact and work as one system of systems.

Displayed on the screen is the second part of the 2020 census solution architecture diagram showing the updated decennial administrative support for this section. I want to reiterate that this architecture is flexible and extensible. And because we made the key buy-versus-build decision and incorporated the changes needed to the architecture based on that decision. Unless there's a

change to the operational plan, the solution architecture can be considered baselined.

We have already gone over the architecture in detail in many of our previous sessions. So, at this time I'll highlight the changes due to the incorporation of the ECaSE platform. And as I do that, we'll also show you the recorded videos of the functionality developed as part of the CCAA process and delivered by the platform for the 11 capabilities needed.

Let us start with the modified legend and survey design and instrument development. So, the red-colored header background in boxes were ECaSE platform based applications can be seen. I'll also know the color scheme -- blue arrows for platform based internal interface.

While we are depicting such an interface, it's all really within the platform which means that the applications on the platform will all be integrated if you consider the interfaces, the data flow, interoperability, data consistency, etc. And these are delivered as an integrated product. Of course we need to build test and implement scalable interfaces between the platform and/or non-platform dependent systems.

The box on the right shows the use of ECaSE platform for survey center's design and instrument development, previously shown as the functionality of COMET, the Content Metadata System and in-house system. Here you see the field and internet components of the architecture.

For field data collection, there is the need for listing and mapping as part of in-office canvassing and update enumerator operations and enumeration as part of nonresponse follow-up and, again, update enumerator operations.

Applications on the ECaSE platform will replace what used to be the solutions of LiMA, the listing and mapping instrument, COMPASS, MOJO/Mobile Case Management. However, a critical component of MOJO, the optimizer, will be reused as a dependent service by the platform as part of its management of field assignments and routing. The internet component of the architecture shows the use of ECaSE platform for internet self-response capability.

The solution will have non ID component integrated and will allow respondents to use any standard web browser to respond. It is to be noted that as part of a telephone questionnaire assistance, the respondent may prefer to take assistance of the CQA which they have to respond, in which case, the staff will use the ECaSE platform internal based response capture system which should be a slightly modified version of the internet self-response system.

At this time, let's watch the video of the listing and mapping capability, followed by the internet self-response capability developed on the platform as part of the CCAA process. The videos will allow you to see a flavor of the capabilities that the platform provides.

For address listing, you'll see in the video an application that was built on the platform and run-on and IOS device -- an Apple device. Note that the case assignments for address listing would have been made by the operational control system when working the route optimizer just like it would do for the assigning enumeration cases.

We'll see how cases have worked in a chosen census block. One of the cases will be in an address verification case. Second one will be adding off a new housing unit, and the third one will be a delete. You'll see the use of a map,

identifying the housing unit location on a map, functionality of the application in an offline mode and the triggering of the sync process in an online mode that ensures case information is conveyed accurately to the operation control system.

So, let's go to the video please. Okay, let's start to play. You open the app on the IOS device we pick a census block 2006, you see the distance warning?

The listed is not where the housing unit is. You navigated the map. Blue icon is for the central of the block, yellow icons for unit detail. The address information by clicking on it.

And the system can be made to go offline. Then we turn off, going to the airplane mode, though it's offline now. We look at the address details and the unit details.

And then we adjust the map spot by dragging the location. And we click verify as a verification done in the field. We'll now see the adding of a new unit.

This is the fact that it was done. And now we take on the next case. It's the details of the added unit. A standard interface that allows you to enter the information about that unit.

Then you click on save which allows you to save the information. The next one we'll see the deletion of the unit.

Still in the add situation. Picked the next unit, who are green as we can see them. That means that the cases are done. This is the third one.

Why did it - so delete. You see it in red. Those are the three cases.

And we switch back to the airplane mode and if you concentrate on the top right corner of that application, you see that they sync well - actually take place when the airplane mode is turned off. This is all complete.

The next we do will be on the internet self-response application. You'll see the application running in the browser with an interface to accept responses on the internet from a housing unit identified by a census ID. The video will show language switching capability from validation and entering of responses including demographic information. Let's start the video.

So, the web page for self-response - and you click on the non-ID portion. It just showing you that, that was an integration of the non-ID aspect of it as well.

But here we enter the ID. This is the information based on the ID entered.

There's online help available from each screen if you click on the help button. And we can even switch languages on the fly.

Let's switch back to English. And then the response information is entered. Name, (Tom Jones), and then there are soft edits as you go forward. The telephone number obviously has to have correct number of characters.

Then this are the typical questions for this - that you will see in the survey. Number of people. Choose to add people, (Stacey Jones) is added.

Did they rent a house or a - owned? The summary, and the introduction of demographic information - name, date of birth - again, edit embedded in the

system to make sure that when it entered agrees with the date that I have entered.

This category is under come and (unintelligible) background and the results.

Relationship question obviously is there. Date of birth for (Stacey), categories that describes (Stacey Jones). That's the result of the survey. We logged off, but then we also show that the same p tries to log in again, we captured the fact that they already gave us the was information. Basic edits...

All right. So, next slide please. Okay. Yes. So, the clipping on the left - on the screen, shows the consolidated operational control system section of the architecture assigned to case platform. If he calls, we had MOCS, the Multi-Mode Operational Control System. And functionality and module for field operational control.

The ECaSE platform will take all the responsibility of operational control from these two systems. The platform comes would reached reporting capability and obviously we want to take advantage of that capability. Oh, you see the architecture diagram of the date of the calling.

The section on the right is showing the updated interaction between the Operational Control System on the platform with C-SHaRPS -- Center Schedule A Human Resources Recruiting Payroll System. And the OCS responsibility to deliver the workload of cases for printing either commercial printing or printing at the National Processing Center in Jeffersonville, Indiana.

Now, let us watch the videos of the OCS functionality. Followed by the enumeration capability developed on the platform as part of the CCAA process.

OCS functionality in the video will show a true leader logging in assessing the stated operations. Look for alerts, view evaluable it and enumerate this. And the first notification to the enumerators. The view cases on the map and update a case based on - at the good alert. So, let's go to the video.

So, log in as the crew leader. There's a big dashboard where we could study the operational reports. A lots notifications and outstanding cases if the reach interface aspect to what the platform provides out of the box.

And we could also see the team evaluability. One of the schedules that they have and by each enumerator as well. And you can locate the housing units of the enumerators on the map as well. As you see, there's a map interface.

Next we'll show a push notification requesting Saturday coverage. This is where the crew leader is entering information and anyone who worked the weekend. And it obviously sends out an alert to all of the enumerators.

If we could look at the status of an operation, let's see the cases. And we could even locate them and review them on the map and zoom in to get more further details as to where these cases are. And you click on the icons and you'll see the information about those cases.

We go back to the dashboard and there is an alert that comes in this example about the language used. One of the enumerators that is visiting a case would require proficiency in Vietnamese, I believe.

Yes. So, we update the case having known that there's a requirement now and that reflects in the dashboard.

So, just give you a flavor of how things are on the platform. The next video is - of the use of an enumeration instrument. You'll see the running of application on an Android phone. The rear one was on the Apple phone. And enumerator and you'll open an assign case, check the address and the fundamental details. And conduct an interview in an offline mode. You'll also see how when connectivity is established. The auto sync process is regard. So, let's go to the video please.

They open the app. The menu shown, they open the case for 600 just to have a new. They can click on the address and appointment details. And again, you'll see from the actual location notice.

This is a personal visit. We're attempting a census address. They may be too far from (600 West Avenue). Continue the interview? Yes.

So, there are in-built checks there. And we go to the airplane mode. Wi-If's disconnected and they can switch languages here. This is Spanish on the fly available again. Let's switch back to English.

And then - the respondent available? The answer is yes. Again, help is available online. Here is the actual collection of information. So there are soft edits again if the correct information of - so the anticipated length is not entered, system warns you. These are based on the guard rails that the platform provides.

Then this is the demographic information of (Tom Jones). And here we show that they refused to give the date of birth. And the system prompts the

enumerator that's a work that needs to be done so that giving an age - but he also can add a note.

And here, you could speak into the system and it'll capture the text for you. It comes with this types of capabilities.

And then we'll go to the next series of questions. So, the categories for (Tom Jones). These are the specific categories as we call him -- typical questions that the survey or the census contains.

So, all information on (Tom Jones) - information about when it will be convenient for them if somebody else from the Census Bureau visits them.

So, we submit it, we capture the contact history, and then we switch off the airplane mode and you'll see that the items actually sync with the status update on the top right corner at synced. We complete the case and we log out.

So, let's go back to the presentation. Next slide, please.

So, we started with this slide if you recall. And let's end with it by going all the systems readiness for the upcoming census test. For this canvassing test, we're going to use the Listing and Mapping System -- LiMA -- that's already in production use. But we have integrated LiMA with module for full operational control and MCM -- Mobile Case Management.

These are the critical CEDCap systems that will support the address canvassing test. All of the non-CEDCap systems: the MAF TIGER system, the BARCA systems, the block assessment research and classification applications, MOCS which is the sampling, matching, the viewing and coding system, UTS,

Learning Management System, etcetera, are all on track and had gone to track testing before deployment to production.

And as you may know, the in-office of this canvassing is already under way with all systems functioning as expected. We have successfully deployed the active block resolution system to production as well in support of the in-office HS canvassing.

The 2017 Census Test is when we'll deploy the ECaSE platform for the first time. And we are actively working with the vendor to ensure that the platform base applications are released on time to be part of track test.

It also working to integrate the non-platform systems with the platform. We've been testing in the Cloud -- multiple Cloud environments -- including standing up data bases, middleware, do a special service, application plus these art of scaling, redundancy for the entire solution with failure capability to multiple geographic locations, security control and monitoring ensuring secure communication between the Cloud and all data center.

In parallel, we are working with the ECaSE render to finalize the optimal solution for the 2017 Census Test as well as for the '18 and to end Census Test. Meanwhile, we have - throughout the ECaSE platform and servers and out datacenter and obtain the totality to test.

As you may know, we're also conducting the Puerto Rico Census Test. And so, in preparing for that test we successfully updated the master address file with Puerto Rico addresses from USPS. And have updated all systems to be able to match Puerto Rico addresses just as we do for addresses on the states side.

It also getting ready to update the master of this file with address information from the local municipal. For the 18 and to end Census Test, we'll -- of course -- using a case platform and the applications based on the case platform will have the technique integrator support for 18 and to end Census Test to work with us to not only integrate the software components, but also systems and hardware either in the Cloud or in our datacenter with scale ability being one of the key aspects of the integration and to provision infrastructure as necessary.

Our anticipated date for the technical integrate to be available to assist us is later this month as Shirin pointed out. that completes the presentation. Thank you.

Woman: Sunshine is our discussant. And since we started about ten minutes late, I think we'll push back everything except the adjournment by ten minutes. So, Sunshine, you actually have a chance to talk.

Sunshine Hillygus: I'm terrific. I will take that opportunity. And do I just say proceed to the next slide and some may do that for me or is there a quicker. Thanks.

So, first before I can have jumped into some comments, I should say that getting the slide back was entirely overwhelming. It was so useful to see and to give some contacts to what was actually in this slide.

And so, my initial thought when I started flipping through - is number 1 is that there is a lot going on - that there's a grammar that's trying to get covered. And there are a lot of acronyms and buzz words like enterprise, capabilities and system integration, operational analyses are not part of my world.

So, often - and so, I really - and hoping that I can comments on things that I can be useful but the other members of the committee will jump in on other - on aspects of the 2020 - that are also critical. So, a few other things just come to mind that I will problem tag at the end of my preferable marks.

So, you know the first thing that we see is - the 2020 is up on us. And in looking at this very comprehensive overview of what still to be done - were a things have been already, you know, there is still always some certainty in where we can be abuse as a committee, you know, I really appreciated that the one slide showed where plans are essentially set and where plans still to be made because that at least give some sense of where we can wait.

But from, you know, everyone here, you know, it would be useful in thinking about how CSAC could be helpful to the 2020 particularly with respect - in my mind, one of the things that we had propose previously with the working group on PRIMUS that is perhaps no longer - necessary, however, are there ways that some of us should be involved on a more detailed level because it certainly what we got here was kind of the bird's eye view.

So, okay, just a few comments that came to mind is "I went to the slide" that you know everyone here knows that when we saw the primary demonstration last time that we raise a number of issue.

And so, on the one hand, you know, it came, you know, quite of relief to see that the plan moving forward is to work with an outside vendor. On the other hand, many of the very detailed questions and concerns that we raise with respect to PRIMUS. We just don't know how are we going to play out with, you know, an office shelf more or modified system.

And so, from my stand point where I feel like I can be most useful is it we had a little bit better sense of exactly the testing plans and timeline, you know, so from my reading of the process like, you know, the explanation to us about, you know, why PRIMUS frankly, you know, how to very - not very user friendly interface that it was about, you know, focusing on the architecture and the limited amount of time and there was available to develop it. But my read was that there was like ten days. That was the sprint time in producing the prototype of the new version.

And so, my concern is just in this, you know, transition or some of the issue raise and empathy during the use ability issues. Are they, you know, are we on track to be able to address those - and, you know, is there time to learn from the PRIMUS experience and end test in 2016 in form 2017 so that 2018 can be, you know, a real role out because thinking to, you know, some of the challenges with respect to none response follow up.

You know, my big question is if we can solve in the front in by increasing, you know, self-response, then that, you know, makes everything easier and hopefully some of the issues that were encountered actually aren't going to be an issues once the self-response is resolved.

And so, you know, the question about this new platform are largely that the same question for me that were posted in the recommendation and the question we post regarding PRIMUS.

What are the specific testing plans in particular regarding user experience? How is it be going to be research who at this interfere is going to be involved in that? I want to make sure, you know, what I'm hoping this is not just kind of architecture people but it's also those individual who are familiar with, you know, survey message and user experience.

And just one of the things I noticed just from, you know, that quick demonstration as over here, you know, come in - I hope that anybody that your hiring to do nonresponse follow-up is under the age of 20 because I think they're the only ones who have the dexterity in their thumbs to be able to actually go through, you know, and use those radio buttons when they're so close together.

What kind of web designs, best practices - would say you actually want to use radio buttons when you're dealing with an iPhone, that you want one of the bigger buttons and that's easier to people to use?

And so, I just want to make sure that there's a process and place. I'm sure there is but it will just be great to hear more about it. Likewise, you know, even if PRIMUS is no longer going to be in the picture, surely there are some lessons learned that we'll be useful with a new platform.

I also wonder - not only in terms of that internet - interface that more genuinely in each stage in the Census Test are there going to be some opportunities to now go back and do some focus groups to figure out - not just those who are doing the data collection but on the citizen side of things, you know, to figure out if there are lessons that could be learn from the 2016 Test, that might inform 2017 and 2018.

And I have said it before and I just - I have to take the opportunity to say it again. Remember that for most Americans, this online interface is their one way that (unintelligible) this Bureau. And so, you know, in my mind getting that right is just so critically important. Getting it right is going to influence the success of the Nonresponse Follow-up. It's going to - it's just such a

critical thing to think about, you know, how the most people will experience the 2020 Census.

And I hope you will involve this - in the process. So, you know, I can't speak for Barbara for another members of the committee, but, you know, - okay. So, you know, we had offered to do a working group for PRIMUS but, you know, my offer still stands that we'd be happy to be involved in a working group to help with, you know, the new systems.

The other thing that I would like to touch on a bit in part just because of my own background and where I can contribute -- I can contribute little in terms of talking about, you know, system integration and architecture. But thinking about the partnership and outreach just a few questions came to mind.

Barbara and I had been involved in looking at some of the messaging work with respect to the ACS. And there was a terrific report and assessment by Don Dillman. I wondered to what extent the research and testing around ACS messaging was being used by the 2020 Census.

And so, this is also something I said before, but remember 2020 is going to be a presidential election year. I would be curious to know if there were any implications in the 2016 Test, say in Texas in terms of, you know antigovernment attitudes and other things with respect to the Nonresponse Follow-up and being prepared for the potential - for the census to be politicized.

I remember one of the ACS messages was a letter that said, "My fellow American." That always addressed as - and I was just thinking like, "Ah, you know, we can't have the letter, you know, have that type of language

particularly in election year - particularly when we're trying to get, you know, non-citizens to respond.”

And I just wondered to what extent the research and feedback that has been given with respect to messaging is being carried over into the 2020 planning and researching. I'm also curious about what type of messaging research is been planned? And once again, any of the details that the CSAC could find out about, we would be - I think more than happy to offer some feedback if it's useful.

I know that in some of the partnership and communication material we've seen in the past there has been, you know, some discussion about micro targeting that could be defined in a variety of different ways.

And so, I would just urge some caution. Not only is it the case that, you know, that the micro targeting means different things but also that there's a lot of people - the very quickly micro targeting can go over into the creepy zone where people feel like there's a little bit too much that is known about them.

And so, a lot of people don't like tailored messaging. And so, again, the question just becomes to what extent and how is the message going to be micro target done in 2020. I guess it isn't - excuse the typo bit but also, it maybe is not necessarily about partnership and outreach beyond the implication I think is that with self-response and increased outreach, there is a potential for increased over-count.

And so, you know, one of the questions that I have was just back to the 2016 Test is looking at those duplicates and how that is been handled both on the automated system but also, you know, just - once every comes in.

With respect to administrative records it, you know, we have seen and had a lot of discussion about the use of administrative records. I guess again, kind of key question is how we can help with the decision, you know, said that the least of information to be use - will be finalize in 2018 but I suspected it that be experience with 2016 that there were some lessons learned. And so I would just be curious to hear and where thing stand in terms of finalizing the least of the administrate records and third party data that will be used and how it will be used.

Again, thinking of the context, you know, to what extent will the use of administrate record, you know, how will that be viewed by the public and are different sources of administrative records going to, you know, create different levels of discomfort with the public?

There's also the question of just record quality and to what extent the 2016 test is able to kind of sort out, you know, which administrator records are - were they abused and that - by the source, or how hold the record is or who it is your trying to fill in for?

There is also ultimately the question of, you know, if we have a really terrific records about people who pay their taxes and have been in their home for very long time? You know, and are older and to have, you know, a lot more data about them, you know, is it possible that at the end of the day, the use of the administrative record ends up exacerbating the differential under-count because those people who are hard to count are also the hardest to find in administrate of record.

And so, I again, presumed the 2015 certainly 2018 will be looking into that, but - I mean these are all questions that in some sense had been discussed

before. The question is just where are now relative to where we were in previous meetings.

Just a couple of thoughts about some of that staff in our field operations - issues discuss. There was a time -- and I couldn't remember the context of which -- data collection of the census bureau we such some material about how doing and testing after training was core related with performance in the COL, but sorry I completely forget, you know, again which census product that was related to.

But I just wondered to what extent on that type of training testing was going to be used in 2020 because from that initial staff that we saw it looks like it could be super helpful. I wasn't sure if you guys are settled yet on the paradata that the interviewers will collect during Nonresponse Follow-up.

And then the other staffing issue wondered about was just, you know, how staffing need are changing given the use of online self-response and this option of being able to complete your census form over the phone if you call with the question. And so, has there been an assessment of what exactly that is going to require in terms of staffing needs?

A few thoughts that didn't really fit under any category but, you know, we had mentioned that in our last set of comments. We wanted about the potential for members of CSAC or researchers through the datacenters to be able to work with some of this - the test data. And so, wasn't sure if there was an update on that.

I wasn't sure the plans to do things with the Standing Rock Tribe, if there are implications for what's going on now in terms of whether that remains the right test case. And then I wonder if they were update final work towards -

from the national content test. From the presentation - I think there was all of my thoughts from the presentation.

The other kind of final points that maybe came to mind is thinking about the North Carolina case, again, 2020 present election year. The potential for the census to be politicized I think is real. And what I realize is that, you know, you actually have a terrific test case with Buncombe County, North Carolina.

This is a country that crosses two congressional districts, one that, you know, has the Ashville and, you know, has a lot Democratic support; another that is very rural and has liberal Republican support, and being able to start the process of getting an advocates on both sides of the aisle for even - that the test I think could be super useful for thinking ahead about how that might play out in 2020.

Woman: Thank you. I think we're going to have actually go on to the point - what? Tommy said we can take five minutes and push that back by 15 minutes. So, Dan?

Daniel Atkins: Daniel Atkins. I'll just run through quickly some things I want to observe. First, actually appreciated your comments about succession planning and how systematically you're doing that. I've been involved with other federal agencies which don't come close to that.

So, on the internet response from the 2016, this notion of partners, did that include libraries community of centers and so forth where you could go and get access to machine and expertise to help and - do you think may that - might be even be beefed up and helped with the response in the future?

I just really want to underscore what Sunshine said about the importance of the user experience testing. And please be sure to allow plenty of time for that for all involve the general public - the enumerators, the supervisors and so forth.

And I also want to say just back up and say that you're seeing all of these a lot more tangible -- it really felt good. I think, you know, Jack and I had been kind of following this for long time in a very abstract way.

And get some UX people involve in particular those that have experience with survey work in the UX world and maybe get some independent people that are made our census or know your vendors.

You're, I guess, leasing or borrowing these devices from some contractor. You showed an iPhone and an Android version. Are you intending to let the enumerator pick what they want or both? So, you really - you going to standardize from one or the other?

Atri Kalluri: We'll standardize on one or the other.

Daniel Atkins: Okay. Have you done that?

Atri Kalluri: Not yet, after the (unintelligible).

Daniel Atkins: I guess you ought to do that fairly soon and stop building for the other.

Sunshine Hillygus: Yes. That would be part of the RFP. And we're going to award based on best GUI.

Daniel Atkins: Yes. You know, in the interest of time, I don't think we have time to get response to this, but I'm sure at some point I'd like to hear a bit more about how are you going to ensure communication, collaboration, timeline harmonization between your various vendors and that census staff that now, you know, they're going to be people resident here.

I mean, all those kind of things are really, really important and I'd like to hear more about it. And then finally, when and how will the scalability and peak load testing occur. I hope not too far. I assume you'll have some way of generating phantom data or something to do that short of the real day. But I really enjoyed the presentation.

Barbara Anderson: Bob?

Robert Hummer: Thank you. I had a question about the Houston or Paris County and Los Angeles County results that were shown. They're quite a bit different results in those two places and specifically for the inter-choice option, you went into that a very little bit about that that may work especially well among certain population subgroups and less well among others, maybe the elderly and so forth. But what specifically did you learn based on those results and more generally what other kinds of lessons did you really learn from those two sides because those results are quite different? We didn't hear much about that.

Barbara Anderson: Noel and then Andrew.

Noel Cressie: Noel Cressie. I just participated in the 2016 Australian Census as citizen and it did not go well for the Australian Bureau of Statistics. On the evening of the census, when people should have been logging on to submit their forms online, there was a huge cyberattack and an attempt to deny service, which was successful and the site was closed down for about 48 hours. It took a lot

of energy out of the census, a lot of -- and it was also combined with fears about confidentiality and privacy. I'd like to know what the Census Bureau is doing about the possibility of cyberattacks, getting in on the web where these devices are trying to access as well. Thank you.

Barbara Anderson: Andrew?

Andrew Samwick: Andrew Sandburg. I want to underscore Bob's point about just these dramatic differences between those two test sites, as well as the internet choice option. I think we would all appreciate a very detailed response to that disparity. It's one of the things where there's a fair bit of expertise around the table.

On a similar point, I continue to be surprised that administrative data doesn't help you with more determinations about non-response follow-up. And in particular, I keep thinking that utility companies have to be a great source of information, right. And with the census, as opposed to other surveys, that you ask for a name is kind of incidental. You need the count, right, and other demographics, name is not even all that important. If there is utility service to that facility then you know you have to follow-up. If there's no utility service to that facility, you have pretty good information that you don't need to follow-up.

So I think we would all really like to sink our teeth into your best ideas about what administrative data you're going to use. Some of them involve privacy considerations. Others of them, like this one, where they're not associated with the name, much less of a privacy consideration. And so it's just another plea for more detailed information about what you're thinking there so we can have some positive input.

Barbara Anderson: We're now 20 minutes behind and what I think we're going to have to do, all these very good questions and comments, stick them in your comments and we stick them in our official comments and recommendations. And all the presentations have been very interesting and we've virtually every time run out of time to -- for the census people to answer our questions. I mean it's nobody's fault. This stuff is fascinating. But I think what we all need to do for things where -- that weren't answered, partly because there wasn't time, we need to put all these things in our comments and recommendations. But I think that we're 20 minutes behind, which is -- this is life. We're doing better than we do sometimes and I think that this means that the next session is pushed back until 2:35. So you're on break.

Some people asked me where did this session come from and I will remind you that at the last CCAC meeting, the committee -- several committee members were expressing a desire for consideration of potentially useful new or improved or restored or whatever census data products. So the Census Bureau being very responsive, put this session on the agenda. So as they say put it -- send in your cards and letters, but actually say whatever you're thinking about. And some people have already made some comments to me about census products they think there should be and how improved, restored, whatever. But Irma, I know you had ideas.

Irma Elo: Well, this is something that come up among the members of the Committee on Population Statistics of the Population Association of America. And that's really -- I mean I think everybody applauds the focus on generating small area data and on the American Fact Finder. But what has also happened is that the sort of historical tables that have been there for a long time have disappeared at the higher level of geography, including some detailed tables by race and Hispanic origin for higher levels of geography rather than sort of small area

data. So here, I'm talking more about like metropolitan area data, national data, so that the historical series is sort of cut.

And so there may be some possibilities, or if there would be some possibilities of generating or reinstituting some of those tables I think would be of value to the community at large. And those would be based on the ACS and probably pull data across some years. So I don't know. There are some people from ACS here so maybe they can comment.

Barbara Anderson: Other thoughts? Well, I have a thought which I've said in other settings is that it would be nice if the group quarter's data from the ACS were published in somewhat more detail and with somewhat more detail by type. People say people don't use the group quarter's data very much, but maybe because it's not presented in a very useful form. And I've said this in a zillion settings, so I thought I'd say it again here. Other people with ideas about possible new improved restored products, you like you want to say something Noel. This could be a very short session if we don't have more ideas, folks.

Andrew?

Andrew Samwick: I may have made this suggestion just at some point in a prior meeting, but I think that the census should take the lead with educational services on how to use their data. I have in mind a (MOOC) or just online courses for people how want to learn how to use census data. And enrolments on these (MOOKs) are in the tens of thousands when you're combining interesting topics with useful skills. And so I'm sure you could find a couple of university partners on any of the main platforms to do that.

Barbara Anderson: (Tori), who now is head of the ACS looked like she was dying to say something in response to Irma and me, and I think that would be appropriate since (Tori) knows everything about ACS now, right?

(Tori Valcoff): Yes, Barbara. As your student, I have learned everything about (ACS) in the short amount of time I've been there.

Barbara Anderson: And everything else.

(Tori Valcoff): And everything else. No, I just wanted to -- (Jeff) will respond to Irma's but to yours, Barbara, we are looking at redesigning our (GQ) products based on the comments that we got from the (unintelligible). So just we're working on it.

Barbara Anderson: It helps to complain.

Jeffrey Sisson: So hi. I'm Jeffrey Sisson. I'm the assistant division chief over data products creation in ACS. Relative to some of the historical tables, we are currently looking at that. Our initial thoughts -- and it's nothing more than thoughts at this point -- is we would produce something on maybe a five year basis. So every five years, we would produce something like the historical tables we had for (unintelligible) at some of those higher levels of geography. And we would most likely do it kind of out of our normal production cycle so that we have the bandwidth and resources to do that. But that's what we're considering.

Barbara Anderson: Other thoughts? Yes?

Ken Simonson: Ken Simonson and this also is something I've raised before but with respect to the construction spending data, I know that the Bureau is in the process of looking for a -- or renewing a contract to provide construction starts

information. And I think it would be great if, when they get going on that, that they have some new categories of structure types or look into whether they can provide some national detail on the construction information. And specific structure types would include datacenters, for one thing, and some more detailed breakdowns of some of the existing categories. I'd be glad to discuss those in more detail if anybody in the Bureau wants to get into it.

Barbara Anderson: Willie?

Guillermina Jasso: In several data products and particularly in the ACS, which is the one that I looked at, in breakdowns, by nativity, they often have also breakdowns by period of entry. And it used to be the case that the residual category for period of entry was before 1980. I think two years ago, this was changed to before 1990. Now, that makes perfect sense from the standpoint of table layout, but there's an important reason why it might be good to preserve the before 1980 category. And that reason is that many immigration researchers believe, including government immigration researchers, that people that enter the U.S., came to stay before 1980 are likely to all be legal. And thus, they would have very different behavioral characteristics and possibly coverage outcomes than those who came after 1980.

So it would be really good to look into the possibility of restoring the before 1980 category to all tables which have breakdowns by period of entry for foreign born.

Barbara Anderson: Well, I have something, which relates to what Sunshine was saying today and various people have said at different times. And this relates to the possibility of census survey data that were collected for looking at various kinds of methodological issues that could be released because of confidentiality in their raw form, whether in some form these data could be --

many which I think now are not available to researchers at all outside of the census bureau -- if they could be released through the census datacenters, which have all these confidentiality protections. I think a lot of these researchers would like to look at and it's possible the Census Bureau could actually learn something useful from having outside researchers look at them.

There's the thing Sunshine talked about but I think there have been several other things like this at previous meetings there, people have made comments along those lines, often with respect to the various methodological tests leading up to the 2020 census. Bob?

Robert Hummer: I just want to quickly third that. I sent a note along those lines through the system here. Couldn't agree more in terms of the test data for 2016. I think the academic and private sector world could do a lot to analyze those data and understand some of the things that we see in there, but I'm not sure we'll have the chance to do that.

Barbara Anderson: Well, I think that I'm really glad we're going to have a briefing on the national content test, but I imagine that the same comment could be made about the results of that. I was wondering, Allison, as a user of local things for policy planning, you usually have good ideas. I mean that in a positive way. I don't mean you sometimes don't. I'm going to dig myself in a hole here. But you know what -- I'm tired. I've been up since 5:00 this morning. You know what I mean. But I was wondering from your perspective as a user and an analyst, if you had any ideas about census data products that would be helpful to you and other people in your kind of situation?

Allison Plyer: I have actually been sitting here racking my brain for what we were talking about around the office in, like, 2011 and 2012 when the data was coming out. And we were going, gosh, I wish they'd broken it down this way. And I can't

think of what that was. So I will, if I remember, I will bring that idea to the next meeting. I think I just have to leave it at that because I can't think of what we were talking about then.

Barbara Anderson: If you think of it during dinner or anything, send me an email about it.

Allison Plyer: I will.

Barbara Anderson: Other thoughts? What about -- Ken had a useful suggestion, but are there other people from the business community who -- of which we have some very good representatives here -- they're all great -- where there are census products that would be helpful for you and your -- the businesses that you're involved in? Krishna?

Krishna Rao: So this less of a specific product but a bunch of the sessions today talked about sort of ongoing efforts to sort of harmonize questions or ways they're thinking about things across agencies. And that's a problem I think all (unintelligible) data also have, how do we combine this answer with that answer and this form to answer the question. So as that sort of work is ongoing, the degree to which it's public and sort of shared, I think, solves a lot of problems that folks have out in the world of trying to reconcile things across different data sources that all produce, but sort of the U.S. government umbrella.

Barbara Anderson: I think that's a great idea and it doesn't sound like it would violate anything as far as I know. Ken?

Ken Simonson: Along those lines, I know the census folks are aware that BLS is attempting to measure a portion of the so-called gig economy and specifically looking at employment situations. And obviously, this ties in very closely to what we

need in terms of measuring the value of that economic activity, which I think falls more in census purview than BLS. So perhaps we could get a presentation next time about what the census is doing to try to capture that so-called sharing, or gig economy, or other initiatives along the lines of how the economy is evolving away from the traditional sources that you've been measuring, and also how you're integrating with BLS, or BEA, or other agencies to get that information and fulfil their needs to the extent that you can be the resource for them.

Barbara Anderson: We had really fascinating presentations today on the economic programs and on big data. And I'm sure these people are formulating really great comments. But for those of you who are experts in this area and also stimulated and/or stimulated by today's presentation, are there any census data products thoughts or suggestions in the area of economic programs or big data? Well, that didn't go anywhere. Ken?

Ken Simonson: Well, let me turn it in the other direction that we know that budget problems are very critical and they're not going to go away no matter who gets elected in November. So would it be useful to the census to use this committee as a sounding board for possible ways of cutting back on what you've historically done rather than -- or just coming up with more ideas for money to spend that you don't have? I know it's always sensitive to raise something that you would cut just as nobody goes to congress asking for a tax break is willing to offer a tax increase.

Barbara Anderson: Noel?

Noel Cressie: Noel Cressie. Barbara's looked at me twice. I feel like I have to say something.

He usually has something useful to say.

Noel Cressie: So just to bounce off a little some ideas on big data, I've been looking at some initiatives in something called citizen science and this speaks to the idea of web scraping or looking at opportunistic datasets. And so in science, what citizen science is where people volunteer to do things, like go out on surveys. They might be untrained or they might be very sort of lightly trained. But the data coming back doesn't have a very high quality associated with them. And I'm thinking, you know, as we're listening to (Carla)'s presentation in the big data area, I was thinking about citizen science and some work that I've been looking at reviewing, actually for journals and really coming to the conclusion that it's very hard to do citizen science, and it would be very hard to do web scraping without a solid foundation upon which to build.

And so there's this notion of a survey -- a source of data coming from a well-designed survey, plus a source of data coming from a non-designed (unintelligible) in science or web scraping source. And the two together can work extremely well, but the web scraping without the solid foundation would be built on a house of sand. And I'm just starting to see those sorts of things coming up in different areas as well. And in my -- in some comments that I'll send you, Barbara, we'll huddle about in the big data area. I've seen a paper by a guy called Michael Elliott at Michigan State where he does a nice piece of work and essentially uses found data or web scrape data and gives them pseudo weights. And those pseudo weights to be comparable to these and does some simulation, and makes it clear that if in fact there is no survey around which to build your essentially found data or web scrape data, these pseudo weights mean nothing. They don't go anywhere. They don't give you very much accuracy.

And so this notion of being able to simply go out and grab data sources, I mean we all know and we're all sitting here, and we're trained in a sense to be pessimistic, right. We're academics but I think we shouldn't get carried away with the idea of, you know, free data out there, just rock up and grab some, that we have to work hard to make sure that the high quality data that we get out of surveys, we might spend less money on that and spend a bit of money on web scraping. But the two go together and that's what I'm seeing. I'm just seeing this general trend coming from different areas where people are trying to do things in the citizen science area without a well-defined science survey and it doesn't work. It does not work.

Barbara Anderson: (Jeff)?

Jeff Lower: Hi, Jeff Lower:. This is just a comment related to big data and the thing that caught my attention the most, and that was the path of collection and integration with things like accounting software and coming from a business owner, that kind of scares me. I come from a very competitive environment. I know business analysts, their entire job is to reverse engineer other company's rates, and burdens, and salary thresholds to determine what that competitive advantage would be.

So I think there's a lot of risks there in terms of the protection of that data and the privacy of that data, and whether or not the census really wants that to get their hands in that. Because that's a really extremely competitive environment and especially as it relates to accounting data.

Barbara Anderson: Since we're allowed to actually have a conversation, I want to make a comment on Ken's comment, which is although I think it was useful, it's also my perception that in terms of making clear the value of the work the Census Bureau does, that some of the things that are sort of most compelling to

people of all political persuasions is the usefulness of Census Bureau products to people in the business community. So I think it is more worth our while in that area than in some other areas to think about what kinds of products people in the business community would actually find attractive and helpful. Because not only is it a good idea to give people products they find useful, but this is some of the kind of -- these are some of the kinds of things that can make the value (unintelligible) clear to people who might be skeptical often about it.

Other thoughts? Dan, you're incredibly quiet.

Daniel Atkins: You know, I don't have the depth of understanding of specific projects. I guess the thing that I was going to talk about more earlier in the context of the economic data, but I think you'd apply other, and Noel touched on it in his comments was whether we're doing enough around data virtualization products, for example, the tableau. They said they were using tableau for the virtualization. It wasn't clear whether that was just for internal use or whether there are versions of tableau that would provide viewing opportunities or even better yet customizable toolkits that external users. And I was also surprised not to hear anything about geospatial representation of the data.

So the emphasis quite properly was on accelerating the delivery of these products, and more timely data, and using a variety of kind of not less burdensome ways of capturing it. But in the context of this kind of brainstorming session, we really want to stretch our vision and aspirations to really -- I also -- in some of our previous meetings, we've talked about educational use of census products and I found that extremely exciting, and never have we needed more data driven decision making and making our general population more fluent in evidence based decision making and so forth.

And so anything we could do there and of course the virtualization capabilities. I mean even the new VR technology that's very rapidly becoming pervasive. I would think even that would have some applicability in the useful products from the Census Bureau. So I guess the general thing I want to say is we've got this day-to-day reality of the 2020 Census and so forth. But if we could have some skunk works activities around more visually intuitive over the horizon ways of interacting with the data, it would be a good thing.

Barbara Anderson: Doug, I and I think everyone else is glad to see when you're back -- now that you're back from your sabbatical. So welcome back and do you have any thoughts on this general topic? Okay. And Jack, I'll bug you.

Jack Levis: I wish I had an answer. A lot of what my organization struggles with is the predictive side, right. So it's not just what happened yesterday but what's -- Christmas is right here. What's going to happen in those days of Christmas and I'm not sure the census data is going to predict shoppers' ecommerce usage during the Christmas days. So I've been struggling with where we're trying to gather data that if we just had the Census Bureau data we could grab it and say we're good. I'm not sure yet I know what that is because I think it's mostly descriptive. It's mostly what's happened and I honestly believe the data you need to predict or to optimize is different than the data you need to describe. And in most of our areas, we're in the predictive world, what's going to happen. We'll do our own surveys on shoppers' sentiments. So we'll go out and gather our own surveys there and that almost becomes a corporate asset, the surveys that we go and publish ourselves.

So I really need to think about it, Barbara.

Barbara Anderson: Although some of the things Noel was saying earlier, not in this session but earlier, related to modeling is pretty related to prediction.

Jack Levis: Oh, absolutely. Absolutely.

Barbara Anderson: Noel and then Sunshine.

Noel Cressie: That's right Jack and I know the Census Bureau has a very fine group and time series. They've sort of looked at prediction for a long, long time but generally without the geography and you need the geography as well as the prediction. And that really speaks to this spatiotemporal dynamical structure that Barbara was referring to.

Jack Levis: And that's what you mentioned earlier, Noel, right, the three pieces.

Noel Cressie: Yes, and then (unintelligible) because nothing works in isolation. Everything is related to everything else. So perhaps, I don't know if -- this is time for us and John (unintelligible) I know is very interested in this and has put some resources into it. And I don't want to be too self-serving because I'm indirectly involved in some of that. But I don't know if this a time where John might just sort of mention what (unintelligible) division is doing on space, time.

Barbara Anderson: Why don't we hear Sunshine's comment and then give John about two or three minutes to fill in on whatever you were talking about, Noel. Sunshine?

Sunshine Hillygus: So this isn't terribly coherent so I'm taking this at face value that brainstorm is okay. But as a political scientist, I don't tend to use a lot of Census Bureau products directly with the exception of the current population survey November supplement. And yet what I pitch about the Census Bureau and its importance is the fact that it's truth benchmark for every survey and poll that is conducted, right. And so the ACS or the CPS is that truth

benchmark and yet has not necessarily -- doesn't necessarily think of itself as producing products for that truth benchmark.

And to give the example is the big debates within surveys, right, is non-probability sampling and how you combine cellphone only households and RDD telephones. And yet cellphone only estimates are not actually coming from the ACS or the CPS, right, but a health survey. And so the question is, is can the Census Bureau kind of take ownership of the fact that they are the truth benchmark for every survey and poll that is conducted and kind of think explicitly about that and the products that it produces.

Barbara Anderson: John (unintelligible). Do you want to address what Noel was talking about for a couple of minutes and then you can transition into your presenter role for disclosure avoidance? See, we like you a lot, John. Go ahead.

John Abowd: So I don't want to take a lot of our time although we can come back to it during the other open time. It is true that I have initiated some explicit uses of multi-variant spatiotemporal modeling and I include either -- in the spatial domain, I include industry relationships. It's a different kind of space but it works the same as aerial representations, trying to encourage teams in economics and in the ACS to use these techniques.

We have brought Scott Holland, who Noel was referring to as one of the people that he has worked with -- on as an IPA. He started two weeks ago and he has three explicit projects along these lines. I think we probably don't need more detail on that but we'd be happy to talk about it at a future meeting and if you ask, we will.

Barbara Anderson: Well, we're good. I think we finished this two minutes early. I think we're -- everyone has either talked or had a chance to talk. And so now, other --

slightly different hat. John Abowd is going to present and we're going to have a discussion on disclosure avoidance and we'll find out what in the world that's talking about.

John Abowd: I think I can see you all better from here and (unintelligible) suggested that this would be a better place to make the presentation from. So I'm coming to give you a brief version of a talk that I've been giving basically around the world. In fact, this is the one I gave to the Australian Bureau of Statistics, Noel. So it's about the challenge that statistical agencies are facing and the big data era, and meeting their confidentiality protection restrictions at the same time as they try to deliver data of the quality that their users expect.

And it's an enormous challenge, and we are undertaking it for many of the flagship products of the Census Bureau as I speak. So it's really my attempt to reach out to the high end of the user community, the academic end of the user community -- I didn't mean that in the high low sensibility -- academic and the part of the user community that I understand best to try to explain it in language that social scientists use to think about these problems. And then hopefully we can have a discussion. I probably can't give this talk in 15 minutes but I am going to just deliberately skip some slides. The accompanying paper that I sent was written as a spoken lecture. So it has actual words that go with a lot of the ideas here. Because I give this in lots of places, it has the standard disclaimer.

So there's some very important concepts that are involved here. There are essentially four. One seems obvious that we should use the confidential data but it's not as obvious as it seems. The next is extremely hard for people who haven't worked in this area to understand. It's called respecting a privacy loss budget. It's a formal quantification of what it is we do. The third follows from the first and the second. So if you're going to respect a privacy loss budget

then you have to prove that you respected it. So that puts requirements on the algorithms that you use that are not obvious in the traditional approach to this subject.

And finally, what I think is the thing that we're most worried about is the release it then forget about it, which is pretty much the only thing a statistical agency can do when it puts things in the public domain is release it and hope it's safe. These methods say you can release it and forget about it, and they show you exactly why. And that's the huge payoff to switching technologies here.

But there's a very, very important social choice question that no matter how much research we do at the Census Bureau, we're not going to be the ones who are making that social choice. We're going to be able to collect data about how that choice is made. There's an explicit tradeoff between published data accuracy and privacy protection and we can show you the technological possibilities but it's a matter of social choice. It's a matter of public policy where you live on that curve and the models for doing that come straight from social (unintelligible) statistics of computer science.

All right, so you would think it would be obvious that we should use the confidential data and it is of course. The problem is that modern SDL methods fail egregiously on this criterion. The most popular, still by far, is the one that was invented in 1972 by (Ivan Faleggi). It's suppression, complementary plus primary suppression, which is probably safe against certain criteria, but criteria that you shouldn't be able to reconstruct a record from the comparison of any two tables. That turns out to be an incomplete criterion and that's why it's not safe all by itself.

But the more you ask for temporal, spatial, or industrial granularity, the more these methods suppress and that is non-ignorable missing data. The suppression rule depends on the values you're trying to see. So it's -- right out of the gate it's not ignorable and you have to understand that. And we don't release any tools for managing that. That's all on the user and good luck to the user on that score.

Almost every other method in use deliberately conceals, as confidential, all the essential parameters for recovering its effect on inferences. And so I had a long paper that was published in an economic journal because economics journals will actually pay you money to write papers like that, whereas stat journals not so much. so I refer you to my Brookings paper with (unintelligible) if you want more details on that.

All right, there are inference valid statistical disclosure limitation methods. I'm going to use the words synthetic data in their statistics and computer science technical sense here and not in the sense of simulated data. Synthetic data have certain statistical properties and if you combine them with validation, they do make effective inference valid use of the confidential data. That is you can publish the results from a synthetic data analysis and the statements you make about the hypotheses under test in those papers are statistically valid. The measures of error are correct.

Privacy preserving data analysis is the way this was developed in the computer science community and has the same property that the statements you make about the underlying confidential data are inference valid. Okay. These are also scalable. We could imagine ramping them up so that the whole population of the United States could use them effectively but they're not plugin tools. We don't have them in our hands right now and so enclaves the Federal Statistical Research Data Centers and virtual enclaves are an essential

part of the transition and long-term solutions. They just don't scale. Anyone who's been associated with setting up one of these and working with the Census Bureau or another agency in using one understands how valuable they are and how statement that they do not scale doesn't really need much more data and I'm not going to provide it.

Okay. So let's get the hard one, respecting a privacy loss budget. What does this mean? I will probably blow past the slide with the technical computer science in it, but there's a theorem in computer science that sometimes goes as the fundamental law of information recoverability or the database reconstruction theorem. And the database reconstruction theorem says that if you release too many accurate summaries of any finite confidential database you expose the confidential database to an arbitrary level of precision. And if you're not familiar with that work, we can talk about it technically but it basically says there's an information limit on how much you can publish from a confidential database. So if the public policy says that the data in that confidential database aren't confidential then of course the way you expose it is you just publish all the micro records. And that's what users would like to have but that doesn't satisfy the requirement that you protect the confidentiality of the underlying database.

If you're going to respect the confidentiality of the underlying database then you have to do things that accept this fundamental database reconstruction theorem. Because it's not going to go away. It's not like it's just somebody's nice idea. It's a provable feature of these databases. So data publication always involves some privacy loss and what modern methods do is they have a precise quantification of that privacy loss and because they have a precise quantification of the privacy loss, you can display the technology for how to manage the privacy loss. That's something that traditional SDL methods don't have and that's why I'm trying to begin conversations about how to move from

traditional ways of doing this to ways that have formal properties that let us work mathematically and algorithmically with the tools that we need.

Okay. So I am going to blow past this slide but I'm going to make one point. If you don't understand how these privacy loss quantifications work, go back to your origins and survey methodology, if you have such origins, and if not, bear with me for a second. Randomized response is a provably safe way of protecting the privacy of a confidential data item and the way that it works is that it reduces provably deniability that the answer that you gave actually represents the answer to the sensitive question. So if I want to ask you a sensitive question, there's a variety of ways of randomizing this. The one that I learned is all the answers are in sealed envelopes that are identical. All the questions, rather, and sealed in envelopes that are identical. That basket of questions contains some percentage, the sensitive question, and some percentage a random or set of random questions that all have an answer yes with a known probability. Usually they're based on birthdates so were you born in the first half of the year or were you born in the second half of the year. Those have answers with population probabilities of a half.

So if you manipulate the design, not the data, this is important, but the design of a randomized response survey, you can show that the protection comes from the fact that the maximum amount you can learn about a respondent, that's called a Bayes Factor in statistics, that maximum Bayes factor is provably bounded by things that only depend on the probability that you ask the sensitive question and the probability of a yes on the non-sensitive question. They don't depend on the answer. Traditional SDL privacy protects about using the answers as its inputs. Modern privacy preserving data analysis protects by using the design. It's just like saying that we used to census everyone because we didn't understand how to sample. When we learned how to sample, we accepted that we were going to have sampling variation in

exchange for much lower cost in creating reliable estimates. Moving to formal privacy is the same conceptualization.

The provable bounds come from the design of the way you did the protection, not from the actual answers. And so in a simple case where you ask the sensitive question half the time and you ask it an innocuous question that has an answer yes the other half of the time, you can prove that if I say yes, the maximum you know about me is that I'm three times more likely than a random person in the population to have done the sensitive thing, say commit a violent crime. If I say no, you can prove that I'm one third as likely as a random person from the population to have done the sensitive thing, again, say, commit a violent crime.

But that's all you're ever going to be able to learn. No matter how much data is published into the infinite future, from those specific survey answers, you're never going to be able to learn more about me. You might learn more about me from other data but that's outside my control. You're never going to learn more about me from those data because my yes has been properly masked and this is the technique that Google uses in its report system for allowing its internal users and cloud suppliers to gather statistics about their users. It's what Apple has said they're going to implement for their gathering statistics for their type ahead systems and other user interfaces. It was invented at Microsoft, this particular way of formalizing the problem that -- not the randomized response version of it.

So the natural question that statisticians ask at this point is, well, if you do this, what happens to the data quality. And again, I'm going to mostly blow through this slide but you can show that for a survey that you did randomized response, your loss of estimation precision, the sampling precision, falls as a factor that depends, again, only on the design of the randomized response, the

probability that you asked the sensitive question. And in this case, if I ask the sensitive question half the time and the answer to the non-sensitive question is yes half the time then I'm only going to get 25% of the sampling precision that I would have gotten from answering -- asking the sensitive question all the time.

So there's an explicit tradeoff between the data accuracy and the privacy protection and you can quantify it. And you really can quantify it. For randomized response that is the production possibility frontier that you get from randomized response. So on the X axis is the logarithm of the maximum base factor, what the privacy preserving data analysts call the epsilon from differential privacy. That's on the X axis. So as you increase X there's more and more privacy loss.

And on the Y axis is how well you do with the sampling precision relative to how well you would have done if you'd just asked the sensitive question straight out. So the Y axis naturally asymptotes at one because if I do just ask the sensitive question all the time then I have no -- and I get a truthful answer, which is the right way for us to think about it because if we're loading in tax records or we're loading in the answers to a compulsory survey, we really have asked everybody the sensitive question.

So you can get full precision, as you can see, with a -- in a randomized response design with a privacy loss of about six. So sorry, I didn't bring my table with me if anybody happens to know what (unintelligible) of the six is we can unpack that into a -- I think it's around 1,000. So at six is about 1,000. I know the ones we've used regularly so on the map has an epsilon of 8.9. So the privacy loss there is bound to a base factor of 18,000 to 1. Okay. But this is just a technology. That's -- this curve is described by economists as a production possibility frontier. It's the best you can do in an efficient

implementation of these algorithms. It's described by statisticians as the risk utility curve and also as the receiver operating characteristics curve. If you think of the privacy loss as an equivalent to a false positive on a particular outcome and the accuracy of the statistic is equivalent to a true positive, and you can interpret this curve exactly the same way as an ROC curve.

But it is just a technology. Any point on that curve is feasible. So that doesn't help us decide what the right one is and here's the crux of the problem. The computer scientists who invented this behave like the marginal social cost associated with privacy protection equals the marginal social benefit of the privacy protection down in the range of privacy protection of some place between 0.1 and 1 where I put the box in the lower left hand corner there. Whereas statistical agencies and social scientists have always acted as if marginal social cost is equal to marginal social benefit at a much higher end, say, up there where the privacy loss is closer to five.

And absolutely nothing in statistics, nothing in computer science speaks to the issue of where to be on this curve. We have in fact as an agency acted like the boxes I put in the upper right hand corner. We have released extremely accurate data but using SDL techniques that are dominated by the ones that are in that graph. So we're some place in that interior and that's just inefficient.

Okay. So here are some examples from the CS literature. The most outspoken and best known advocate of these procedures is Cynthia Dwork who worked until last week at Microsoft. She's now at Harvard. And she says epsilon, that's the bound on the Bayes Factor. Epsilon is public. The choice of Epsilon is essentially a social question and beyond the scope of this paper. That was her 2008 summary of differential privacy. She said it again (unintelligible) literally in 2011. The parameter epsilon is public and its selection is a social question. We tend to think of epsilon as, say, 0.01, or 0.1, or in some cases

Log 2 or Log 3. Log 3 is 1. So the Census Bureau is the only organization of statistical agencies in the world to have ever implemented one of these systems. We were actually the first organization to implement one period without qualification when it was applied to the residential side of the on the map back in 2008 and the Epsilon is 8.9. So that's about 18,000 to 1 is the bound on the Bayes Factor.

That was chosen to get accuracy like the users have come to expect of data that are published by the Census Bureau up on that high end and not like the accuracy that Google gives its rapport users, which is down in that lower box and the accuracy that Apple will have to accept for the type ahead system if it wants to use the values of the parameters it's been publicly talking about. Okay. So this is really an important social choice problem and this is where I need your help because we don't have much data or even good conceptual ways of doing this. As it turns out, we do have some data and in fact, as it turns out, the best data for the subject were collected by the Census Bureau in the Center of Survey Measurement where every night on the Gallup overnight poll, we ask a battery of questions about how our citizens feel about the trustworthiness of statistical agencies, the accuracy of their data, the usefulness of the data for making policy decisions, the problems that they might see with record linkage in terms of the privacy.

And so we use those data to parameterize the population preferences on privacy loss and data accuracy. And that lets you choose an optimal point because you can use those data to represent the preferences of the population. If you're an economist that straight line there (unintelligible) a social welfare function. If you're not an economist you can treat that last sentence as gobbledygook. But it does show you where, if you use the social science here, the survey evidence suggests that the marginal social benefit is equal to the marginal social cost and it's at a level of privacy protection that's quite a bit

higher than the computer scientists have been assuming but at a level of accuracy that's somewhat lower than what we have been used to publishing.

And this particular parameterization is about a privacy loss of  $X^{3.5}$  so the logarithm is 3.5 and accuracy of about 85%. I think that's actually on the next slide. So yes, there's the (unintelligible). And you should ask the -- in the randomized response context, you should ask the sensitive question about 92% of the time. So you can do this and in fact, we have teams working on the 2020 Census, the American (unintelligible) Survey and the Economic Census, trying to do it as we speak. And they're using algorithms that allow you to prove that the privacy loss budget was respected.

They are acknowledging the database reconstruction theorem. There's the slide I said I was going to go past. I might not blow past the last bullet on. This isn't just idle speculation. There's already a paper in the statistics literature that does an essentially complete reconstruction of the quarterly census of employment and wages by using just the table structure that was published along with the data that were published and the knowledge that the suppression rule was to suppress an interior cell and not a marginal cell. And so once you understand the (unintelligible) then the time series for Noel's benefit, it uses the spatiotemporal relations among the tables published by the BLS to reconstruct the quarterly census of (unintelligible) employment wages to a much more accurate level than you reconstruct them from the comparable data that are published by the Census Bureau based on a similar implementation.

And just so I don't hold all -- one sister agency up without admitting (unintelligible) difficult, in the same paper that I cited earlier, (Ian Schmudie) and I take apart the two publications that the Census Bureau has using similar technology. So our disclosure limitation methods are vulnerable. None of

these papers, not Holland's, not mine, had any exact disclosure identification. So it's not a matter of the law says that you can't have an exact one. There aren't any exact ones and so everything out there is perfectly legal but the amount of inference you can draw about the records in the underlying database is not controlled by these procedures. And once you control it then you can say we're allowing inferences up to this level of accuracy and not anymore and we can prove that that's what we did and you can come in an audit the algorithms. We're going to show you the entire code that was used. We're going to publish all the parameters. You can calculate the inference correction directly or we'll calculate it for you and you can audit our calculation.

Okay. So that is -- you can prove that these algorithms are resistant to all future attacks and that's the basic -- in the new world, where more of the data is outside the firewall than inside the firewall -- that's the kind of insurance that an agency needs to buy, assuming that the statutes aren't going to be changed to say that we no longer have an obligation to protect the confidentiality of the underlying records. I don't think that would be a good change in the law and so I think that we need to, as a community, think about how to get past traditional SDL and move into the formal privacy area.

I went a little over but that's really all I had to say. So I can stop here.

Barbara Anderson: We have I think until 3:35 in total, but don't give more of a talk.

John Abowd: No, that's fine. That's my...

Barbara Anderson: Because we want time for questions and comments. Sunshine?

Sunshine Hillygus: So I guess the thing that comes to mind for me is not -- so this is cool stuff. In practice, right, is it enough to find the one point, right, using aggregate survey data when in fact there's tons of heterogeneity in any given item, right, is not sensitive for all people in the same way. And at the end of the day, what matters potentially is not so much about the actual risk of disclosure for perceptions of such. So that not only is it how much are we saying that we can protect somebody's political information or income information but also are they willing to even provide it in the first place and do they understand that they have protection. And I don't know if that's a piece of the puzzle that you're thinking about at all. But I mean when you're talking about data collection, at the end of the day you also want to know if people will be willing to give you the sensitive information in the first place.

John Abowd: So of course we're thinking about that problem but we did focus at the beginning on mandatory surveys and another place where there's been a lot of focus on things that are built up from administrative record systems where there's serious money spent on enforcing your participation, but not because it's a survey but because it has to do with some legal interaction.

So it is the heterogeneity and people's preferences that determines that curve that I put up there and so that's actually captured in the formal model. To the broader question of the difference between what you actually do and a perception about it, this has been a hard question and Jennifer Hunter Childs and the CSM lab have been trying to struggle with it, a comparable but much smaller scalar lab at Cornell tried to deal with it too. It turns out that people understand this for health data. They understand that these big medical databases contain a lot of very sensitive information about them and that if it's aggregated and the models -- I won't use the world model -- but the summaries are used to give better diagnostics and to better predict treatment efficacy that they will benefit.

So they get that enough people have to consent to have their data used in this way for the whole thing to have its public good social value and they're willing to tell you how they feel about it. And so one of the papers I cited up there has some analysis of these medical outcomes. What we lack at the moment is a good way to teach a survey respondent that there is a tradeoff and then elicit preferences about it. And that's still a high priority laboratory exercise to try to figure out how to do that.

Barbara Anderson: (Jeff)?

Jeff Lower: Hi, Jeff Lower:. Two questions. The first one, are you able to quantify and in terms of dollars or budget percentage how much of the budget is spent protecting against disclosure?

John Abowd: Yes, but let me clear up a misconception. Some of my own colleagues had this. When I talk about the privacy loss budget that's not a dollar budget. That's an information budget, okay. So you're holding constant the amount you spend on the survey and in terms of the amount of resources that are spent at the Census Bureau on this part of the problem, let me just characterize it as quite a bit less than I've spent on many other parts of the problem.

Jeff Lower: And I ask that because...

John Abowd: It's a safe statement.

Jeff Lower: I was at a...

John Abowd: I'm trying to corner the market on people who know how to do this. I've been making offers as fast as I've been allowed to. HR has been making offers that I recommended as fast as we've been allowed to. Completely clear.

Jeff Lower: I was at a speech with Jim Medlock who's the founder of Intergraph and he was asked this last year whether or not he supported manned space travel. And so this is a guy who was the -- designed the navigation system for the Apollo aircraft and his answer was no and the answer was -- the reason that he gave was he says you spend 90% of the time, and money, and effort in protecting the human and you're only left with 10% for the actual exploration. So he said he'd rather have the budget allocated to exploration versus protecting the human and that's kind of where that came from.

And then the second question is can a respondent waive their right for the protection. Like I give an example there like (FERPA) laws, the protection of your child's educational records. As a parent, I can waive that right and allow my child's records to be disclosed or used in whatever manner. So can a respondent -- I mean is it their right to be able to provide information and give some sort of waiver that releases the Bureau in terms of their obligations to protect it?

John Abowd: So I defer to our policy officer -- answer to the question about whether the current law would allow that under some configuration. I don't know the answer. I will say that we have already encountered a comparable situation. So there's a product that's put out (unintelligible) administrative records on employment and it was combined with data from the Office of Personnel Management on federal employees. And as it turns out, the federal employee data was in the public domain. We approved this by forwarding the exact data that were used in the census product and delivered another copy to an external environment.

So that's the confidentiality protections couldn't be applied to the data from OPM, because if you did that would expose the parameters that you were using for the rest of the world. So two separate systems had to be developed and two separate systems can be developed in the situation where some of the records are declared public. Again, it's a public policy decision. Doesn't have anything to do with the math or the statistics. It's public policy conditions that can be accommodated in these -- in the design of these systems if such an environment occurs.

Man: Our title says that we have to protect the privacy of individuals. So that's what we do. So if you give us your data, any table we put out, we try to protect your information. Nothing's perfect as John mentioned, but even if you were to say I waive this, we don't do that.

Jeff Lower: It's kind of getting at more of a way of maybe a proactive collection of data and things like -- and I mentioned this the last meeting, but like in TurboTax, I fill that out every year and it probably has 80% of the information that the Census would require. And if there was a way for me to add the additional information the Census needs and say, okay, I'm going to submit this to the IRA and by the way, the portion that census needs goes to census and I'm fine with the convenience on my part of being able to provide it in that manner.

John Abowd: That would be okay if we could ever work that out because we're -- once we take it in, we're not giving it back. Just to push the limit a little bit, we get the post office's delivery sequence file, which they will sell to certain vendors or give -- I won't say sell -- give it. That is covered by our Title 13 and we can't release it. Once we get it in and put it in our files that's covered by our Title and there's a Supreme Court case that actually determined that.

Woman: Just one clarification. So there is a distinction though between the Title 13 data and things that are -- when we're talking about privacy -- just to be clear. Because people can waive the Privacy Act. It's titled, so just to be clear, it's Title 13 data that we're talking about that under the law, we don't really have exceptions to how we protect Title 13 data. But in the Privacy Act, people can waive their right to privacy for, like, their PII for a variety of purposes.

Barbara Anderson: Willie?

Guillermina Jasso: What's the state of knowledge on the following? As you know after -- is it 75 years -- complete census records become available. So what's the knowledge about those data reveal information about data ten years later, 20 years later?

John Abowd: Well, there are some very interesting projects underway to link those data to current data but I haven't seen any of the fruits of that research yet. I will also say I make the argument in the paper that I circulated that if you interpret equal protection under the law as applying to the Title 13 confidentiality protections that one of the big advantages of this way of thinking about disclosure limitation is that it provides equal protection under the law for everybody. The bound on the privacy loss applies to the entire population. If you relax it, everybody takes the hit. If you contract it, everybody gets the benefit and there's no rivalry in the consumption of the privacy budget.

Barbara Anderson: Thank you, John. We're going to have to move onto to the next thing.  
Well, gang...

John Abowd: Thank you all.

Barbara Anderson: Thank you. Thank you for both of the things you told us and both the topics you spoke on. I've been thinking about how we're actually going to

manage this recommendation procedure and a few -- couple of hours ago, it occurred to me what I'd mentioned to some of you and I think is about the only way we can come up with recommendations that aren't totally simple minded. And that is we're going to do sort of a sub-group project, folks. You're not in second grade but it may feel like it, where I ask -- and the topics that we talked about today where we potentially could develop suggestions or recommendations are the economic programs, big data, 2020 Census updates, the session on potential census data products and disclosure avoidance.

So I asked a person the discussion if they're there, other people who all agreed to be sort of the organizers of collecting potential recommendations on the topic and for economic program, the discussion, Krishna, for big data initiatives Roberto was on the phone and that's not going to work. So Noel is going to be the organizer for that. On census updates, it's Sunshine, our very good discussant. On the brain storm session, it's Ken and on disclosure avoidance it's Willie.

So what I want you to do is we're going to spend the next half hour -- when I talked to (Sara) a long time ago -- not so long ago -- about this new procedure, she told me that what I'm proposing now is actually, it's legal. Isn't that good? We try not to violate any law we can possibly avoid. So what I want you to do is that people who are especially interested in one of these topics, and you can move around, the organizers stay where you are and people go to where they are and you develop a draft set of comments and recommendations on the topic, having -- on the given topic. There are five topics today. There's two topics tomorrow and we may actually get a set of recommendations by the end of tomorrow.

And you're going to do that for half an hour and then we're going to -- each going to talk about what you came up with and other people who weren't

involved in the given discussion are going to contribute what they're thinking. And then everybody individually who wants to, but especially the five people, email me what you came up with and I will try to put this together tonight and we're going to try to come up with something coherent and not totally obvious and stupid.

So does that sound all right, folks? So we're going to be doing this until, say, 10 after 4:00 and then we have until 5:00 when we adjourn, although we need to stop the general discussion a little before that so people can email it and the IT people can help people who are having a problem. So go for it, gang, for the next half hour.

Noel Cressie: Do you want to say who's who?

Barbara Anderson: The names?

Noel Cressie: Yes.

Barbara Anderson: Again? Krishna for economic programs, raise your hand in case people forget who you are. Noel for big data. Sunshine for 2020 updates. Ken for the brainstorm session on census data products and Willie on disclosure avoidance. You can locate yourself however you think convenient but you can all distribute yourselves and start inputting great ideas and those of you who are the organizers paid special attention so you can start drafting what you think on your given topic. Be sure at the top of it to label what the topic is you're thinking about so I don't have to play guess a topic, like a student who doesn't put their name on the top of the paper. I read this recommendation, which was this about. So okay. ))

The IT guy is going to give a quick refresher on how to actually email this stuff to me when you're done. So helpful IT guy, do it whenever you want.

(Raymond Lee): Good afternoon, everyone. My name is (Raymond). I'm with TCO and I just want to take this opportunity to reiterate some information that my co-worker, Kyle, provided you guys in the morning session on how to email your recommendations. You can do that from here, which inside the note taking app and in the upper right hand corner, there is a box with a tab...

Barbara Anderson: Attention folks, he's explaining how to use the system.

(Raymond Lee): Thank you. Inside the note taking app, and actually many of you have already recorded your recommendations, you can email it to the chairperson by selecting the square with the arrow in the middle. When you select that icon, you're going to be presented with an icon of an envelope. You select the envelope to mail your recommendation. In the To field, just begin typing CCAC and the email that you (unintelligible) should auto-populate. So you select it, and in the subject line put in your name and hit send and you will email your recommendations.

If you have any questions, I'll be right here. You're welcome.

Barbara Anderson: When you get ready to do it, he will help you. My thought was that we would have the organizer person for each group to go through what the group thought and then anyone else put in comments they want. And if we get through all of this quickly enough then the five organizers would actually have time to, in something resembling English, to -- because it's hard in this kind of stuff -- to incorporate, and whatever, reflect on the comments from the other committee members and get help to send the stuff to me. And maybe I'm too optimistic but I think this might actually work.

And so the first one is Krishna on economic programs. So if you can say what you all thought that would be supper.

Krishna Rao: Sure. Well, actually and before I do that maybe an ask because I was trying to play back the tape in my head of the discussion we had afterwards. And I think I got some of the comments but I think I'm missing some as well. So the ones I remember the most were -- so Noel talked a little bit about this idea of sort of trying to think of everything in this spatiotemporal framework, is that a lot of these sort of retail trade issues can be thought in that same space.

I remember, I think it was the same -- Andrew and Ken had comments about sort of areas of the economy that are maybe under measured by some of these things or not measured -- they're changing, right, and we need to think about the right measurement problem both in sort of ecommerce with these large physical locations to the Amazon warehouse and the idea of the gig economy. And then some questions I think from Allison around sort of the speed of data release and especially around -- I thought it was sort of the survey of business owners in particular and that being quite slow to update and ways to get faster.

And I think I remember both Roberto and Sunshine having a comment in this session, but I don't remember what they were. So if it's not top of mind, we can probably move past it and I think we might have lost Roberto but those are the comments I had from the committee. And then some sort of -- let me open up (unintelligible). Yes, and then a bunch of comments I had that sort of didn't make the slides around for sort of being the discussant here, which were something that also came up later -- this idea of the conversations around standardizing and harmonizing data questions across different agencies and work with the BEA to speed up releases, sort of making as much of that public as possible and sort of updates on sort of how that is working because I think

that's sort of -- we talked a little bit about this idea of not asking the same -- I now remember Roberto's question, which is about company burden -- but not asking companies sort of the same question over and over again, and sort of trying to minimize by not duplicating in some sense respondent burden were some of the comments.

And then also had a few comments on this idea of alternative sources of data in some of these spaces. So is there anything else -- any other things people have to add? I can...

Barbara Anderson: Yes, I think -- we're allowed to say nice things to them. I was and I think a lot of people were impressed by the progress they've made and the improvements they've made in lots of areas by going to an internet version of stuff and that the improvements to that are really impressive. And I think that they -- the Census Bureau should think about what other parts of the Census Bureau -- what lessons they can learn from the progress that the economic programs have made in this. I found it was just incredibly impressive what they've accomplished and I think we should tell them that.

Anyone else on economic programs? The next one is Noel on big data initiatives.

Noel Cressie: Thank you, Noel Cressie. So Allison Plyer and I met and we talked about it. I know a number of you here have interest in the area but we couldn't all be in the places we wanted to be at once. So the comments that I will send to you, Barbara, eventually come in three trenches. The first trench is this notion of gathering building permit data from localities. The problem is one of definition and I think this is, as a general issue, is as you go down in granularity and perhaps going to other forms of government on federal

governments and get into counties, and parishes, and local government definitions change.

And the ability to collect big data at a national level, for national reasons but at local levels, gets harder and harder. And so I guess what we did is identified an issue as to how to do that is how to standardize some of these definitions. There's a suggestion that Allison made that perhaps the statistical agency itself steps in and makes some standardization and the local agencies sort of prepare their reports based on it. It might be difficult simply because those local agencies would then have to employ somebody or there would be a time issue about sort of filling in yet another report in a way that would satisfy the Census Bureau.

But there is that issue of common data definitions. Now, the next issue was based around web scraping and (Carla) went into some detail about how this going to be done. A lot about software, a lot about whether they could do it and the issue is what is the goal in the end. It became clear the interest is in small area estimation. I made a comment about earlier in the economics side of things but it applies to the big data side of things as well that small area estimation doesn't apply just to geography. It applies to also different categories and has a multi-variant field to it.

I guess one other comment that needs to be made is that web scraping by itself without some sort of design survey is potentially dangerous and I'm going to include a reference to Michael Elliott's paper (unintelligible) certainly a referee publication (unintelligible) located right now included in my comments to you, Barbara. But in that paper, he makes it clear that if you try to do some sort of so-called pseudo survey but you don't have any series survey weight, you'll run into serious trouble. The idea is to take, in a sense,

web scrape data, combine it with survey data where you've got reliable co-variants, then link the two and then you can potentially combine strengths.

So that's two out of three and then Roberto's comments, I wouldn't do justice to him in the next two minutes, but it's very interesting what -- it's interesting what Roberto was referring to and it's really -- it's attention, I think, between public and private. All the micro data versus some summary of the micro data and this sort of tension leads to issues about, in the end. Right now, you see the big data working group is hearing about goodwill or hearing about perhaps there is some financial engagement between the Census Bureau and the credit card companies. But we're hearing about, well, this is an experiment. We're going to try it. But what Roberto points out is that is it sustainable. Is there a long-term working model by which statistical offices can work with private companies to mutual benefit. And would that mutual benefit to the private company simply be that statistical offices have a contract with them and give them money in order to get the information.

My personal feeling is that there surely is a mutual benefit that the work that's being done by the Census Bureau could surely benefit the credit card company itself, financial company. And I imagine in these early stages of big data model that they're proposing, that is the interest of the credit card company. They're trying to see what the Census Bureau, what more the Census Bureau can get out of their data that they don't -- they haven't employed a quant for.

I also imagine that once they see what could be mined from it, they might go off and employ several quants and end up doing it themselves. So Roberto's comment about sustainability is serious and it needs to be looked at it in a very hard way because some sort of business model is needed for how these web scrape data could be used. So the issue about all the micro data versus some sort of summary, Roberto called it sufficient statistic. Sufficient statistic

has a formal definition in statistics. It means that you can simply, for a particular parameter that you want to estimate, you can throw away all the data and simply keep that statistic and it has all the information on that parameter.

Unfortunately, we ask many more questions than one parameter and again, my feeling is that if you tie your horse to a sufficient statistic, you only ever get to ask one or two questions. You will surely want to get back to the micro data eventually and I think Roberto implicitly implies that. So that's about it. I haven't really done full credit to Roberto's discussion, which went on and full of lots of (unintelligible) information, but I did my best and I think I'd like to ask if there is anything that I can add in my report to Barbara.

Barbara Anderson: I thought your comments were very good and one way to put, in a slightly less nice way than what you put it, is that the big data census -- and this is about having some weights to put things -- to try to avoid, in fact, an imperceptions, one of the main criticisms for big data is that it's often an inch deep and a mile wide, and not clear what at all its useful for. And that's been my skepticism always about big data that if you can't do what you were talking about it's close to useless and I think they really want to avoid that.

Noel Cressie: So it's interesting. You know how big data working group calling (unintelligible) feel like the curmudgeon who basically says so now you can do it, but what are you doing. You know, what is your goal.

Barbara Anderson: I think that's what we're saying.

Noel Cressie: Right, and they always say small area estimation is our goal. We want more timely estimates at higher granularity.

Barbara Anderson: But of what?

Noel Cressie: Yes, that's right. So first of all, what's the question. I think a number of those questions are framed and formulated within the economic census. So in a sense that's -- it's a good exercise to start with is because you've got a well-designed survey. It's just a bit out of date and it has granularity that you don't want. And now, you've got credit card information and the questions moderately well defined.

So I think they're poised -- it's taken some time -- a bit longer than I thought it would take but they're poised now to do the sort of small area estimation that's needed. And I actually hope that they get in touch with John Abowd and...

Barbara Anderson: It's really closely related.

Noel Cressie: Right because John and through Scott Holland, who is now an (unintelligible) personnel agreement, have that capability to do the small area estimation at that sort of level. And very exciting, by the way, to get it going. Admittedly, it's what you might call a pilot study and in a sense it's kind of perfect, too perfect, right. Because you've got a credit card company who's willing to give you the information. You've got an economic survey, which is well designed. If it can't work for that, you know, obviously, it can't work in worse situations.

But I like the idea. You know, they've set up things designed to succeed but they haven't yet succeeded.

Barbara Anderson: Other comments? Moving right along. We go to the 2020 Census update and Sunshine. And I think as we were talking about besides on this, there were some general points that you were thinking about making, which I think you should make also.

Sunshine Hillygus: Okay. So I mean, some of these general comments actually apply to the 2020 but also apply. So some of those general comments are just that they're - - I think we want to try and make CCAC to be as effective as possible that we, again, would encourage the Census Bureau to limit the length of the presentations and to provide white papers or back up information behind those presentations. I thought, you know, John's presentation was really nice that he gave that background information and that was super useful.

And second is, again, thinking about broadly how CCAC can be useful. It would be really nice to be able to be brought in and actually get access to data. And so it is again this question of -- that was one thing that they were going to check on feasibility and we want to, again, bring that up.

With respect to 2020, there was a lot of material covered, but again the devil is in the details and we wanted the details. We're far less interested in the process and hearing about what the process was that got us to where we are. Instead, we want to know what are the decisions still left to be made and can we be useful in that. And so what we have are a specific list of request, questions to be able to find out some of the details. So what is the analysis that already exists for the Harris County versus in L.A. There was 85% that we didn't -- that the undetermined that we wanted to better understand what that meant.

We wanted to know (unintelligible) records were that were used that were ultimately used in 2016 and then which ones are being considered for 2017 and 2018 to be able, again, to know exactly what's going on. Ultimately, I have a long list of here's specific questions that we have that would be useful for us to be able to make an assessment. And our recommendation is that we have three working groups, whatever they are called, regarding the 2020, one

on user experience, one on non-response follow-up, which would also include the use of administrator records and the address canvassing. And then one on kind of system integration, stress testing, contractor management and integration, cyberattack security issues, all of those. Because one of the kind of concerns is the management of now all of these different outsourced functions and how they're going to be integrated.

For the user experience, that was one that I'm particularly interested in. I think that just from the very short demonstration we saw, it made me want to be able to go in and actually do a little bit more. There was -- quickly I noticed a typo, I noticed that your date of birth could go back to like 1830s. There appeared to be some lack of flexibility in terms of how you entered a telephone number. And I suspect, right, this is all just a function of it was put together in ten days' time. But it was, again, this question of what is planned to make sure that, for 2017 and 2018, that this interface is going to be functioning and set up the non-response follow-up team to have a less daunting job.

So whether we call it a working group or just a set of questions that we might divide into three, I don't want to say webinars, because again I'm just -- I'm very worried about us hearing more about process when really if we're going to be useful, I think what we want to see are the details, the dirty details and we should divide up as committees to hear different parts of the sausage making in those different areas.

So I think that was mainly the gist. I don't know if I missed something in terms of people's feedback on 2020. Oh, another example for the user experience stuff. I would just love to get a copy of things like what was the brochure that was used, the letter that was used and is it what is going to be used in 2017 and 2018 or is that something that's already being tested. And

maybe the response is to say, well, that's a done deal. There are no decisions left to be made and I'm fine with that, and I can move on and say, okay, well, what are the decisions still to be made that we can be helpful on. And that's the piece that I just feel uncertain about is how we can be most useful in terms of the decisions that are still to be made.

Barbara Anderson: And in what you wrote you have the more detail that you talked about on those other -- great. Dan?

Daniel Atkins: If I could just add, I just wanted to underscore what I think was implied in what you said, but to make it more explicit, that I think many of us in the group have a willingness to help out beyond just these meetings if there were mechanisms for doing that, of whatever working groups or whatever. And there's a set of complementary expertise that could be brought to bear on this.

Barbara Anderson: Great. So Ken on the brainstorm session on potential census data products.

Ken Simonson: I heard three types of areas that could be either restored or added to what census does and in terms of a restoring or going deeper, restore the pre-1980 immigration information to series that now show pre-1990 as the earliest arrival date that was -- I think I characterized what you were saying, Willie, correctly. And Irma on behalf of the population association, talked about restoring some of the demographic tables at higher levels of geography. Longitudinally also, okay, thank you, Doug. And going in the other direction, provide more information at lower levels of geography, which may require using some of the anonymization methods that we heard about from John Abowd. I think you were talking Krishna about building permits.

Krishna Rao: There are specific. I'm going to get the name of the dataset wrong but it's sort of the linked location of your employer and living (unintelligible) dataset

where it's very difficult to go to very narrow (unintelligible) geography because it's quite revealing to know where someone lives and where they work. But if there are ways to get around that, people look at it at a much finer level.

Ken Simonson: And Andrew, you talked about using the (PUMS) data to -- it would be good if that could be aggregated by civil unit of government or different classifications than we have now for the geographic units. And then I mentioned wanting more detail or a different classification of structure types for the monthly report on construction spending.

A second category of brainstorming as I heard it or tried to group it is looking for ways of harmonizing definitions across agencies. And I know there are some things already happening with that. We heard from Carma about looking at definitions of products or retailer categories when looking at the various retail information. But certainly, if -- and I think it is ongoing that census is trying to work with other agencies to make sure that we're not picking on data providers to provide the same information to different agencies and using different definitions to get that same information.

And then looking for ways of making census even better at putting its information in front of the public through data virtualization methods or providing (MOOKs) or other kinds of educational tools that people can use to learn from census how they can extract data from the census site or to perhaps census to be the go to guy for learning how to use government statistical information generally. So that's not so much a census product in terms of doing something with its own data but telling the public how they can use data from census or other places.

So again, if there are other ideas for things that census could be doing that I haven't picked up, please let me know and I'll add that to our recommendations.

Barbara Anderson: Well, another thing was what had come up various times about various surveys that the census has done often, to answer methodological questions, and they don't want to let out the raw data because of confidentiality. The idea of putting these in the research datacenters, which have very good confidentiality protection for non-census researchers to be able to analyze the stuff in that setting in a way that wouldn't really threaten anything. And we brought that up related to a variety of studies, surveys that the census has done.

Willie on disclosure avoidance.

Guillermina Jasso: There were no recommendations but one option is to write the following. So everybody tell me what you think, including John. CCAC commends the census bureau for this valuable work in statistical disclosure and looks forward to updates on its progress.

Barbara Anderson: I think that's great. We're certainly allowed to say nice things to people. Yes, Noel?

Noel Cressie: So I think it's an indication of John's -- we're talking about John Abowd's -- John's technical -- we're talking about you, John, technical presentation and the backup work, which requires the amount of time and effort to get into. I didn't get a chance -- we didn't get much of a chance to follow-up with John and perhaps with John here, it might be good if -- can I just sort of ask you, John, whether these Bayes factors and the Bayes factor curve was the sole way that you were assessing, much as you would with an ROC curve, or a

power curve, or that type of thing, was the sole way that you were assessing disclosure avoidance?

John Abowd: Historically or going forward?

Noel Cressie: Going forward.

John Abowd: So going forward that's the methodology, yes. But the choice of formal privacy system that you apply the Bayes factor bound to is the decision -- (unintelligible) make decisions about for the X axis and the choices of accuracy measures are the things you have to make choices about for the Y axis. And so there's -- there are an array of formal privacy systems that we were considering. We're starting with the most restrictive, which is differential privacy and there are an array of accuracy measures that we are considering.

The computer scientists like L1 and the statisticians prefer means (unintelligible) error, which in their world is L2. So we're just basically doing both. And they're all stated as a percentage related to the best case outcome with no privacy protection.

Noel Cressie: So I wondered if there was a way to look at this, a criterion, right, where you're looking at false positives and false negatives. So you think you found something but you haven't or you haven't found something but in fact you think you haven't found something, but you have. And this is something that says, well, we like Bayes factors. They have a very nice interpretation, but in the end when the rubber hits the road, if I do a simulation of this, how often am I right and how often am I wrong. And these are false positives, false negatives, there's sometimes false discovery rates involved. So what I'm getting at is the Bayes Factor the only criterion by which, in the curve, the only criterion by which you're going to assess these?

John Abowd: No, and in fact there is an emerging literature that's got the first real participation by statisticians and computer scientists actively collaborating with a methodological contribution coming from adding how protective of the inference validity are these systems. And there's some very interesting results. So the first of these papers showed up in 2015. So this is really a new literature. Some of them have made (unintelligible) really quite intriguing.

So there's a paper that Cynthia Dwork is a co-author on that makes the claim that just by enforcing the rule that you have to use a differentially private algorithm, never mind the privacy level, you can simultaneously get a false positive -- a false discovery rate control, multiple comparison control. And actually Google implemented that in their implementation of (Rapport). They implemented that.

Noel Cressie: So there is thought about false discovery rates, false non-discovery rates. That's excellent.

John Abowd: Yes. I usually summarize them as the inference about a particular population quantity is valid, including the disclosure limitation rules, which is the same basic idea.

Noel Cressie: Thank you. So I don't want to continue this. It's obviously something John and I can do offline. I agree with what Willie said, which is that it's ongoing and it looks really interesting.

Barbara Anderson: Tomorrow, Jack is going to play the organizer role for the evidence based policy making and Bob is the sort of discussant designated person for the ACS. But Bob has to leave pretty early. So I've been on all these committees with Bob and I think I'm going to play that role for the ACS. At the end of

tomorrow, we will review what happened about the five today and add in the other two and hopefully finish everything up. If the five people who gave presentations here could have, like, three minutes to add in anything they wanted to after this discussion, and then if we can make sure with the help of our wonderful IT people that you can all actually email these things to me, I think we'll be in good shape.

So you have a -- it's 4:42 now. You have until 4:45 to finalize your stuff or until the IT person comes by to help you or you decide to do it. I just don't want us to run into any kind of trouble in terms of deadlines today. So you've got three minutes, folks.

Woman: Can't I just email it to your email address? I can't just send it to your email address?

Barbara Anderson: I think it's better to send, from what I understand, to send it to the CCAC chair email. We have wonderful IT people here if you're having any problems and tomorrow we'll go through all this stuff and we'll have time. So everybody except Noel has sent theirs? When you have, we can wander off toward the bus. When Tommy can adjourn it, we can wander off to the bus. We can't do anything until Tommy does his thing.

Tommy Wright: Has Noel sent his email?

Barbara Anderson: He's not going to until 5:45.

Tommy Wright: Oh, the meeting has ended. Thank you very much.

Coordinator: Thank you. That concludes today's call. Thank you for your participation. You may disconnect at this time.

END