

**Census Scientific Advisory Committee  
U.S. Census Bureau  
4600 Silver Hill Road, Suitland, MD 20746  
Spring Meeting, April 16-17, 2015**

**AGENDA April 16, 2015**

- 8:30 AM Opening Remarks**  
Tommy Wright, Designated Federal Officer
- 8:40 AM Chair Remarks and Member Introductions**  
Guillermina (Willie) Jasso, Chair, Census Scientific Advisory Committee
- 9:00 AM Executive Remarks**  
John Thompson, Director  
Nancy Potok, Deputy Director
- 9:45 AM 2020 Census Update (2015 Census Test, ROCKiT, CEDCaP)**  
Presenters: Lisa Blumerman, Associate Director for 2020 Census  
Maryann M. Chapin, 2020 Research and Planning Office  
Brian McGrath, Associate Director for Information Technology
- 10:45 AM BREAK**
- 11:00 AM 2020 Census Update (2015 Census Test, ROCKiT, CEDCaP) Cont.**  
Presenters: Lisa Blumerman, Associate Director for 2020 Census  
Maryann M. Chapin, 2020 Research and Planning Office  
Brian McGrath, Associate Director for Information Technology  
Committee Discussion
- 11:45 AM LUNCH**
- 1:00 PM Federal Advisory Committee Act Briefing**  
Presenter: Hector Benitez-Solivan, Department of Commerce  
Committee Discussion
- 1:30 PM Center for Enterprise Dissemination Services and Consumer Innovation (CEDSCI)**  
Presenter: Rebecca V. Blash, Research and Methodology Directorate  
Committee Discussion
- 2:00 PM BIG Data**  
Presenters: Ron Jarmin, Assistant Director for Research and Methodology  
William (Bill) Bostic, Associate Director for Economic Programs  
Discussant: Noel Cressie, Member, Census Scientific Advisory Committee  
Committee Discussion

- 3:00 PM**      **BREAK**
- 3:15 PM**      **CSAC Committee Discussion (Moderator: Chair)**  
Working Groups: BIG Data Working Group Plans  
Overall impressions of topics discussed
- 4:00 PM**      **Meeting Adjourned**  
Tommy Wright, Designated Federal Officer

Tommy Wright: Good morning. Good morning and welcome to the Spring 2015 meeting of the Census Bureau Scientific Advisory Committee. We're very happy that you have joined us as members of the committee and we look forward to a productive discussion these next two days.

In view of events at the Bureau of Census recently, it seems fitting that we pause for a moment of silence. And I'm going to ask you to - to do that, in remembrance of Officer Lawrence Buckner who lost his life actually about a week ago, doing his job as the director had said, here at the Census Bureau. Thank you.

As you know, my name is Tommy Wright. I am the designated federal official for the Scientific Advisory Committee. As such, I am required to preside over the Advisory Committee meetings, as specified by the Federal Advisory Committee Act.

And we will hear a little bit more about that today. But there are a number of things I must do each time, even though they are repetitive. So please bear with me. Before we begin, please note the sheet at your seat outlining the emergency exits and safety procedures.

The proceedings are being recorded and transmitted live on ETV and on webcast by way of the Census Ustream Channel. Please be advised that any side conversations will be heard. Every time you're ready to speak, turn on the microphone and clearly state your name for the record.

And I will interrupt you as needed. I hope it's all right to remind you of that. All meeting materials have also been posted on the Census Advisory Committee Web site for the public viewing, online. I'd like to introduce the people at the head table.

To my right is the Committee Chair, Willie Jasso; to - next to Willie is John Thompson, the Director; Nancy Potok is not here now. She will be joining us later on this morning. Next to - where Nancy would be sitting, I think, is Jeannie Shiffer, Associate Director for Communications.

Next to her is Lisa Blumerman, the Associate Director for the 2020 Census. I think I saw Enrique, but maybe I didn't. And no one else there. So to my left is Bill Bostic, Associate Director for Economic Programs. And I did see Brian McGrath somewhere here but he should be sitting there.

Man: He's trying to fix the echo.

Tommy Wright: He's trying to fix the echo. Am I giving an echo?

Man: Oh yeah.

Tommy Wright: I - okay. Yeah, this - I'm aware of the echo problem. Last week we had one. Next to where Brian should be sitting is Tom Louis, Associate Director for Research and Methodology. Next to Tom is Ted Johnson, Associate Director for Performance Improvement.

Willie and I will chair in facilitating your deliberations today and tomorrow. Between the two of us, we will do our best to keep the discussion moving, being mindful of the schedule. We'd like to acknowledge all staff who are participating by way of ETV and the public participation in person.

If - I don't know if I see any faces from the Department of Commerce, CSA or any Congressional staff members. I'm going to pause. We would like to acknowledge your presence. Any regional staff from the six regional offices, I don't see.

We're going to have a brief demonstration of the iPads by Kimberly Vines-Weathers. Kimberly?

Kimberly Vines-Weathers: Good morning and welcome to the Census Bureau Scientific Advisory Committee members. For just a moment, I want to take some time to review with you, the layout of your tablet that you will use to access the documents during this conference.

Located next to your tablet you should find a tablet overview document that details the applications that you will use to access the documents during this conference, as well as the credentials that you will need, to access the internet.

If you have your tablet before you, you should have it open to a (MAS) 360 application which details a CSAC document. And it should be open to a CSAC folder. We - okay. Good morning again. Okay.

Just to briefly recap what has already been discussed, you should find your tablet open to a (MAS) 360 application. In this application you will find a primary folder entitled CSAC.

Under the CSAC folder you will find subfolders listed that will provide you access to the documents that will be presented during this conference. The primary folder is presentations. This presentations folder is a folder that you will use to access all of the presentation documents.

You will also find other folders that capture additional information that will be discussed during this conference. The (MAS) 360 application is the application that provides you access to the documents in read only mode. If you look to the right center of your iPad you will find a square button.

We call it the Home button. It returns you to the main screen. On the main screen you will find two additional applications. Centered you will find the pages application. If you touch the pages application you will find there a notes document.

This notes document, if you touch it, will provide you the ability to capture notes that you want to scribe during this conference. This is the pages applications. This is where you can write and you can type and you can document information that you want to capture during the conference.

Again, returning to the Home button, it returns you to the Home screen. And thirdly, you will find the application entitled Safari. The Safari application is the application that you will use to access the internet.

On your tablet overview document that you have located near your tablet, you will find the credentials that you will need to utilize to access the internet. It provides you the username and the password.

Throughout this conference we will have analysts standing by to assist you with any problems or challenges that you may have accessing any of these documents. Simply signal an analyst and an analyst will be directly over to assist you.

Again, to view the documents that will be discussed during this conference, please access the (MAS) 360 application, access the CSAC folder. And under the folder you will find the nine subfolders contained in the documents that will be discussed during this conference. Thank you.

Tommy Wright: I guess we could ask, are there any questions? Well thank you very much Kimberly. 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 you've recommended on critical program areas, research and methodology.

Sessions have been allotted time for discussions, presentations and committee member discussions, as well as an opportunity for you to jot down any notes you might have following the presentations.

And there is also if it - the ability to print out during breaks if you - if you would like. All presentations, papers, supporting materials and note sections, are loaded on your iPad as was just mentioned.

Before moving onto agenda, we want to thank today's presenters, the discussants, working group presenters, the advisory committee coordinator and working group subject matter experts for their diligence in collaborating prior to this meeting. A lot goes behind the scenes, as you know.

First, on today's agenda, will be our committee chair, Willie Jasso, who will bring remarks as well as introduce CSAC members. Following Willie, John Thompson will provide some executive remarks on important Census Bureau programs and activities.

Lisa Blumerman, Brian McGrath and Maryann Chapin will present the 2020 Census updates with details about the 2015 Census Tests, ROCKiT and CEDCaP. We'll take a break at 10:45. After the break we'll continue the 2020 Census discussions and we'll have a committee discussion as well.

Before breaking for lunch we'll have a professional photographer come - coming in to take a group picture just before lunch. Lunch is on your own and starts promptly at 11:45, as the Census Bureau is located just down the hallway.

At 1:30 we will come back again. We'll have the Federal Advisory Committee Act briefing by Hector Benitez-Solivan from the Department of Commerce.

At 1:30, Rebecca Blash will present the Center for Enterprise Dissemination Services and Consumer Innovation overview, followed by a committee discussion.

At 2:00 pm Ron Jarmin and Bill Bostic will do a presentation on Big Data followed by commentary from Noel Cressie and then a committee discussion. We'll break at 3:00. After the break, Willie will moderate the CSAC committee discussion, to include the Big Data working group.

We'll have an opportunity to talk about its plans at that particular time. Today's meeting will end at 4:00. If you're planning your calendar ahead, the fall meeting is September 17th and 18th. Next year, we said we would advise you in advance as much as possible and to mark your calendars.

Next year, in 2016, the spring meeting is April 14th and 15th and the fall meeting is September 15th and 16th.

As a reminder to the audience, during any of the question and answer sessions during - occurring later today, only committee members are permitted to ask questions and to make comments on Census Bureau panelists. The public will have an opportunity to comment on tomorrow, at 10:45, during the time set aside for public comment.

If anyone intends to give public comment, please leave your name at the registration desk. A few housekeeping things before going onto Willie - committee members must stop at the registration desk sometime today to pick up your travel reimbursement materials.

Due to federal guidelines of governing meetings and conferences, the refreshments provided are for committee members only. Please remember that the bus will be leaving at 4:15 today. For those committee members needing a taxi, please check at the registration desk.

And of course, the rest rooms are behind us. And now Willie, welcome.

Guillermina (Willie) Jasso: Thank you so very much Tommy, and welcome to everyone. It's a pleasure to be gathered here again today, for this, the spring CSAC meeting. I have a few - just a few things to say before I turn to introduce all the members of the - of the committee.

The first is to the Census Bureau and to the family of Officer Buckner, on behalf of the committee we extend our deepest condolences. And we are always grateful for everyone who serves the Census Bureau. And this is an occasion when this service goes to its extreme.

My second remark is also about service. Probably everyone in this room and certainly every member of the committee walking either today or at the last meeting or an earlier meeting, to our meeting, everyone has passed a beautiful wall with the photographs of Census directors going back to Thomas Jefferson.

Today, as I came in and stopped there, I was so happy to see a new portrait. John Thompson is now on the - on the Census Director Wall. It's a - a very, very big pleasure.

Third announcement - committee members, if you would\ like to sign up for dinner tonight, we're going to be passing around a sign up list. Our idea is that we could have an early dinner at a place of our choosing. If you - if you're available and can make it, please circle your name on the seating chart.

So I will pass this starting to my left and then we will come back and do the committee members on my right. And at this point, it is my very great pleasure to introduce the members of the Census Scientific Advisory Committee. So let's begin on my left with Irma Elo.

Irma Elo: Yes. I'm Irma Elo. I'm from the University of Pennsylvania where I am a Professor of Sociology and Director of the Population Aging Research Center.

- Dan Atkins: I'm Dan Atkins. I'm a Professor of Computer and Information Science at the University of Michigan.
- Roberto Rigobon: Roberto Rigobon, Professor of Economics at MIT.
- Doug Massey: Doug Massey, Professor of Sociology and Public Affairs at Princeton University; Director of the Office of Population Research.
- Steve Ruggles: I'm Steve Ruggles. I'm Professor of History and Population Studies at the University of Minnesota, and Director of the Minnesota Population Center.
- Gary Gates: I'm Gary Gates. I'm the Research Director at the Williams Institute at UCLA.
- Ken Simonson: I'm Ken Simonson. I'm Chief Economist for the Associated General Contractors of America.
- Babs Buttenfield: I'm Babs Buttenfield. I'm a Professor of Geography and Geographic Information Science at the University of Colorado.
- Bob Hummer: Hi. Bob Hummer - Professor of Sociology at the University of Texas at Austin.
- Sunshine Hillygus: Sunshine Hillygus, Associate Professor of Political Science and Director of the Initiative on Survey Methodology at Duke University.
- Jack Dangermond: I'm Jack Dangermond. I'm with an organization called ESRI. And I'm interested in geographic information systems.
- Noel Cressie: I'm Noel Cressie. I'm Distinguished Professor of Statistics at the University of Wollongong. I'm also a Director of the Center for Environmental Informatics.
- Peter Glynn: I'm Peter Glynn. I'm Chair of the Department of Management Science and Engineering at Stanford University.

Barbara Anderson: I'm Barbara Anderson. I'm Collegiate Professor of Sociology and Population Studies at the University of Michigan.

Jack Levis: I'm Jack Levis. I'm a Senior Director of Process Management at UPS.

Guillermina (Willie) Jasso: And I'm Willie Jasso. And I'm Professor of Sociology and Department Chair at New York University.

Tommy Wright: And now we'll hear remarks from John Thompson, the Director.

John Thompson: Well good morning - good morning again. I'm delighted to be here with you this morning. And unlike previous meetings, as I'll talk about a little later, I'm going to be missing part of this meeting. But I think you'll understand after I describe what I - what's going to - what I'm doing - why it's appropriate.

So let me start with thanking Willie. Willie - I think it was probably a year ago - approached us about saying that it would - it would - while she enjoyed being Chair, it would be good if we started looking for someone else and let her focus more on being a committee member.

So we've been doing that. And fortunately for us, Barbara Anderson agreed to - thank you Barbara very much - agreed to become the Chair. So this will be Willie's last meeting as Chair. There'll be an overlap with Barbara and Willie.

And then at the next meeting Willie will - I'm sorry, Barbara will be Chairing the meeting. So I think we all want to thank Barbara for - for this high paid duty.

There are a few updates at the Census Bureau that I want to make you aware of. Lisa Blumerman has accepted our offer to become the Associate Director for the Decennial Census. So she's no longer acting now. She is the Associate Director.

So we congratulate Lisa and we're very pleased that she has agreed to take on this responsibility. I also note that we - we have - we've been doing a lot of webcasting. So last Wednesday we had a project management review for the 2020 Census program. It was webcast.

I don't know if you watched it or not. But if you want to, it's available on our Web site. You can go to it and you can flip through it so you can get to some of the parts that might be of more interest to you. And as you know, as Tommy mentioned, last Thursday we had a shooting at the Census Bureau.

One of our guards, Lawrence Buckner, who was just - just basically doing his job, was killed and it's very sad. We've had a couple of town hall meetings at the Census Bureau to talk about it - the incident. It was a very tragic incident. It was also very trying for people at the Census Bureau.

I won't go into details but there was a lot of - there was a lot of activity on our campus. But part of what's going on is tomorrow is - there's a funeral for Officer Buckner.

And Deputy Director Potok and I think it's appropriate that - that we go attend the funeral and express our condolences directly to the family. So we - we won't be here for some - for some of the meeting. But I think you understand why that has to take place.

So let me talk about something else that's been of great interest to me recently and that is the American Community Survey. There - there is a - there are rising concerns among our important stakeholders in the Congress, about the American Community Survey.

And these concerns are - take on two themes. One theme is - is the mandatory nature of the survey, and that is we have on the envelope and - in very bold letters, your response is required by law.

And they - some - some members philosophically think that it's just not appropriate to have someone fill out a very lengthy form and compel them to do it by law. So we - we've heard that.

We've also heard that there are concerns with the length and intrusiveness of the American Community Survey. And there are also concerns in this tight budget environment and Nancy I think is - she's going to talk about the budget. But if she doesn't, I can talk about it a little bit.

But there are great concerns with - with the funding of the American Community Survey as well. What makes this even more of a - of a more urgent issue to us is that the House and Senate are in the process of doing their markups in the very near future.

I don't know exactly when but in the next couple of weeks. And so there are several things that are going on here. Undersecretary Mark Doms and I, are trying to meet with a number - well with all of our members of our oversight and appropriations committees.

And we are basically explaining to them one, the value of the American Community Survey data; the efforts we are taking already, to address the concerns that - that they may have with the survey. And we're explaining to them the critical nature of the 2020 Census.

And that is we - we have asked for an increase of over \$300 million in FY '16, to support the 2020 Census. The President and - the President's budget includes this.

In fact the Department of Commerce and the Office of Management and Budget have been very supportive of our efforts to - to change the 2020 Census. But FY '16 and FY '17 are just critical years in this process.

And we need to do a complete end to end test in 2018, because we are reengineering, reinventing the way we take the Census. We've never written the kind of computer systems that we're going to have to write, to conduct the Census.

And so we - we really want to do an end to end test in '18, of our systems and we're going to, so that we can be prepared to go live in 2020 without suffering something like happened with Healthcare.gov. So anyway, we are doing this.

And as a result of that, this afternoon I will be on the Hill visiting with Senator Capito of West Virginia and the staff of Senator Lankford of Oklahoma. So in - and I - there's a great urgency to do this. I apologize for not being here.

But it's - it's very difficult to get on the schedule of these folks. And - and we don't have very much time before we - to talk to them. So I'm going to be doing that. Let me just say a little bit about some of the things that we're doing on - on looking at the American Community Survey.

What you're going to hear later in the meeting, I believe, is the results that we've gotten to date on our content review of the ACS. And the content review has really gone in depth. We have information now on all of the citations that require the use of the American Community Survey.

We have good information on the geographic levels of detail at which the survey is needed for particular questions. And we can now begin a process where we can look at reducing the overall respondent burden of the survey, by say some questions are only needed at the state level.

So we cannot ask it on every questionnaire. We found that there are some questions that are only needed every three years, which again says we can do something with that. We're in the process of really working through this very heavily.

We're looking at how we could use administrative records to supplement some of the questions on the American Community Survey. And we've gotten a very good report on that from NORC. One of their senior researchers, Pat Ruggles, did a very good job on that.

And we're also looking at a reduced - well we're looking at whether we can optimize the way we visit people in the field and knock on doors. We might not have to knock on doors as much as we do. We're looking at how we can reduce that. And we're doing a test in May.

The mail out will go out later this month to see what is the effect of taking the mandatory language off the ACS. And the reason we're doing this is we don't - we have two data points on this.

We had one data point that was collected back when I was the associate director for - I don't even know if I was the associate director. So there may have been - anyway, but - but that found that if - there was a 10% gain in response - self-response rate from the mandatory message.

But that was in the '90s I think. And then the next most current data was collected in 2003 when we did an experiment with the American Community Survey. And that data is 12 years old and that showed there was a 20% response.

But we want to understand, in today's society, what the current effect is of not asking the mandatory, to get us a baseline as to what - what that effect is. And it is very fortunate that we could do that test. So that's - that's some things that are going on right now at the - with respect to the ACS.

Now on the positive side, on April 20th which is coming right up, I have a hearing with our Senate Oversight Committee and that will be Senator Johnson from Wisconsin is the Chair and Senator Carper from Delaware is the ranking minority member.

So I'm looking forward to this hearing. It's a great opportunity to have a conversation with - with the Senate. And although there might be some tricky questions but I'm still really looking forward to it because it's a tremendous opportunity.

And I'm glad that they're interested in talking to us. I should also note that GAO will be testifying at the hearing as well. So before we really get into the meeting, let me go over a few of the committee's recommendations and some of our responses.

So one recommendation was with respect to our development, our ROCKiT and our CEDCaP development. And the committee recommended a working group on this. And right now I'm - I'm pleased to say that we are going to start a working group on this.

And we are going to staff it with committee members and more on that to come. Okay. There is also a recommendation on 2015 administrative records modeling I think you're going to hear today when we talk about the tests, some of exactly what we're testing on administrative records.

And hopefully this will get us to some of our major design decisions in - that we'll be putting out this fall. And the committee has requested an update on our activities surrounding Big Data. And as you can see, from the agenda, that's - that's on the agenda.

So in conclusion, well let me first say thank you very much to the committee for being here. But the conclusion will be I want to recognize Willie by giving her one, a round of applause; and two, by giving her a very small token of our appreciation.

So Willie we have a plaque for you and I'm very pleased to present it. We have a photographer. I think we do have time for a few questions.

Man: John, Willie wants to say something.

Tommy Wright: Oh. Oh, sure. Of course.

Guillermina (Willie) Jasso: Thank you so very much. This is a huge - the plaque is a huge surprise, the succession isn't. It's been well planned, well deliberated and I think will just work out beautifully. There will not be wars if succession here. Thank you so much.

Tommy Wright: All right. John - I think John had called for questions or comments.

Kenneth Simonson: The couple of comments about the ACS - I had participated in a briefing that the National Association for Business Economics held a month ago to inform media and the public about the importance of ACS, the threat that it was under. I think that was successful in generating a couple of media stories.

I've also participated in several visits to staff of members of the Census - the subcommittee that funds the Census appropriations.

And I'd have to say that there's awareness of the value of ACS but also that these committee members are very much responsive to constituents who complain about the alleged intrusiveness.

And some of them also have the attitude that while we know it's bad to make it voluntary or de-fund it but the Senate will take care of that. And I'm not confident that under the current make up, the Senate will.

So I think the threat is much greater than it was in the last couple of years when there were riders passed in the middle of the night by voice vote at the House. Nobody actually on record for their stand on this. One point that I think there's confusion about, is what is the penalty?

I had heard figure of \$100, \$1000, \$5000 and I think to the extent that you can clarify that, particularly if it turns out that it's "only \$100" that might reduce the uneasiness at least to some members of Congress. I know members of the public still would think that that's a lot to pay for not filling out a form.

But it does sound a lot less Draconian than \$5000.

John Thompson: Let me just say that we - one, the Census Bureau cannot fine anyone. We're not an apportionment agency. We would have to ask like the Department of Justice to do that.

Two, we have never taken any action to penalize or threaten or do anything to a person in - that refused to answer.

We believe that the best way that we can convince people to respond is to explain to them the value of the American Community Survey to their communities. So that's not something that - that we use right now. But in answer to your question, the fine is \$5000.

Peter Glynn: I think I've heard recently that in Canada they've also moved to non-mandatory report...

((Crosstalk))

Tommy Wright: ...would you continue to state your name?

Peter Glynn: Peter Glynn. So let me just repeat what I said. I think in Canada they've also moved to non-mandatory reporting on long forms. I guess that's also potentially another data point to look at in terms of what the - what issues will arise as a consequence.

John Thompson: Yeah, we - we've talked to our colleagues at Statistics Canada about that. So in their last Census which was 2011, they did - they didn't have to go volunteering for their long form which is the equivalent of our American Community Survey.

And the effect of that was that they couldn't publish data for quite a large portion of their rural areas just because the quality in their - as they looked at it, wasn't sufficient to publish data. So it had - it had a fairly dramatic effect on the Canadian Census data from their long form.

Irma Elo: You mentioned that you...

Tommy Wright: State your name.

Irma Elo: Oh, I'm sorry. I'm Irma Elo. You mentioned that you're also looking into the use of administrative records with data. And I'm wondering whether you could say a little bit more about the types of records, because - for two reasons.

One is I hope it doesn't restrict access, public access to the data if there are some IRS records or some income records that - or whatever the records might be, that you're using.

And I think this issue has come up before how the public would react to if they think that their administrative records are being used instead of survey data that they would report themselves, if the administrative records - if it's not clear to them that that was the purpose of those records.

John Thompson: Yeah, so let me - let me first say that if we were - we - we would not restrict access to the data, any more than it's already restricted. And that is when we publish data we don't, you know, tabulate the data, we don't reveal - we tabulate it in such a way that we don't reveal any individuals' information.

And if you want to access the micro data for scientific purposes, that has to be approved by the Census Bureau. And if you want to use other administrative data sources then you have to get approval from that - we have to get that also approved by the agency, from which we're using the data.

Many of the housing questions on the long form, seem pretty amenable to being supplemented by administrative records. The - the question that - that causes the most concern among our respondents and is most difficult to collect, either in person or - or on - or on our self-response, is income.

But we, you know, to - in order to use income - administrative records, we would have to do some very, very careful and serious and thoughtful work with - with the Internal Revenue Service. So I don't know that we'll be able to do that.

But we certainly can explore that. But that's...

Irma Elo: So one potential restriction could be the PUMS data then. You know, you said you tabulate but the individual record data, there might be some restrictions on the PUMS that you would have to the Census Data Centers or something like that. You know?

John Thompson: No. I think we could solve the PUMS problem.

Bob Hummer: Yeah, Bob Hummer. The results from Canada sound worrisome, that you just mentioned.

So I wonder if you could talk a little bit about the May tests here and the different kinds of options that - that are being proposed on the envelope or letter inside or however that's working, so we can kind of think through that a little bit.

John Thompson: I'll let Lisa Blumerman, who's doing the test, talk to you about it.

Lisa Blumerman: So I'll answer briefly. But just as a - a kind of preview, Jim Treat will be talking about this tomorrow in more detail. So the May test is a very simple test that we're proposing.

What we're initially proposing to do is for about 24,000, and Jim will correct me if I'm wrong, but for about 24,000 cases in our production sample that mails on April 27th from that front of the envelope only, we're removing that box and the black outline that says your response is mandatory.

So just a very simple test where we're only addressing the envelope itself. In August of this year we have a more extensive test that we're planning where we're looking at softening the language. I shouldn't use the word softening. We're looking at modifying the language to make it more friendly, perhaps.

In addition to that we're also looking at some of the individual mailing pieces. And Jim is planning to go into great detail on that tomorrow.

Jack Dangermond: Yeah, Jack Dangermond. John, I didn't quite understand what you said. In the '90s without the mandatory language you had a 10% reduction. Then you said something about a 20% reduction. And then this test. What was the 20%? And maybe you could just clarify that.

I didn't - I just didn't - I didn't quite hear it right.

John Thompson: Sure. So in 2003 the Census Bureau did a test to see what would be the effect of not putting the mandatory message on the envelope. And in that experiment and this was on the ACS. This was on the ACS. The other one had been on the Decennial Census and the short and long form.

And so this was just on the ACS to see what the effect would be on the American Community Survey. And that was a 20%.

Jack Dangermond: And then my second question is what - April what is going to be the hearing? What is the date?

John Thompson: April 20.

Jack Dangermond: Thank you.

John Thompson: I think it's at 3:00. Yeah.

Jeannie Shiffer: Hi. This is Jeannie Shiffer, Associate Director for Communications. We'll make sure that we email the hearing information to all of you so that you can watch it. We'll be broadcasting live.

John Thompson: So let me - let me just briefly say a little bit then about - Nancy - Nancy Potok is not here. She's tied up. And she's going to be here at 11:00 but in case the agenda gets a little tight, let me - let me talk for just a few minutes about the Census Bureau's budget.

So we've - we got - we got an appropriation in 2015 that allowed us to continue to stay on the critical path for the 2020 Census. And you'll hear today about some of the testing we're doing. And we're on track to make our major design decisions in this fall.

And - and so that's very good. We've also been able to keep making progress in our economic programs. And so our planning for the 2017 economic Census.

And we've been able to fund, as you know, our big CEDCaP initiative, which is building our enterprise wide IT systems that will support not only the 2020 Census but all of our survey and Census data collections.

In 2016 the President's budget is pretty exciting because it fully funds our work to continue to do great things on the 2020 Census. And Lisa will probably talk about what those are, so I won't steal her thunder. But - but we're excited about that.

We are also excited because we can continue to move forward with CEDCaP. We can continue to move forward with planning for our economic Census since Bill - Bill is going to do the next economic Census totally online. So we're really excited about that.

And finally, we've received \$10 million which doesn't sound like a lot of money if you look at our total budget. But we received \$10 million to start work on a clearinghouse for administrative records use for the federal statistical system.

So we're really, really excited about that because it recognizes the - the really good work that we've been doing on administrative records. So with that, I think I'll turn it back to Tommy.

Tommy Wright: We are a little - thank you very much John. We are a little bit ahead of schedule but I think that may be okay. Next on the agenda we have a presentation by Lisa Blumerman, Maryann Chapin and Brian McGrath. I think all three are - Maryann are you - who's - who's talking?

Or there's a question? Sunshine?

Sunshine Hillygus: If we're - if we're ahead of schedule maybe I could get in one last question...

Tommy Wright: Okay.

Sunshine Hillygus: ...for the Director. So my understanding is that the NSF Census Research Networks, will not be going beyond their - their initial term. This was a funding initiative that - that tried to encourage interaction between the Census Bureau and academic researchers.

And so in light of that and certainly the funding situation, did you - could you speak to a little bit to how the Census Bureau is looking ahead as - as another way to kind of engage with academic researchers? Or are there other initiatives, you know, that - that - that you thought about?

John Thompson: So I - I wouldn't - I wouldn't say that they're going - that they're going to go away. We're looking to see if we can find ways to continue them, because they're doing really good work.

And I don't know exactly when - they're having - they're having a meeting in the near future and it's going to be a pretty good meeting. So I don't think that it's clear that we're not going to be able to find some way to fund them. We're trying to do that. We think they're very important.

Barbara Anderson: This will come up tomorrow when you're not here.

Tommy Wright: Say your name Barbara.

Barbara Anderson: Barbara Anderson. This will come up tomorrow when you're not here. But I just wanted to ask, is there any rule that says in Census Bureau decisions about what they're going to ask on surveys and such, that the importance for research cannot be considered at all?

John Thompson: It depends on the survey. For the Decennial Census and the - and the American Community Survey, as part of the Decennial Census, there has to be a clear federal programmatic need for a question to be included on it.

Now having said that, I think we all know that the ACS is a treasure trove of valuable data. And it's used by researchers, it's used by city planners, local planners, counties. It's used by a number of business to make decisions. So it's a very important national resource.

But the requirement to get on - to have a question on the ACS is there has to be a federal programmatic need for a federal requirement like (unfortunately), civil rights. So that's - that's the case for that.

Barbara Anderson: Well the questions that are already on the ACS are there because of some governmental need.

But in terms of what is removed, since they already met that or they wouldn't be there already, I was wondering - and this will come up tomorrow in the discussion of the content review, is there any rule that the importance for research, much of which is useful to the federal government, can that be one of the considerations?

John Thompson: No. We - I mean I'll just take the heat for this, but there has to be a federal programmatic need for having a question on the American Community Survey. And some of the questions - some of the laws and rules have changed.

And there are no longer requirements to be on the American Community Survey.

Gary Gates: This is Gary Gates. But I - I guess adding a little bit - conditioned on there being - obviously there's lots of federal needs that are not on the ACS.

So conditioned on that, I guess what we're trying to say, is could there be a requirement or some consideration - very explicit consideration that, assuming it is - there is some kind of federal mandate for it that also the broader utility of it from a research perspective could be a factor in decisions about what goes on and off.

Because, as I said, I mean there has to be other federal things that people want on the ACS that could meet your criteria that don't make it on there. So there has to be sort of sub rules there.

John Thompson: I'll just repeat - and I think you'll talk more about - there are three types of needs for work.

There's program - there's mandatory which means there's a citation that mentions either the Decennial Census or the ACS is being used as a programmatic, which means that the agency - well I shouldn't go down that path. I'll just let some other people talk about it tomorrow, in content review.

But the important thing is, is that - that - is that there is a - there is a process. There's also a committee that's jointly chaired by me and the chief statistician which right now is Kathy Wallman, of the United States, that reviews the results.

And if there were any question that would go on or come off the American Community Survey. And they - and we could circulate the charter of that committee, right? And you can see the criteria that we use. It has a number of functions.

It's whether it could be used for a sampling frame, it's whether it can - put questions on and take them off. But we'll send that around so you can see what the committee considers.

Steve Ruggles: So in the past when questions were removed from the long form - the decennial long form or the, you know, going back to the - the early 19th century and up until just now, there was always an extensive consultation with the - with the - with scientists and - and I

mean Jefferson went to the American Philosophical Association and pleaded for not - not to have a narrow Census that just covered the - the minimum mandate of the Constitution.

He - he - he wrote that the Census had a - has a higher purpose, offering an occasion of great value, not the otherwise obtained - obtaining sundry facts interesting and important to society and to ascertain more completely, the causes which influence life and health and to furnish a curious and useful document of the distribution of vocations to our fellow citizens.

And I think - I mean I've had students and colleagues investigating the history of how these decisions are made in the Census.

And it's really kind of unprecedented to, you know, not begin these things with a discussion with the - with the relevant organizations - the American Statistical Association; the Population Association of America and so on.

And - and so I was wondering why - why this current - the current decisions were undertaken without - without talking to experts.

John Thompson: Steve, we appreciate that a lot of - three's a lot of science, a lot of research and a lot of use of data that's on the American Community Survey. But we have to have a process that makes sure that - and this was set up back by my predecessor. It was a very good process and I supported it very clearly.

That the functions of determining what goes on the American\ Community Survey now are going to be under the auspices of OMB and the Census Bureau. And they have to follow certain criteria.

It sort of clarifies - and like I said, we have to send that document around so you can see what - what criteria that we do use. But that's - the American Community Survey is at a point where it is - it has to have a federally supported need, have a question on it.

It's the same thing with the Census short form. And Steve one more point. So just so you understand what's going to be happening - in 2016 - I'm sorry, in 2017, we're going to be sending the topics of the - of the Census which includes the content of the American Community Survey due to Congress.

And in 2018 we're going to be sending the actual questions. Now we don't send them for approval. But if the Congress decides to act they can act. And so what we believe is that we need to have a pretty good business case for every question on the American Community Survey that - when we send it up.

Steve Ruggles: Yeah, the - and I think though, in terms of the legal requirements, that case has - has been met that you have documented the mandatory required and everything very - very well.

So once, you know, once that minimum standard has been met there still has to be a judgment made as to what is most important. And the problem with - and then - and the procedures are very well documented. It's on the Web. The procedures the committee used. But they just don't make any sense.

I mean essentially the - the assumption is that the only criterion that's really important, is small area analysis, sub state analysis. And so the questions proposed for elimination are all questions that are not appropriate for sub state analysis, that can't be used for sub state analysis.

And so - but still require very large samples and cannot be moved to another survey.

John Thompson: Well I mean the purpose of the Decennial Census long form was to produce (RAC) level data. That was the design. And that's the - and that's in the criteria. So I mean I think we're probably going to disagree on this point.

But I'm hoping that there is still an American Community Survey to argue about next year. I mean actually, that wasn't the purpose of the Decennial Census long form, I'm sorry. You should learn a little more about the history of the Census.

Tommy Wright: Seeing appear to pause, I'm going to see if Maryann is all right to come forward. We'll hear an update on the 2020 Census update, along with Lisa Blumerman and Brian McGrath.

Lisa Blumerman: Bear with us for just a sec as we get resituated. All right. So good morning everyone. I - I'm Lisa Blumerman. I'm going to start us off this morning. We have a full your to talk with you about the 2020 Census, CEDCaP and all of the exciting things that we have underway in our planning.

I actually have the easy part today. I get to just simply refresh you with where we've been and kind of bring you all up to speed on where we are today.

Then I'm going to turn the floor over to Maryann, who is really going to talk in some detail about all of the testing that we have underway in '15 - 2015 is a very, very critical year for us.

And as we were putting this presentation together, we kept trying to - to cut things out because we only had a short period of time. And (Sara) kept telling us to put it back in, that you all wanted to hear it. So then we got a long period of time. So it's very exciting.

After Maryann kind of walks us through all the tests, Brian is then going to take the floor and talk with us about CEDCaP and where we are in building our enterprise systems and how those systems will integrate with the 2020 Census so we're excited about that.

And then I'll circle back to kind of close us out. And I'll talk with you about where we are with the Operational Plan development and then where we go from '15 in terms of getting from here to the 2020 Census. So in terms of where we are today - oh, you have it. Do you want me to do it?

Okay. Maryann's got it. In terms of where we are today - let me jump right in - I think you might have seen this slide or a variation of this slide previously. It does give us a nice

kind of picture and it's nice to see that it's actually clearer on your iPad than it is on the screen, which is good.

It gives us a good picture of where we are and where we've come from. We're actively working right now, very actively, on finalizing the results from our 2014 test. Our test papers - our analysis papers, are actually scheduled for release the end of May.

So we're really in the last stages of that with the review and approval processes. And I know Maryann is going to talk a little bit about some of our findings. We're also actively in the midst of our 2015 testing. We have two site tests that are active right now, underway.

We've already completed two tests this year. So it's a very busy year for us. And we have a 5th test later this fall, so we are active in Savannah - the Savannah, Georgia area, the 17 counties in Georgia and three counties in South Carolina for a test that's focusing on self-response.

We're also active right now in the Maricopa area, for with the self-response period right now. But that test really focuses on nonresponse follow up, as we get a little bit further along with that.

So both of those tests are live and they've been live since mid - well the Savannah test since February, both tests since March. But where we are is we really spent the last few years really doing our extensive research, our extensive testing, our extensive design.

Really leading us to being able to determine what the Operational Plan is - those major design decisions are for the 2020 Census. What will this Census look like compared with other Censuses?

We know, and you all know, from everything I've said before from what (Frank) said before as me, and from what John has said, that this Census will be like no other. So how does it come together?

How do - how do we get from where we are today to where we need to be, which is really just around the corner? We were talking amongst ourselves last week or two weeks ago actually, on Census Day, April 1. And we had a little celebration for ourselves here.

And it was at that point that we realized that with this past Census Day we're actually closer to the 2020 Census than we are to the previous Census and that's just a monumental undertaking for us and just really kind of exciting.

As you all know, our goals for the Census include designing and conducting a Census that costs less per housing unit than the 2010 Census, while maintaining the high data quality that we saw.

I - and we have accomplished this by really identifying those major cost drivers of previous Censuses and then designing and testing and soon to be implementing innovative enumeration methods and other innovations aimed at reducing these costs.

And our early research and testing has really focused on that. And if you look at that graphic which is part of a larger info graphic that we have, where we are is really coming out of that red which is our research and testing period, into what I refer to as our operational design.

So we still have a little research and testing going on, but it's really about the refinement of methodologies. We're now actively in production. We are building the systems that we will use.

Our research and testing to date has been one off systems, kind of recast as systems that we used in 2010, or previous surveys, with some tweaks for what we think we want to do for 2020 proof of concepts starting with the tests that we're rolling out right now.

We are starting to use those systems that we expect to have in place for 2020. That doesn't mean we won't continue to make improvements. But what it does mean is we're starting that integration now and that's (with) the most critical test.

Just as a reminder, there are four areas that we really identified. We call them our innovation areas. But there are four areas that when we think about it, really were the major cost drivers for the 2010 Census and it's where we focused our initial effort.

And we'll be talking about each of these in greater detail and I tend to veer off the slide when I use it because I'm so excited about what we've been doing and I don't want to steal from Maryann. So I'll try to be careful with what I say.

But in terms of reengineering address canvassing, this is a really exciting area for us. You know, in past Censuses, the last two in fact, had us walk every block in the nation, all 11 million blocks. We had boots on the ground and we walked them looking for all of the - all of the housing units.

What we've determined and what we really believe, and we released this at the end of last year, and what we've been working on all this year, is how we refine our methods so that we have an accurate master address file. But what we really believe is that we no longer need to walk every block in the nation.

Technology has come so far. The use of aerial imagery has come so far, it allows us to do so much more. With that said, one of the questions I hear often when I talk about this, is how do you know what's in those large, multi-unit buildings?

How do you know in areas of economic decline that those housing units that you may or may not seem, aren't really active housing units with residents in it.

And the answer - and it's a very important answer - the answer - and this is true for not just this but for many of our operations is conducting the Decennial Census is never a one size fits all operation. There are always special operations that we have to field.

And with address canvassing this is true and it's very true. And so what we look at with the innovation that we're trying to use - the combination of statistical models, of aerial imagery and change detection, we're looking at where we can identify areas that we feel meet the quality that we need, so we don't have to do in field canvassing of those areas.

We'll be doing in office canvassing of those areas. That will allow us to use the resources that we have, to be on the ground those areas where we need to walk the ground. We're also looking at how we can partner with private sector on some of this.

And I'm not sure if that's what Maryann will cover. If she doesn't I'll come back to that. But we have some exciting work going on - some RFIs that have recently gone out as well as at least one RFP that has gone out and one that's soon to go out, about how we might be able to partner around this.

So I think there is just a lot of opportunity here for us to really actualize what we intend. Optimizing Self Response - our second innovation area - again, what this Census is all about is encouraging people to self-respond.

Making the Census as easy as possible for them to get them to respond early, on their own and hopefully through the internet. With that said, again, it's not one size fits all and we recognize that.

But we're looking at how we can optimize self-response how we can optimize the use of the internet, how we can engage respondents early through partnerships, through advertising, through different techniques. How we can make the Census mobile.

One of the things Maryann will talk about is our non-ID processing. That's what - this is a technique where we're looking to be able to remove or have the opportunity to remove the use of a traditional Census identification number and allow people to respond anytime/anyplace.

So what's most important to us about that because it seems really easy in concept? Sure, I can respond to a survey. I don't need an ID. But I need to be able to take you, put you back at your appropriate address unit, put you back at that address.

So I need to be able to do two things with non-ID processing, in addition to just collecting that Census. I need to be able to validate you as you and I need to be able to validate that response and link it to an address. Our test in Savannah right now is testing this.

We're testing it real time which is very exciting. I say real time because it's an important component. We've done non-ID processing before.

Those of you that are, you know, data junkies an Census junkies that know, we've had programs called the Be Counted Program, where people during non-response follow up could go into a library or community center, pick up a form, fill it out if they didn't think they'd been counted.

We would collect those forms. We would key them and we would process them. And in bath mode or offline we would then match them back to our master address file and do that validation. What we want to implement here with optimizing self-response, is our ability to do this in real time.

If I can do it in real time, if I can take that response that came in without an ID, match it back to the address - master address file, ensure that it is an appropriate response, then I can immediately remove it from that non-response follow up workload. And I can save money that way.

And that's something that we want to be able to do and we're very excited about it. The third area is utilizing administrative records. And I tend to put adaptive design in this area. The last two areas have a nice blend to them.

This is where we really are looking at data that have already - information that's already been provided to the government or third party information, as ways to help us remove non-response follow - as ways for us to help us remove cases from the non-response follow up workload.

There are lots of different ways we can look at this. In its most simplest fashion you can think of that is having almost two buckets - one bucket is really looking at how can we identify vacant housing units and move those from the workload.

And then how can we use administrative records and potentially third party data, to help us enumerate those housing units that we believe to be occupied but where people have not respond. And Maryann will talk in a little more detail about both of those.

Our last area is reengineering field operations. And here when I talking about this, I really am focusing on both the technological innovations we have around this as well as some of the human interactions, the staff management.

We are really taking a hard and whole look at our field staffing structures, the infrastructure, the blueprint so to speak; the number of offices we need to have; the management ratios between those offices; the number of enumerators.

And how we can successfully integrate technology into this. And I know we've been working with a number of you on this.

So you'll get a bit of a preview about MOJO- our new enhanced operational control system, which will allow for very efficient case management as well as the routing and of - of our numerators so that we can be very efficient in how we're doing it.

As well as COMPASS, which is our application that we've developed that is device agnostic for the use of non-response follow up.

No more paper and pencil. But we did show in the 202014 Test that the use of the COMPASS app or some - in this case it was the COMPASS app, but some app on a handheld device, we can effectively conduct non-response follow up.

Those two innovations in combination with how we plan to manage the staff or how we're looking at reengineering field operations. And we are just very excited to see what the increase in productivity will be, coming out of the Maricopa test.

So as I said, 2015 is absolutely a critical year for us. We are on the fringe of research and production, moving into operational design, operational development, into actual production for the Census. We are conducting five tests in calendar year '15.

The first test that we did was a human in the loop in simulation. And I believe you'll see a video. Some of you were here for it. But I believe you'll see a video of that as well, to see what that was. And this was very exciting for us. Maryann will talk a little bit more about it.

But we were able to use a SIMEX, a simulated experiment instead of launching and incurring the expense of a field test, to do some early testing on our new enhanced operational control system, as well as prove in some of the staffing ratios.

So it was very exciting. And we did it - we hosted it here at headquarters with some work at - partnering with MITRE, as well as using about 80 people throughout the country, our enumerators and field staff that were all active in this simulation that we conducted in November.

We've also recently completed the address validation test which is - was our test around the master - the accuracy of the master address file as well as the use of imagery and change detection. And Maryann will talk about that.

We are in the field now for both our Optimizing Self-Response Test and our 2015 Census test. And in September - in August actually, we'll be launching the 2015 national content test which has a September 1 Census Day.

Also, the end of this fall and this fall we will be releasing our 2020 Census Operational Plan, which I think we used to refer to as our major design decisions. At the end of this discussion I'll give you a little more information about what will be included in that.

The version I saw yesterday, because it's a work in progress and nothing's complete yet, what - we're up to about 200 pages of text without appendices. So it is coming together very nicely.

And just one of the things to think about and I'm sure I'll say it again later, is that this release at the end of this year, with our Operational Plan, is three years earlier than the release of the 2010 Census Operational Plan. And so we know it's a living document.

But the fact that we have a very concrete plan and a very concrete vision for where we need to go from today through 2020, puts us in a very good shape for the 2020 Census. We're also beginning our core programmatic work this year, such as LUCA, the Local Update of Census Addresses.

We're doing some research on our partnership program in combination with the - the 2015 Optimizing Self-Response Test, as well as some additional programmatic work. And we're going to continue defining the testing activities so that we can get to the 2020 Census in 2016 and beyond.

With that whirlwind of an overview just to set the stage for you, I'm going to turn it over to Maryann who's going to talk specifically about our tests.

Maryann Chapin: Thank you Lisa. I'm pleased to report that we are live in the 2015 Optimizing Self-Response Test.

At the end of Marcy we went live with self-response in the Savannah media market which is comprised of Savannah and neighboring counties in South Carolina and Georgia. Savannah was selected as a sight test based on several criteria.

It is a medium sized media market; its population is racially and ethnically diverse; it includes households with a variety of levels of internet access and use; and it offers a mixture of address types from city style addresses to rural route designations.

This kind of area can help the Census Bureau test new forms of digital advertising and targeted promotion, to increase response particularly via the internet. The test is critical for conducting early research on the use of advertising and outreach to engage and motivate respondents.

For determining the extent to which we can use preregistration and for testing the operational feasibility of real time non-ID processing and the potential resulting workloads, for system development.

And finally, for determining the extent to which non-ID responses will contribute to the national self and internet response rates. We're looking at how we can engage people. This is the first time we will use real time non-ID processing.

Generating a movement to make the Census mobile, allowing people to respond anytime/anywhere. The Savannah media market covers approximately 400,000 households.

There are three panels that include 90,000 sampled units who will be contacted using our internet push strategy, which consists of an invitation letter, followed by two postcards and lastly, a paper questionnaire. One of these mail panels is testing an early announcement offer.

A postcard was mailed to approximately 30,000 housing units. These housing units as well as others, not sampled, may have been exposed to the media outreach related to the

Notify Mecampaign. The postcard mailing precedes the invitation letter that is part of the internet push strategy.

If the housing unit chose to be notified and provided an email address or a cell phone number, they did not receive the letter of invitation or any of the reminder postcards. Instead, those messages were sent using the preferred method - either an email or a text message.

Housing units that did not respond to the Notify Me offer, are being contacted using the standard internet push strategy. The remaining two panels differ only by whether their initial letters included a Census ID or not. Both received a standard internet push contact.

Our Census Day was April 1st and we will continue our data collection through the end of May. I should stress that this test does not include a non-response follow up component.

Since the 2010 Census technology and the advertising landscape, have continued to evolve along with the public's consumption of technology and advertising.

On this slide, we show several of the print ads that have been used in Savannah, tailored to specific audiences, with messages to motivate self-response.

And now we'd like to share with you a short clip of one of our TV ads used in Savannah.

Man: My Census my job.

Man: My Census my road.

Woman: ((Spanish Spoken))

Man: The US Census is vital to your community and the future. Population counts are used to determine our political representation and resources for transportation needs; for schools; health services; new business development; emergency preparedness; and so much more.

And now the Census Bureau has chosen our community to prepare for the next Census in 2020. Just fill out the test Census online. It's quick, easy and safe to stand up and be counted.

Woman: My Census, my business.

Man: My community.

Woman: My future.

Man: We all count. So go online to [Census.gov/2015](https://www.census.gov/2015) and complete yours today. It's vital to our future.

Maryann Chapin: As Lisa mentioned earlier, the 2015 Optimizing Self-Response Test, is providing the opportunity to better understand response rates when housing units can respond without a Census ID. This diagram outlines key differences in the implementation of non-ID from the 2010 Census to the 2020 Census.

While non-ID processing is not new, the real time processing is. We expect that the opportunity to self-respond via the internet, will impact our non-ID workload. Historically, non-response was a small fraction of our overall response.

However, we anticipate that by promoting internet self-response, the non-ID workload will increase. We'll get some measure of that increase in our real time non-ID response in Savannah.

For the first time, we'll be comparing respondent provided addresses to our frame in real time, meaning during the interview.

Real time non-ID processing in 2020, will allow us to remove workload from our non-response follow up efforts, which has the potential to significantly save some funds for us.

And we'll continue to perform the batch processing and manual processing we have done previously, in order to increase our match rates and further remove cases from the non-response follow up workload. Finally, for 2020 we are exploring further mechanisms to validate the non-ID responses as Lisa mentioned.

Recently, a project was established referred to as the customer experience management project or CEM, to build a prototype of a centralized customer experience data store.

One aspect of this pilot project was to provide real time insights into key operational and communication activities for the 2015 Optimizing Self-Response Test.

To that end, we created a dashboard for the 2015 Optimizing Self-Response Test, specific to the advertising campaign in the Savannah media market. The CEM project will provide key insights into the success of the digital and micro targeting - targeted advertising efforts to increase self-response rates.

Insights from the test advertising dashboard, will enable us to provide information on the advertising campaign, and will inform key analysis questions that are considered after the test conclusion.

The CEM dashboard allows us to combine the information we are receiving from our digital advertising efforts, traditional media buys and social media outreach, into a single consolidated dashboard.

Because this is updated in near real time we are able to track our progress and see what campaigns are being most affected - effective for the targeted respondents. I'm not sure what just happened with the slides.

The interactive dashboard allows us to - to dive in by specific panel, medium and targeting, to see which campaigns are driving arrivals to the landing page and completions of the online form.

The operational view of the CEM dashboard, provides a near real time view of the actual response patterns throughout the test. By visualizing the data in an interactive manner, we're able to see which demographics are responding to the Census and which are lagging behind.

The information is critical for making real time adjustments to our advertising, communications and outreach programs, to drive response to those - to those geographies and demographic profiles that are lagging.

To set the stage for the 2015 Census test, the next few slides provide background on our efforts to leverage technology and build efficiencies into our field data collection operations. Lisa mentioned the human in the loop or SIMEX testing, earlier in the presentation.

The SIMEX tested the proposed devices, systems and field structures for staff and management processes. The results from the SIMEX have informed the approaches we will utilize and evaluate in the 2015 Census test.

Through the SIMEX we learned that the enhanced operational control system or MOJO is intuitive. Users were able to use the system with only a small amount of upfront training. The smart devices were usable by all people.

Even those with little technology experience were able to adjust, adapt and ultimately, embrace the use of the smart devices. We'll show a video in a moment that provides more information about the SIMEX.

But as a little more context, the elimination of paper and the introduction of automated enumerator assignments, has created a need for a different management structure. With this new organizational structure, we are looking to separate the management of staff from the management of workload.

Looking at this pyramid and working from the bottom up, at the foundation we have our enumerators who will perform the data collection. At the next level we have our local supervisor of operations or LSOs, who will supervise and manage the enumerators and staff resources in the field.

At the next level are the field managers of operations or FMOs, and the area manager of operations, or AMOs, who will work out of the area operation support center or AOSC.

The FMOs will manage operational performance and will supervise the LSOs. Each FMO has a primary geographic area of responsibility but they will work together as a team, to ensure that all work is completed for the entire area - operation support center.

The AMO will manage the AOSC, supervise FMOs and ensure completion of their workload. Oops. I'm sorry.

Managers for admin, recruiting and QA and technology and where they reside, has yet to be determined, and will depend upon automation and centralization of certain functions related to those operations.

Through the automated data collection, the automation of payroll and the increased automation of enumerator assignments, we are able to explore and test revised staffing structures, roles and responsibilities.

For the 2015 Census Test we will test a supervisor - I mean an enumerator to supervisor ratio of 23 to 1, compared to the ratio used in the 2010 Census non-response follow up operation of 8 to 1.

We conducted the SIMEX using a combination of scenarios, real world systems and synthetic data. Sixty-four of 87 SIMEX participants were from regions and had varying experience and job positions.

The simulation provided the ability to alter staffing, roles and systems within a controlled environment. It is giving the Census Bureau the ability to change variables and quickly learn from the new concept of operations.

The SIMEX reduced the risk associated with implementing the major change - changes incorporated within our concept of operations by allowing us to - for issues to arise and be resolved prior to actually being in the production environment for the 2015 Census test.

So now we'd like to share a short video with you about the SIMEX.

Woman:

Welcome to the Census Bureau's ROCKiT simulation experiment. This is what operations will actually look like on the ground, as we manage the two ROCKiT panels for Maricopa County, Arizona. It is extremely different than the way we've done Censuses in the past.

Before we would meet at McDonald's in the morning and they would pass out binders and maps. And we would give out work assignments to the enumerators. We would also meet with them every morning and they would select their G308 payroll forms.

Now what we're - actually giving their assignments to them via the COMPASS device. They'll receive their text message in the morning. They'll open their device. They're already routed the way they should go for the day of all their case assignments.

Man:

We've got our simulated set of enumerators. So whenever they have a problem they're escalating that to the LSO level and sometimes the LSOs are escalating that to our FMOs here in the office, if there's something that they can't handle or they need approval for.

Woman: That one is actually a work in alert. So I will alert in a timely manner. We're looking good here.

Man: We kind of input a lot of real life situations that will come up in the field into the simulation. So these scenarios that are coming up, we think are probably pretty common things that are going to come up in the field. And all this data updates on the fly.

It all - we're not looking at, you know, a day old data; we're not looking at three hour old data; this is happening. This is probably 10 seconds old.

Woman: One thing that's really good for is to track changes in enumerators and their work schedules and availability. This is good data to use to kind of monitor your team's performance and progress.

Woman: This MOJOsystem is amazing. Coming from the 2010 Census as an area manager, I can see how far we've come from paper to automation. I can really see this being a cost savings effort.

Man: These are my LSOs up here. We have alerts that are generated; alerts that are resolved; and alerts that are live.

Man: It's amazing. SIMEX is flawless. I've had previous experience from three decennials in 1990, 2000 and 2010. We had more paper. But with SIMEX and what we call our MOJO system, we are able to connect actions to thoughts.

We can send it out to those local supervisors and they can impart that data on the enumerators and we can resolve things quickly and very flawlessly.

Man: (Unintelligible) has additional resources that we can move over and that can help participation. That was never before. It would take days and possibly weeks.

Man: There's a key paradigm shift. We're no longer expecting people to dig through reports and figure out where the issues are. The system ought to be smart enough to tell them you've got somebody with a lot of short interviews, there's something wrong.

You cannot possibly enumerate somebody in two minutes and get quality data out of them.

Man: We never had an idea of when enumerators were actually out there working or not. So the faster we got this real time data coming in about whether someone has synced their COMPASS device, has picked up the assignments and are actually out there working, is a huge step for us.

This gives us a better feel for whether work is actually getting done in the field and where we need to assign resources in the future if work is not getting done

Man: What if you don't need two people here. What if there was just one?

Man: My only advice - don't set your targets too low. At (UPS) ten years ago, we did a reengineering effort and reduced 85 million miles driven a year. No one would have believed we could have done that. And I think there's more to gain than you think.

You're going to find more benefit and you're going to make bigger changes to this organization than you expect.

Woman: I love it. I love it.

Woman: For anyone in the future, in the role of FMO, I would definitely say let MOJO work for you. Let MOJO guide you on how to do your work. It is the best and most efficient way to get your job done.

((Crosstalk))

- Man: With seeing how well the system's working so far I'm excited to get out there and see what successes we have and see what we can fix as well.
- Woman: Up here we've got a pocket where we're not hitting very well.
- Woman: I did send out a message asking informants...
- Woman: It brings us mass efficiency. It brings us into the 21st century. It gives us that gateway to the future. Let's walk before we run. But we're actually starting to deploy that technology early on.
- Maryann Chapin: Okay. The data collection tool for the 2015 Census Test is the COMPASS. COMPASS is a mobile application that will help enumerators conduct non-response follow up.
- For the upcoming test, enumerators will also be able to submit their time and attendance sheets from their phone, further eliminating paper from the decennial operations.
- The three pictures that are shown, with blurred out data, are what the enumerator will see on their handheld device shown as the entry screen and two data capture screens.
- COMPASS includes seamless integration of all on device functionality required by the enumerators, which includes managing, scheduling, locating and navigating to and between assigned cases, collecting response data and collecting (para) data.
- The goal of reengineering field operations is to use technology to more efficiently and effectively conduct and manage the 2020 Census field work.
- The Census Bureau is developing the operational control system that manages tasks and makes some of the decisions that have typically been made by humans such as case assignments in the timing and number of contact attempts.

This screen shot was pulled from the operational control system or MOJO. I should note that all data shown on this screen are test data. What you see here and hopefully it's clearer on your - your iPads, is the route for one of test enumerators.

The route was generated based on a series of inputs including the enumerator's home location, the location of the addresses for the non-responders and response propensity models taking into account when certain people are most likely to be at home.

Every night MOJO considers all of the staff who are available to work the next week; the location of the units who still require a response; and then optimizes the routes for the enumerators as appropriate. The output of this optimization is a sequenced route for every enumerator.

To explain further what you see in the diagram, the green pen is the enumerator's home location. The orange markers are the units we want to - this enumerator to visit.

The numbers inside each of the orange markers indicate the sequence or the order in which we would like the enumerator to make those visits. And the blue line is the path of travel we expect from this enumerator.

At the end of every night we consider the day's planned work and compare that with the work received from that enumerator. We are checking various data markers and comparing expected performance with actual performance.

When we see areas of concern we provide alert to the supervisory staff who then follow up with their enumerators, to take corrective actions.

Likewise, if our comparisons show that the enumerator is performing as expected, MOJO generates a happy alert where a supervisor will call the enumerator to thank them for a job well done.

Our focus for the 2015 Census Test is on testing innovations and collecting data that will inform the preliminary design decisions for the 2020 Census. In the 2015 Census Test we will test the reengineering of roles, responsibilities and infrastructure for conducting field data collection.

We will test the feasibility of fully utilizing the advantages of planned automation and available real time data to transform the efficiency and effectiveness of our data collection operations.

We will further our exploration of using data that households have already provided to the government and third party data to reduce the non-response follow up workload and increase the overall non-response follow up productivity, through the use of administrative records, field reengineering and adaptive design.

We will test the operational implementation of a bring your own device option for enumerators. And we will conduct focus groups to explore reactions to the contact methods, administrative records use, privacy and confidentiality concerns and how the Census Bureau might react to those concerns.

The 2015 Census Test will be conducted in Maricopa County, Arizona. Maricopa County was selected for three primary reasons - the diversity of its population including a high Hispanic population; the mobility of its population, being a relatively transient population and potentially having some high vacancy rates; and the close proximity of both urban and suburban areas.

Having these characteristics in a test site will allow us to more fully understand the impact of the new methods to assign and manage staff as well as allowing us to test the quality of administrative records with a diverse population.

The Maricopa test includes approximately 165,000 addresses. Approximately 161,000 addresses are in the initial self-response universe with another approximately 4500 addresses set aside for the bring your own device testing.

The bring your own device addresses are not included in the self-response universe. These addresses will receive an advance letter about the possible contact during the bring your own device component of the test.

I'd like to stress that the focus of this test is on non-response follow up and the methods and technologies to reduce the workload and increase the efficiency and effectiveness of conducting non-response.

To that end, the initial universe - an opportunity to self-respond using the internet push strategy, enables us to get to the non-responding universe. To enable our research and our ability to answer research questions, there are three panels in the 2015 Census Test - a control and two experimental panels.

A major objective of the 2015 Census Test is to evaluate the difference in productivity rates between the control panel and the experimental panels. The experimental panel treatments vary the approach by removing cases from the non-response follow up workload prior to field work.

And change contact strategies for cases that are visited in the field. It also varies the way that cases are assigned and managed in the background. Data collection will be automated using the handheld device and the COMPASS application.

For the 2015 Census test, the non-response follow up panels will consist of approximately 60,000 non-responding cases that are relatively evenly split between the three panels - the control panel and the two experimental panels.

The experimental panels are referred to as the full removal panel and the hybrid removal panel, and we'll discuss those in some more detail in the upcoming slides.

The control panel will employ non-response follow up procedures that were very similar to those used in the 2010 Census. The work will be managed out of a traditional local

Census office, located in Maricopa County. Enumerators will be instructed to make no more than six contact attempts.

The first contact attempt must be made in person with no more than two additional in person attempts. The main difference from the 2010 Census is that enumerators will use an automated instrument instead of paper, for the field data collection.

The full administrative records removal panel implements an adaptive design contact strategy and will reduce the initial non-response follow up workload to exclude any addresses identified as vacant or occupied based on administrative records, prior to any contact attempts being made.

All remaining non-response follow up cases, will be visited at least once. The work will be managed out of the area operation support center running out of the Denver regional office and will use the enhanced operational control system or MOJO.

In the hybrid removal panel, the initial non-response follow up workload will be reduced to exclude any addresses identified as vacant using administrative records, prior to any contact attempts being made in the field.

For all remaining addresses, enumerators will make one personal visit attempt. After that initial attempt, any remaining non-response follow up cases will be further reduced to remove any addresses that could be enumerated using administrative records.

The cases that still remain after the administrative records removal, will have at least one additional contact attempt made.

The work of the hybrid removal panel will also be controlled and managed out of the area operation support center in the Denver regional office, using the enhanced operational control system or MOJO. It's not on this one.

In summary, our implementation activities for the 2015 Census tests are well underway. We are currently in the self-response period and are preparing for the non-response phase which will begin on May 14th and which will run through June 23rd.

The last of the planned 2015 tests is the national content test. This is a nationwide test with a Census Day of September 1, 2015, and will include 1.2 million housing units in the sample.

The 2015 Census Test will enable testing of key content areas as follows - race and Hispanic origin questions where our focus will be on several key dimensions that include question format - a separate versus combined question approach.

Response categories - ways to collect, tabulate and tabulate data for respondents of Middle Eastern, North African and Arab heritage.

Wording of instructions and terminology specific to optimizing detailed reporting and improving respondent understanding that more than one group may be selected. Using Web based technology around enhancing question design and optimizing reporting of the detailed racial and ethnic groups.

Tommy Wright: This is Tommy Wright. I could just apologize. We've had a slide out and that's why you're not seeing something here. But it will be fixed.

Maryann Chapin: Specific - specific to the relationship question - is it - we will examine the most effective way to gather information on relationship among household members.

Looking at the 2010 relationship categories and new relationship response categories that will expand the husband or wife and unmarried partner categories. And to distinguish between same sex and opposite sex relationships.

Specific to the within household coverage question, we will examine the most effective way to help respondents provide a more accurate household roster.

Looking at a rule based approach and a question based approach, where respondents not shown residence rule instructions, are asked to create an initial roster and then subsequently it will be guided by additional questions to identify potentially missing people.

In the 2015 national content test, we will continue testing various contact strategies.

Different panels will test the timing of reminders, delivery of the paper questionnaire at various points in the response process, ways to further encourage self-response even after the questionnaire mailings, and the impact of sending a postcard as the first mailing instead of a letter; not sending a mail questionnaire at all and using mail to supplement - or email to supplement postal mail.

At this point I'd like to turn the presentation over to Brian.

Brian McGrath: Good morning. My name's Brian McGrath. I'm the Associate Director for Information Technology and CIO, here at the Census Bureau.

As you've heard from both Lisa and you've seen - witnessed on the screen, 2020 Decennial Census will take on a look, approach and reliance on technology that no other Census that we've ever conducted here at the Bureau.

As a result of that, we as an organization, took a step back and thought, how do we deliver services from a technological perspective, to not only the Decennial Census but the other important surveys and Censuses that we conduct here at the Census Bureau, within the construct of a shared services model and at the enterprise level?

The 2020 will look like nothing we've ever done before. As you've seen in the videos, there's a significant reliance on both technology and our ability to transfer data in a real time format so that on the ground decisions can be made to optimize the activities of our

staff and the performance of our systems as well as the input that we're looking to receive from the American public.

There will be new technologies. Lisa and Maryann both talked about an operational control system that we call MOJO. They also talked at length about a mobile application that was built here at the Census Bureau, called COMPASS.

There are a lot of great lessons learned from the 2010 Decennial Census. As a benefit and privilege of starting here at the Bureau in May of 2009, my first activity was to focus some efforts on the paper based operational control systems and some of the challenges that that's facing.

Looking to mitigate risk - John was kind enough to mention the challenge that one of our sister agencies in states with a deployment of a Web site. I do not care to be in that same framework. But we want to manage the risk of how we deploy important technologies.

And CEDCaP will be a key ingredient for - from a program perspective, how we accomplish those goals. Without question we want to reduce costs.

We believe that by applying a shared services enterprise strategy for the delivery of technology, that we can significantly reduce the costs of IT and IT systems that have been used for Censuses in the past, as well as leveraging the wisdom and experiences from the existing Censuses and surveys and the decennial; marrying those activities together to deliver really high value technology.

Obviously we want to continue on our path of improving our security posture. As we rely on the American public to give us their information, we have to be able to convince them that the data they are providing to us is secure.

IT colleagues around the room can attest, this is an ever growing an ever changing challenge. We recognize that.

But we think CEDCaP again, taking an enterprise approach, allows us to connect all of the various and systems so that we can look at security availability integrity of our systems in a holistic approach. Last but not least is to improve agility.

One of the aspects that I saw when I first arrived here at the Bureau for the 2010 Decennial Census, was a lot of just in time development and delivery of systems. We cannot be in that mode as we move forward with 2020.

Through both government resources and through our strong partnership with the private sector, we - we want to ingest agility into the processes.

And both Lisa and Maryann shared with you how we are iteratively developing defining capabilities; extrapolating requirements; building information technology systems, deploying them and testing them in an interactive and agile manner throughout the decade.

And we've done so already from 2014 into '15 and beyond. Really emanating here at the Bureau, out of the strong efforts in the research and methodology directorate around the Center for Adaptive Design.

And marrying that initiative with our enterprise architecture program here at the Bureau, we took a step back and said, from a capability perspective, what are the systems that we have in production today that support the varying stages of the frames, development, data collection and processing which resides within the scope of CEDCaP.

The grade box on the left shows you the current state. And I won't bore you with all the details. But every blue oblong circle represents an IT system.

As you can imagine, there are significant costs and absence of agility and security concerns around developing and delivering and sustaining the infrastructure in systems necessary to support that model.

What CEDCaP's target architecture or 2B state looks like, is the box on the right of screen where we show a significant reduction in the number of systems, but we don't reduce the capabilities and actually augment the capabilities that will be delivered to our business partners.

What we illustrate on this screen is a graphical depiction if you will, across the top of the survey lifecycle. And then under it, with all the pretty circles and pretty colors, is the mission enabling capabilities that we as an organization need to deliver to support that survey lifecycle.

Again, this is the CEDCaP segment of what we will be delivering in support of not only decennial but the other activities here at the Bureau. So how are we going to do this? Traditionally it's a marriage of people, process and technologies.

The CEDCaP program is governed by an executive steering committee here at the Census Bureau which is Chaired by the Deputy Director, Nancy Potok. And is - has full voting membership of each of the Associate Directors from across the Bureau.

We have a broad - a broad based team with great experience from across the Bureau, that's actually managing the execution of the CEDCaP program. For example, the Chief of the CEDCaP program comes to us from many years of fine work in the economic directorate.

More recently, she worked in the IT organization where she spearheaded the development of our systems development lifecycle. And most recently, was the Application Portfolio Manager for the Census Bureau.

There are other members of the team that come from across the Census Bureau that we have a broad expertise. We also have the benefit of external oversight and participation.

The Department of Commerce, under the direction of the Undersecretary, is leading an effort of the CEDCaP program.

And we are providing input and data for their independent assessment and validation of not only the technical aspects of the program but the programmatic budget and risk management.

We have received a letter of intent from the government accountability office that while they have not instituted a formal audit program as of yet, they do intend to be structuring an audit engagement. So we will be receiving the benefit of their oversight as well.

Moving onto process - we have a disciplined and structured project management methodology here at the Census Bureau. We are implementing a program lifecycle management capability. The CEDCaP along with our partners in decennial, will follow both of those methodologies.

As current state all of the 14 projects that constitute the CEDCaP program will be following that structured methodology for project management.

As I stated earlier, the individual who is leading the CEDCaP initiative was at the forefront of the development of our enterprise systems development lifecycle, which is the key pillar for forming and guiding all of the activities that each of the projects in the program within CEDCaP will follow.

Last but not least, we're deploying an agile methodology for the development of all IT systems where we have content engagement with our customers, receiving their initial capability requirements, fulfilling requirements.

And then routing back to them in a real time manner, the capabilities and systems that we're building for them. As you heard in Maryann's presentation, a number of those systems that were built – MOJO and COMPASS - were done so within that construct.

And they are in production and being used for the - the 2015 Tests. Last but not least, the foundation for all of this is industry best practices technology.

We're using a suite of proven technologies today that we have experience with both here at the Bureau, but are also state of the market across the best practices in the private sector. We know and have the challenge that the technology that is available today will change. It will improve.

And we're making risk based decisions as to how we will adopt and implement new and emerging technologies in the coming years.

Last and not least we will be engaging with Carnegie Mellon who will be establishing an industry panel of technological experts who will be providing us technological feedback on the architecture and underlying infrastructure in systems and development capabilities that we'll be engineering for the CEDCaP program.

On the next slide what you have is a graphical depiction of the timeline and the capabilities that we're building for the CEDCaP program and how they marry to the Decennial Census, the economic Census and the other key initiatives here at the Census Bureau.

Along the left hand column you'll see, what in essence are the capabilities that we're delivering through the 14 projects, and the timeline through which we're delivering them. Focus for the CEDCaP initiative is in fact a Decennial Census.

And again as you heard, many of those capabilities have already been built. Some are in a prototype phase but we're using them in a production format to deliver and test how they will respond in a real world setting. And I'll turn it over to Lisa.

Lisa Blumerman: My term again. Everyone take a deep breath because that was a lot. Right? I know I like to take a deep breath when I get to that point because it makes me think about everything that we've done but also think about everything we have yet to do, which is what I'm going to talk about now, are those things that while we're busy doing all of the tests, all

of those other things that we're doing to really ensure that we are ready for the 2020 Census.

So the first thing I wanted to spend some time with you all on, is the 2020 Census Operational Plan. This is the key decision document that we will be releasing the end of this year.

And what you have in front of you on the slide, is a graphical representation of what this plan really will look like and what it will encompass. The concept of operations - the 2020 Census Operational Plan, will describe how the 2020 program will operate to allow us to execute the Decennial Census.

It will describe the operational intent and the context and the key decisions that have been made and will - and still need to be made, around the 2020 Census. Within the Operational Plan, the - includes a number of key pieces.

It includes the concept of operations which is the real key focus I think, coming - what most people are looking for although the other pieces are just as critical.

The concept of operations itself, will have a heavy focus on the four key design areas - the innovation areas - reengineering; address canvassing; optimizing self-response; utilizing administrative records; an reengineering field operations.

Additionally it will include information that we - the information that we know today, regarding the rest of the 2020 Census operational areas.

The narrative for both the key design areas and the remaining operational areas, will have varying degrees of maturity based on the research and implementation work that we've completed to date.

This is a really important point when we think about what the Operational Plan is and where we need to go. As I said earlier we're releasing this document three years prior to the release of the 2010 Census Operational Plan.

I also should point out that the 2010 Census Operational Plan wasn't finalized until after the 2010 Census. And this will remain true for the 2020 Operational Plan.

One of the key points about this is that it's a living document, meant to continue to change and to be expanded upon as our decisions are made, as we refine our operations, as we execute those operations, so that we can document what it is we actually did do.

Because we all know that in the heat of the Census we change some things and some things occur and we work on it, we fix it and we do the best that we can do. But we want to make sure that we capture all of that information.

It is important to note also, that the operation plan itself, the body of the document, is not the - it's not the execution it's not the implementation it's the design right, it's the design of the census.

The execution or the implementation plans of the operations for the census will be added to it as they are finalized and developed as attachments to the plan.

So for example for some of our early operations such as address canvassing and LUCA we are currently now working on draft implementation plans or execution plans.

That's the details of how the census will be executed how those operations will be executed. The Operational Plan itself is at a higher level and it talks about the design of the census and encompasses all 34 operations of the census.

Additional things - I lost it okay is it back? Maybe, no I can be really loud though. I can be really loud okay, so and maybe it will come back okay.

In addition to the documentation we'll also include the documentation will also include a description of all the proposed operations. So as I said this will include the 34 operations.

Thirty four operations for those of you that know the details of previous censuses that's a change from the last census. I believe in the 2010 census we had 44 operations and I honestly don't know the number of operations that we had in the census 2000.

But some of the things that we're looking at and some of the things that we're testing are reconfiguration of operations. So for example, we're looking at how in the course of our innovation area for our field reengineering field and utilizing administrative records at how we could potentially combine our non-response followup operation with our field verification operation.

In the past we had two separate operations that were doing very similar - it will come back. That were doing very similar things with the use of MOJO and the technology we have we really think that we can combine those two into one operation using the enhanced operational control system to be able to tell an enumerator whether they're doing field operation or whether they're conducting a census enumeration at that particular address and that's some of what we'll be testing in the years to come.

We'll also include the key decisions that have been made and those that still need to be made. Just as with everything that we work on there are some key decisions that drive our methodology that we need to know now.

We want to have this course on them we want to have - there we go. We want to have this course on them, we want to have a clear understanding internally and externally about the decisions that were made.

There are other execution related decisions that we have more time to finalize. And so it will be clearly articulated what decisions have been made and what decisions need to be made and when those decisions need to be made.

We'll also have a discussion of the research that's been completed to date. We've done a lot of tests we've had the 13 test in the Philadelphia census test in the Philadelphia area.

We had the 2014 - sorry I'm so sorry. We had the 2014 census test in parts of DC and Montgomery County and now we have 5 tests in 15 - maybe if I hold this. We have 5 tests in 15 that we are working on now.

Good, Brian tells me it's not me I'm not special it's just the mike. Good to know. Okay, it also will include a discussion of the cost and quality trade off and if you're interested in more information about how we're planning to do cost and quality trade off that isn't something we're going to go into a lot of detail here on.

But at the recent program management review that we had last week (Pat Cantwell) actually did a presentation where he walked us through the tool that we're using as well as our approach.

And while we're happy to answer questions I really would encourage you if you didn't see it to go online because there's a presentation and it's very nicely, he very nicely articulated that.

And of course we'll also have a discussion about the IT capabilities that are needed to support the operational implementation of the 2020 census.

So I wanted to talk briefly and I think this is probably my last blast slide about what's next for the 2020 census. Following the finalization of the design decision our operational we are quickly pivoting and we're doing it already in all fairness.

In fiscal year 16 to really focusing on the operational design, the development and the systems testing for production for the 2020 census. Our efforts in 16 will allow for the development and further integration of our production systems with our operations that we need the operational implementation that will allow us to have a successful census in 2020.

The development of our systems and our interoperable production systems must be largely completed in fiscal year 16 and 17 so that we can conduct the end-to-end test in 18 and allow us time to make continued revisions in 19 before the census is largely under way in 2020.

Other things that we're doing in 2016, we're currently working on our plans for the 2016 census test, we're working on site selection now and that's some of what we're working through.

Some of the information we're looking at for site selection would be things like an urban area with language diversity. You may remember from previous presentations there were some discussion on the budget but one of the areas that we slowed down our efforts on and descoped a little bit in 15 as a result of the budget was our effort on the language program.

The language is a very strong criteria for us for site selection in 16. We know what to do with this but we also really want to make sure that we can get out and really test some of the innovations in those areas.

We're also looking for areas with high vacancy rates as (Marianne) talked about. One of the key components for utilizing administrative records is for us to be able to potentially remove vacant housing units before starting non-response followup.

So we need to find some sites that have large vacancy rates to allow us to prove in that methodology. Again we'll continue to look for address variation that combination of urban and rural addresses so that we can continue to refine our processes in general but also for non-ID processing.

And to me the critical objectives for this test will be allowing us to really finalize soup to nuts are proving into the methodology for non-response followup.

We're doing a lot of testing in 15 and some ways the testing that we're doing in the 15 tests are really looking at the extremes with the full removal versus the hybrid removal.

There's a lot of things we can do along that scale and in 16 we'll be putting the best of those out into the field so that we can then assess what that methodology should look like.

We'll also be continuing to work on our bring your own device as well as the research for our targeted communication efforts, partnerships and language support.

In 16 in addition to the census test with an April 1 Census Day we're also looking at a test in the fall with the September 1 Census Day and this would be an address canvassing test.

Again this would be a nationwide test or that would be a difference from this it's again for me but this would be a nationwide test where we again are looking at the accuracy of the maps.

But at this point we'll be refining and taking that draft implementation plan that we're putting together now on what address canvassing would look like and putting it out into the field to see what the execution of that would look like.

Looking for a number of things here but it will be nationwide likely with over sampling and a couple of particular areas so that that over sampling can then lead into our plans for 2017.

So what we do is in those communities that we over sampled we would then use those as the sites for our census test so that we could see the integration of the operations from address canvassing and the impacts of the changes to the designs of our methodology on self-response and non-response followup and that's very important to us.

Not only do we need to look at the interaction and the integration of our systems but with all of the changes that we're putting in place for the methodology of the census we also need to look at the integration of the operation and that's very important and the methodologies and that's very important to us.

In 2017 we are again looking toward at least one April 1 Census Day Test. Again this test would now be focused around the integration of operations, bringing more tools, more technology online.

Every test we're building upon the last so we're adding our production systems in so that we know we're ready to go for our end-to-end test. And then in 20 we'd also in 2017 be bringing in some testing of our group quarters populations and special populations, which while we've done some research on starting now we really haven't done a lot of field testing on. So we'll be looking to focus on that in 17.

All leading to 2018 which would be our end-to-end test and then in 2019 we'll tweak some things, fix some things and be ready also to launch the infield address canvassing what's remaining for us to do and ready for the 2020 census.

While I focused through this discussion on testing I also should point out that there are some key core programmatic work and core programmatic decisions that need to occur along the way.

So for example in 2016 we will be, we anticipate awarding the communications and partnership contract. We also anticipate in 2016 awarding our questionnaire assistance contract.

This is previously what we called we're not calling it census questionnaire assistance because we recognized that with the use of the Internet and the use of modern technology we're not just looking at communication tools via the telephone.

But we also need to expand that to include Web things like Web chat and other potential mediums to help encourage and provide people with the answers to the questions that they've been asked. And we've learned a lot both from the 14 tests about the kinds of questions that people ask as well as the 15 tests that are underway now.

In 17 we will be delivering the topics to Congress and LUCA gets underway in 17 and then in 18 we will be delivering the questions. There's a lot more that we are doing within that but those are just a few of the highlights to give you a sense of what's to come from today through 2020.

And with that we are amazingly on time. So I will pause and see what you would like to do Tommy.

Tommy Wright: Thank you very, very much Lisa, Maryann and Brian and because we are on time let's continue to be on time and take a break. Let's return at 11:00 thank you very much.

Part 2 starting @ 11:15/56:36

Guillermina (Willie) Jasso: Thanks. More questions. Yes.

Sunshine Hillygus: Sunshine Hillygus. And actually this is a perfect follow up because while your question is focused on evaluating the effectiveness of operations I had essentially the same reaction in thinking about the campaign the communication campaign. And that is, is that, you know, while it sounds like there is so much progress that's being made what I didn't hear and what was a little worrisome to me is that comments were made about evaluating effectiveness but I didn't actually see the design that would allow for that.

And in particular where I was worried was there were comments like, you know, we will be able to adapt and target those groups that are under performing, which to me sounds like Savannah will be able to generate great experience but I'm not sure if the research

design is embedded to actually be able to evaluate the effectiveness of these different parts of the communication campaign.

And so what comes to mind for instance is that, you know, TV ads for this are incredibly inefficient these days in trying to reach low propensity responders and so then it becomes a question of, you know, is that worth the dollar that you're spending on them.

Social media doesn't have the same boundaries as, doesn't have media market boundaries and so, you know, how has the test been set up to account for that. And so I know these are all details that you couldn't necessarily get into but I would love to be able to see some of those details to make sure, to see if I can help in any way in making sure everything is on track for being able to make some of those statements.

Lisa Blumerman: Sunshine, thank you we would welcome your help. First to reassure you yes we do have those details and we do have a research plan that articulates it. And I can give some examples of some things that we're seeing and some things that we're doing in response to it.

Maryann started to hint at it a little bit by showing you what we can refer to as the CEM the customer what is it the engagement, its Customer Experience Management tool that we have.

That tool allows us to drill down into the various components of both the communication campaign, the social media as well as the print advertising and begin to assess the effectiveness of it.

But with every communication, you know this but with every communication and partnership campaign it is hard to do this. But looking at the data, looking at the response rates as they're coming in, looking at where those are responsibilities are coming in in combination with the different contact strategies that we've employed for those addresses.

We actually are making some changes. For example early on when we were in the notify needs the early engagement technique which was the month before we opened self-response, we had some advertisements that were on social media that we replaced those advertisements.

We just found we weren't getting the response that we were looking for. So those were pulled and different ones were then released in response to it. During the self-response phase and we're about 2, 2-1/2 weeks in, looking at it now and then looking at the self-response data that are coming in we've decided to increase some partnership activities that we have going on there.

As well as to increase some of the advertising that we're doing particularly for a couple of the demographic groups. So we are able to look at it. But this is very hard and it's always hard to decompose the impact from a partnership event or a communication in advertising from what we see in terms of what comes in.

But compared with previous emphasis I can say and previous survey's that I've worked on I can say we have more data today than what we've had before and that's allowing us to make informed decisions as we're going forward but we would welcome your help as we look at this that would be great.

Guillermina (Willie) Jasso: Yes.

Jack Levis: Hi this is Jack Levis. First, we've had many meetings and I want to start with I'm extremely impressed with the work you've done. I'm glad you mentioned that you're closer to the 2020 census than 2010 because like the mirror in your car about objects dates are closer than they appear.

So I'm glad you have that on the forefront. We've seen and I think you recognize a lot of similarities and parallels in what you're trying to do and what UPS has done what I've been working on for 15 years.

And again I've got zero criticism of things I've seen but I do want to point out a few things that I know from my experience what was hard for me to do if you can keep that in your forefront so you don't have a blind spot there.

I know you're counting on automation of the sites and I've counted on that too. But from my experience I spend less time thinking about the happy path, which is the automation and more time thinking about what I do in the unhappy path.

So I generally like to design my systems where it works without the automation and then the automation is a help and if you focus on the non-automation component and say I know that works, the automation can only help. So please be careful about assuming that automation will work. The unhappy path can really hurt you.

Similarly and we even said it again the optimization is going to be harder than you think is my gut. I, we've been in the news a lot lately we announced our latest optimization of a savings of \$400 million a year and we've announced that.

Keep in mind we deployed it now and it looks great but it took us five years. So what we thought was an off the shelf purpose it took us five years to get us there.

I just came back from a very large optimization conference and as I told (Stephanie) last night I'm very willing to help you evaluate those off the shelf components but the optimization is going to be harder than you think.

And if you noted and one thing to keep in mind with that, you have to know what the rules are before you say I'm going to optimize it. So, you know, when we used to have to remember we were looking on the board when we came here for SIMEX you have to first design what are my business rules, how does this thing work and then let the optimization take over.

Don't start with the optimization and throw your rules in. I don't want you to wait five years as I did.

As I said in that video I struggle with this about setting sights to low. I was looking at your things on the big board there and I was looking why is it doing this, why is it doing that because I think you got a lot more to gain.

So you got to make sure you get it don't make it just a hurdle this is all I'd need to do to get the 5 billion. I think you got a lot more to gain and if we do all this right I think you're going to hit a huge, huge homerun with that.

Finally, I know you're working on it but you need to manage risks, you know, if something can go wrong, you know, if you have a 50-50 chance that something is going to go wrong it will about 90% of the time.

So I'm glad you're looking ahead. You need to manage the risks and a couple there is making sure the stress and performance side is covered and I know you guys worry about that.

It's one thing in the size of the SIMEX it's a whole other ballgame and, you know, I live with that when we scale things up and all of a sudden you can't get it through.

I know we're going to do the things in the field so don't forget the telecom side of it. I've seen way too often where everything looks okay but I can't get the communications up through that thin pipe somewhere from the field through the telecom network and you're counting on real time.

So that was about everything I had - sorry, from my notes.

Lisa Blumerman: Thank you so much.

Guillermina (Willie) Jasso: More discussion on express - yes Jack Dangermond.

Jack Dangermond: Yes just again but I've already said it I'd like to add to the previous speakers in acknowledging you for taking what is a very difficult situation. I mean I don't know what happened in the last six months but it's transformational that's the way I would describe it.

It's really great and reinforcing a couple points that I think Jack you brought up in the last meeting was let business be the driver and IT supportive and a good relationship between those two teams and that's what I saw back and forth in the iteration here.

Second, the distinction between designing the com ops and designing the implementation. And in a way the com ops done up front and then the implementation iterating through prototypes is just like makes my heart patter because you just don't see it very often that people are mature enough to make that distinction they usually throw them all into one.

And then Brian, reinforcing your sort of seven rules the last one of which was, we don't relinquish that the CIO has to be the ultimate integrator in other words do it in house not out house.

And every instance that we said I think in the last meeting when it was outsourced it flopped by a large enterprise system. This is really, this requires leadership.

My last comment is I hope the three of you are going to stay employed here for the next five years because this is not a joke. It's I think talking about risk and Jack brought up the issue of risk.

The biggest risk is that you guys get distracted or, you know, so this is really, really, really important. You should bond with each other and really don't, you know, whatever that means. Don't get into trouble.

Guillermina (Willie) Jasso: Barbara Buttenfield.

Barbara Buttenfield: Let me reiterate - wow, wow I have three comments and I think that these are comments that you didn't have time to present but I'll put them on the record as far as my voice will carry.

The first one has to do with the previous meeting where we recommended that you bring in experts in the middle of sensing and pattern recognition to help you figure out the automation of change detection.

And as you reengineer that address canvassing and I didn't hear you talking about that so I wondered if you could comment on that.

The second has to do with something a comment I made before. The issue of all of these dashboards and systems and trying to get the real time is a huge step. Jack just said it was transformation and he's right.

It's important that you pay attention or that someone on your team is paying attention to usability especially if you are trying to work with this data in real time.

So I would like to hear what you're doing to work on usability and I would also like to I guess I'm suggesting this not to Willie but to Barbara to consider if there's another opening on the CSAC that we bring somebody in who works with real time usability evaluation just a heads up to help you folks.

I've done some work with usability on Web sites but nothing with real time and it's a different, horse with a different color. But related to this you are expecting your colleagues here at the census to be working with data in real time.

And so one thing I do know is that our eyes and our brains work really, really fast and the more that you can develop innovative tools for visualization the more efficient your staff is going to be at working with this data.

And the visualization can be in the form of mapping but it would also be in the form of exploratory data analysis. And so I wonder if you could talk about that so the address canvassing and the pattern recognition and change detection, the usability and the visibility.

Lisa Blumerman: So where to begin. Those are great questions, great comments thank you. Let me start with address canvassing and hopefully this will address it but I'm not sure how much detail I can go into right now on that.

We completely agree with the recommendation that we do need to bring in external experts. Both individuals to help us as well as partnership with the private sector on what tools may be out there or what services may be out there that we can utilize as we're moving forward.

On the field as you know the field of change detection is very new and it's moving very rapidly. And as Brian talked with some of the other technologies we recognize that where we are today, which is a little bit of an interactive mode we won't be there tomorrow it will be much more automated and the tools are just growing.

So we've partnered with several others to just statistical agencies on that and I'm looking at what they're doing and on the development of tools around that. So we have experts from the geographic community that we're working with.

In addition to that we also work closely with the standing committee of the National Academy of Sciences that has several geographic experts that are working with us day-to-day on what we're doing.

In fact Don Cook was here I think my weeks are a blur right now it might have been earlier this week or it might have been last week sitting side-by-side with Mike Ratcliff and our geography division really working hand-in-hand on what we're doing and he's one of the members of our standing committee. So we have a number of experts that area working with us on that.

The third area is and I glossed over it and I can't provide a lot of details at this point but certainly at a later meeting we'll be able to provide more information.

Over the last few months we've issued two RFI's around items related to address canvassing. So the first was about data availability, addresses, road features, aerial imagery those kinds of things.

The second RFI that we put out was about services. So things like change detection, things like availability of updated data. From those two RFI's we've already issued one RFP, which I can't talk about.

We are analyzing the information and moving forward with that. And we have a second RFP around the notion of the services and change detection and things like that coming out the end of this month.

That was only one of your comments, wait there were two sorry. Usability, you know, you brought up fabulous points it is hard. What I wanted to talk about on both maybe I'll tie the together and just briefly and then I know there's lots of other comments so maybe we take some others.

I can talk about some of the things that (Stephanie) has built in to help us with the analysis of data real time and it's from and these are tools that are baked into MOJO our system.

So what the screenshot that Maryann showed you was the route. What she didn't show you were the screens that our supervisors have available to them. What we don't want them to do is to have to dig into each and every individual record.

What we want to have them do is to manage what we're calling alert where we have parameters that we've established, reports that we've developed that will ping our managers when things seem array.

So for example the kind of an alert a manager will get in real time. At the end of the day an employee and numerator may, they'll have to put in their timesheet their payroll request.

And they may say they traveled 6.7 miles and worked eight hours. And we know from the route we had provided them in the case management workload that maybe that eight hours should we estimate it, it should be four.

They'll get a, the supervisor will get an alert that gives them this information as well as a report that they can then look at to determine if there is an issue do they need to have a followup call.

There's lots of reasons that the time might take longer but it's a probe, it's a prompt for them to have that conversation. You saw in the SIMEX video Ali was talking about the length of time in interview tapes.

That's another probe or an alert that a supervisor will get if we have an interviewer that's generally completing interviews and a very small amount of time.

They'll get a prompt, they'll also get an alert if someone is taking way too long based on what we expect an average interview to take. And those tools are baked, built right in, into the system.

So that's one area but you're right that's only related to the use of MOJO and the operational control system and what we're looking for field reengineering.

There are other areas that we need to look into this as well. And so this is an area we do need to do more work but I think we've got a good foundation to move forward from.

Ken Simonson: I share my colleague's admiration for the presentation and all the moving pieces that you seem to be handling well. Two comments, one, I hope that you'll be able to capture information from the field, from the enumerators about buildings that aren't on the list or ones that have disappeared and so forth.

Going back to the 202014 Test are you incorporating that information as you go forward with 2015 or do you need another year because you are capturing lessons learned and what are some of those if you can talk about it briefly.

Lisa Blumerman: Sure, so to that first point which is are we capturing information from the field and incorporating it in. That is something that we are planning to do. I'm going to segue into the ACS for a second and then I'll come back to 2020 because I'm really pleased with some of the progress that they've made on this particular issue.

Over the last year we've actually been able to integrate some of the findings from the ACS enumerators back directly and adjust them into the master (unintelligible) and that was a real, that was the thing that we've been working on for a long time and that process is now in place so that's helping us move forward.

For 2020 we're looking at a couple of different things, how we identify that information and how we feed it back. So through address canvassing for example and we didn't talk a lot about it but the tests that we conducted earlier this year the address validation test, one component of that test was what we looked at with a partial block canvas.

But the staff that went out to conduct the partial block canvas were using our new listing device, which is part of our CEDCaP program so it was the first time we had that device in the field.

One of the findings they had from the use of this for the first time was that we needed more fields to capture exactly that information so that we could integrate it back in.

So that actually is one of our lessons learned that we need to - and we're already making those changes so we'll have it for the future. So I think we're working toward that.

Also in our 16 tests but not in the 15 tests but in the 16 tests we're starting to look at the integration of a couple of the operations that will also help with that at least that's part of the plan today.

The second component of your question was what did we learn in 14 and how are we building that into 15. So, you know, I can go back to 13 and talk probably for the rest of the day about all of the different findings that we've learned from the 13 tests, the 14 tests and the 15, leading into 15 and we've already learned some things in 15.

A couple of key things that we learned from 14 that we either incorporated or didn't incorporate or we incorporated in a different way. The 2014 Test involved a panel that used email as a primary contact method. That really did not work.

So we didn't incorporate that in 15 in fact I think it was 60,000 email addresses that we utilized for initial notifications for the first and second mailing in the 14 tests and of those about half of them bounced back.

And then we had an additional problem because everything was moving so quickly that by the time we sent a paper notification that paper notification said you've received two contacts from us why haven't you responded.

People were calling us saying but I didn't receive anything because we had contacted them by emails that they never received. So we actually removed that as an initial contact strategy.

And the 15 tests I think it's the 15 test that we're using email as a supplement. Is that in 15 or 16? Then I believe it's in the 15 test that one of the panels the contact strategy panels is adding email as a later supplement to the mailing pieces.

So we're not using that as a primary but potentially a supplemental way to reach people. So that's one lesson. Another lesson from the 14 tests that we did experience was we had some challenges with our telephone questionnaire assistance and the call volume and the type of call that we were receiving.

So this was really important to learn. For example some of the things we were seeing and once we did a little more exploration we learned that this was pretty true for any Internet data collection was that we're getting a lot more calls now using an Internet data collection tool than we had from just this a mail out mail back and those questions are very different.

So the questions are we started the questions about is this a legitimate survey, why do I need to participate can you help me. But we're also now getting my browser doesn't seem to be working and I don't know how to make it work or I clicked the link and the link, you know.

So more technical questions that were not in a position or space to handle because we don't know what their device is. And so we're working on introducing additional materials around that. We incorporated that into the 2014 Test I mean to the 15 test, sorry.

We also had a problem in the 2014 Test with our estimate of call volume. We didn't estimate well what that call volume should be. So for the 15 test we introduced some demand models to help us predict it.

The demand models have worked better they're still not perfect as we found in the 15 tests we had about a day blip that was a challenge for us so far and we've made some adjustments to the model and are moving forward with 16 now. So those will continue to improve over time but those are just a few handful.

Irma Elo:

Irma Elo.

Brian McGrath: Yes could I offer one other hand back to you. In the 14 tests with the handheld device we had layered some security controls on top of the device that challenged some of our users to use the device.

And the team the comfort team did a great job. Lisa and I got together quickly and made some decisions with the guidance and assistance of Tim Ruland and our chief information security officer.

And we were able to turn another of the software around with a re-rack of the security controls so we didn't diminish the security of the device but we made it easier for people to use. So lesson learned that we implemented right away.

Irma Elo: That sort of - this is Irma Elo. That sort of gets back to my question. I think, you know, everybody agrees you've done an amazing job moving forward and coming up with systems and I hear a lot about discussion of what's happening in the census and then we saw that center was information was coming really fast.

But what I haven't heard any discussion of how is this going to filter down to your (AMO), (MO's) and (LSO's) and it seems to me like the technology is going to be very different than what at least the past interviewers are used to.

And people are resistant for change they also may not have the skills to implement using these systems or adjusting. So what are you doing because you're going to have to have a huge workforce by the time you implement the census.

So what, I haven't heard you much, discuss much about what are you doing assessing the capacity at the local level and the medium level of some implement policies?

Lisa Blumerman: So that's a - did you want to add onto that or? No okay. So that's a great question and there's lots that we're doing there. I'll talk about it at a high level and we'll see if we've covered it and Stephanie Studds certainly can add more to it.

First to talk about the handheld device though. That was one of our risks, one of our concerns with the 2014 Test and that was why we did go out with the 2014 Test with the use of the app on the handheld device.

Our enumerators were generally pleased with it. We did not experience the challenges or problems or the risks that we thought we would have with it. It was very smooth, they were very, they were able to work with it very easily and it was a success from that perspective from the device perspective.

From the kind of workforce perspective at the supervisory level the use of the operational control system, the use of managing alerts. So, you know, we've worked really hard to work with our field directorate and the staff at all levels of the organization throughout the development process over the last year.

So our ROCKiT team lead by Stephanie Studds, one of the things that they did is they actively reached out and went to each region as they were developing to solicit the ideas from the staff that would need to work with the tools as they built it.

And as they built each iteration they then took them back out for them to work with them and to engage with them and to see if it met their needs. And that's where a lot of the ideas for the improvements and the innovations came in.

When the staff, the enumerators, the supervisors they said but I need this and we were able to either make, build it in right then or build it in in a later iteration.

Also the SIMEX which you saw the video for that involved about 85 actual people from the region. And everything we heard, which is all qualitative about this piece is as soon as the tablets which is what they were managing from arrived in the region there was a clamoring for people to get their hands on them.

So they're really excited about it. So some of the kind of resistance to change I think what we're seeing and what we're hearing is they're ready for it, you know, every not

everyone but lots of people are using a smart phone, they're ready for it in their day-to-day work.

And we are doing focus groups as well so we're having continual discussions to get a more systematic feedback but in general it's going well. I don't know Stephanie if that was sufficient if there's anything more you would - it's okay? Okay.

Roberto Rigobon: I was going to say something else but let me react to what you are saying. I think it would be good to have a plan B in case you have the resistance. Now I mean this, I understand I think Irma's point is very, very good.

You're dealing with, you know, half a million of way more a million in New York, everybody will comply is not clear and success depends on adoption.

So coming up plan B for those would be advisable just more reaction. Let me my reaction was to the previous comment on the demand, you know, a good part of me is that I am an engineer the best part is that I'm a professor in economics that tells me how good of an engineer I used to be.

So yes I was a musician first and then an engineer I'm then economy. So it tells you a little bit about my (unintelligible). So but my point is let me help you with demand information so you - if this can be shared maybe a difference (unintelligible) can help you to stimulate a more robust information manner would be more than glad to help you if that helps you the information. So that was just an offer.

((Crosstalk))

Peter Glynn: Peter Glynn, just as with my colleagues I'm very impressed with the path that the census is on. So I guess I just wanted to raise one issue that has certainly been an important one at Stanford over the last year or so, which has to do with data security issues.

And you're talking about extensive use of the cloud you also mentioned a bunch of times bring your own device. And those particularly bring your own device comes with all kinds of potential security issues. I'm just wondering how you're addressing it.

Brian McGrath: So all federal civilian federal agencies are required to adhere to the NIST, National Institute of Standard of Technology publication for how to secure information system.

We obviously follow all of those guidelines and rules and regulations. We've got a couple other things that I think are somewhat unique here at the Census Bureau that Tim Rulands team really gets credit for.

When we start a project an IT project now Tim embeds the security engineers into the process from day one. So this is not a paper pusher security person this is a skilled engineer in software engineering and network architecture that is on the ground part of the team from day one.

So there's a constant feedback loop of and an assessment of what we're building, how we're building it and what are the security risks and vulnerabilities that are adherent to the particular strategy or technologies that we need to build the solution.

That has paid significant dividends for us. As far as the cloud we're watching now frankly. A lot of my counterparts around the federal space we see a lot of what people are doing in the cloud and there are infinitely more people talking about what they're doing in the cloud than people that are actually doing something in the cloud.

So our first strategy into cloud architecture was to build a private cloud in our own data center. And we learned an awful lot about what that means and what and how the infrastructure in a cloud architecture interrelates with the application.

So I think we're pretty well positioned that whatever the cloud solution looks like I frankly don't think we're ready from a societal perspective for the census bureau director

to come out and say we're going to collect all your data and we're going to put in the public file.

I just I don't think the American public is ready for that. That being said there's other flavors of the cloud that I think we can leverage and we're doing some really interesting testing now.

The question about scalability is obviously one we're concerned about because we're going to go from stop to go really fast and then probably back to stop again.

Clearly I don't want to be the business of buying all that infrastructure and trying to provision it in my own data center or my own series of data center. Probably even if I started now I wouldn't be done in time for 2020.

And from a cost perspective it's just an unsustainable business. So we're watching we meet with our counterparts at NIST on a regular basis their CIO and myself are pretty good buds and we talk a lot.

The defense community has been very willing to talk to us about what they're doing and they have some significant challenges in that space as well and I don't mean to minimize the data that we capture but, you know, in their world if somebody makes a mistake people die.

So they're hyper sensitive and they've been great. The folks at DISA and just fantastic about inviting us out to the lab and showing us what they're doing.

We've had them out to our lab and we've showed them some work that we're doing around the concept of derived credentials if you back up.

So and like I said we're doing some really interesting testing now, you know, in a private cloud that's not ours and the next step which I think will happen week after next we'll

actually add from this private cloud into the public cloud only for the compute layer but the data will still be retained in the private cloud.

So, you know, you guys know that this is a space that's just changing rapidly and things that look impossible today are going to be commonplace in six months.

So we fight this battle, you know, people want us to say well what's the Internet solution going to look for 2010. Well I don't really want to tell you right now because I'd like to take advantage of these great technological leaps that are going to occur in the next couple years.

So we're managing it through a risk management framework but it's a little bit challenging. I hope I answered your question.

Guillermina (Willie) Jasso: More discussion, comments, questions. Yes... Sunshine Hillygus.

Sunshine Hillygus: Sunshine Hillygus and I think you've kind of addressed some of this already and one of my questions having looked at the adaptive design decisions and process through 2013 and 2014 was a little bit more about both the lessons learned but also, you know, a little more detail on that.

And again I know you couldn't cover it all and I wondered if there was a more detailed document. And in particular I mean it sounds like ultimately in terms of an adaptive design right it's pretty narrow right.

And so we're talking about vacancies and if you get contacted but not necessarily using information collected during initial non-response calls. And that's the part that I wasn't sure for instance what type of para data were being collected on that first contact.

Even if that doesn't go into the, you know, stopping rules having that information for later evaluation could be valuable.

Lisa Blumerman: So you're right the points you've raised are all good. In terms of sort of where we've gone in terms of adaptive design from 13 to 14 to 15 just really over simplifying it.

You know, some of what we learned in the 13 test was the propensity models that we had built didn't work as well we'd like. So in the 14 tests we didn't use propensity models.

In the 15 tests we're using something similar so we kind of skipped that test and have modified it a little bit. We've had the same kind of iteration in our thinking around stopping rules and models and how we intend to address it.

Really in terms of a variable contact strategy is the way I, keeping it simple the way I like to think about it which is that for a particular housing unit looking at some sort of average number of contact attempts across an area based on its propensity the area's propensity to respond so that we have high quality data.

What would that contact strategy be within it? Would the housing unit receive, you know, one, two, three, six types of visits. And so we've done a lot of work around that and a lot of research on that.

Tom and I saw him earlier I can't turn around right now to see if he's still here but his team has really led that effort and we can certainly kind of walk you through where we are.

And also we know what we're testing in 15 but we also already have ideas for where we think we need to go for 16 with it. We need some of the data. In terms of data coming in in real time and feeding the model.

Some of that at least in terms of our tests it's rather fixed right now we're using information that we already have, information from the ACS to help determine what we want to do.

Certainly there's some information feeding in around the optimization and the use of MOJO and how we're routing cases but from the design strategy right now it's fairly fixed.

I don't know if that helps and there's a lot more detail.

Guillermina (Willie) Jasso: Jack Dangermond.

Jack Dangermond: Jack Dangermond, yes this is along the lines of what Jack was getting at is technology change and thinking big. I have just two thoughts. The first one is drone.

We're seeing an enormous uptake not so much on the U.S. but outside the U.S. for the concept of persistent surveillance. And you think well that would be wild to instead of sending out a person to check you just send out a drone. How much would that cost?

Well not much if you had 1000 drones or 100 drones for the major metropolitan areas it would just zoom on in and check out a situation night, day whatever and give you back information.

All the forecasts that we are sort of tooling up for is persistent surveillance with drone information like you have with satellite information. And a lot of our customers like utilities are actually tooling up to fly their lines, fly their pipelines with drone information to get much more up to date data.

So just put that in the back of your mind. The second thing occurs to me is what we went through with Ebola. With Ebola starting last fall we began to look at using devices for field survey tools that would capture the death or disease incident in the field.

And West Africa it was deemed to the cloud and looked at in real time looking at heat maps trending patterns and overlays of those against demographics in West Africa and Geneva using a kind of dashboard.

So this like Jack said we can't imagine how this instrument that you are inventing will affect the science of both initially data exploration and then its application.

Well I'm wandering a little bit here in my conversation but it seems to be that it would be enormously interesting that as the data is being collected if you're really collecting the entire thing on a device kind of like the survey one, two, three notion some of you have seen.

That that actually can be examined both from the research perspective as it's being collected even if it's internal to the bureau so that like in 2016 and 2017 wrapping up we're actually getting real time observational data that's leading to radical shift in adaptive design of the methods and followup.

I think technologically what I'm saying is that that's all going to be available to you and not 20 years or 10 years but like next year as a commonplace thing. I'd just like your thoughts on the notion.

Lisa Blumerman: Sure I think those are great points. Drones kinds of worries me a little bit but...

Jack Dangermond: Yes, Congress will shut down I know.

Lisa Blumerman: ...I don't know we, you know, gyro copter yesterday drones I don't know. It worries me a little bit but I think the point about the - so maybe I take the drone comment in a slightly different direction, which is a direction that we have been thinking about, which is there's a lot going out in this area now with private industry that where it's their business model to be out there all of the time.

And is there a way we can also get that information without doing it ourselves. And so we're looking at that and we're thinking about that and we're thinking about how we can leverage that to help improve the census taking and help improve our master address files.

The notion of looking at data in real time and then using it (unintelligible) I absolutely agree with you. In six months, in eight months in a year that information will be available.

And the census bureau has recently stood up an office in particular to look at this and to look at data analytics and to look at how we can integrate real time data analytics into our field operations in particular.

So we're dipping our toes in it its very early stages but it's important to us.

Jack Dangermond: It seems to me that the idea that I collect the data and then I put it into a format and then I can use it that sequences changes where I'm really measuring and doing exploratory data analysis with maps or charts dynamically.

And I can show you some videos that are prototyping that right now right directly from the cloud. So you can simply say we're going to keep the analytics in house for adaptive designs modifications and imagine that we were able to going back (John) to the vision to the Census Bureau it's just like the weather channel or the weather service.

I'm giving as I'm measuring I'm beaming out so that people can interactively explore this data dynamically. It sounds like a dumb idea and probably statistically it has some implications.

But, you know, I would just invite you to think about that notion of if we measure and we use in real time we collapse the whole cycle. That's going to be technologically possible certainly within the timeframe.

And it's one of those Jack visions that he's saying, you know, think bigger than simply automating in the old modality the same work. Just throw a lot of the stuff away and conceptually go more to the new modalities.

Like not reading, not automating the PDF files for dissemination but you're automating the book, you know, in real time.

Dan Atkins: just want to kind of recap slightly what both of the Jack's have said. You obviously your primary mission is what the core mission of the Census Bureau is the 2020 census.

But I think what you're doing has some really important secondary impact entailed. First off you're creating an existence booth and some best practices that would be adopted in other parts of the Federal Government and, you know, setting new expectations and efforts for doing it.

But the thing about feeding the real time data is an example of what on the NFS side I refer to as kind of dual use or multi-use. So in the NFS I've been advocating for years that these hundreds of millions that they put into research infrastructure yet intentionally used for education as well.

And so you, this is an opportunity to do that in a very limited basis. And the drone thing I had actually thought about as well. And so the question is whether the drones or the dual use or anything else like that could be done as a very contained pilot, you know, kind of attached to this but not in the critical path and not the way that would disrupt the critical path.

Just the fact that you used it as a proof of concept would be worth thinking about.

Jack Dangermond: One little addition is there is new, there's a new class of satellites coming out that are little tiny things that cost almost nothing that are like drones but are going to be feeding persistent surveillance.

And this again is in a timeframe of people are talking now about total earth coverage every four days with resolutions down to ten meters. While ten meters can give me my classifiers so I can actually subscribe to a commercial vendor maybe it's not drones but a

commercial service that could do classification of change analysis dynamically sort of bypassing the manual style.

So again just thinking in the timeframe and this is not science fiction this is actually committed dollars going up in the sky and real stuff. So it's just thinking about how will that change the instrument and the science.

Tommy Wright: Thank you very much. The photographer for the group photo was here as scheduled at 11:45 but we just could not stop. She has another appointment. We will aim to do that the beginning of the session when we return.

The question is when will we return? Can we return at 1:00? Can we return at 1:00 that's about 45 minutes is that all right? So we will return at 1:00 thank you very much.

Tommy Wright: Good afternoon and welcome back and I promise you that this photo will not be sent to anyone's university as proof of attendance or non-attendance as Roberto was suggesting. We will not do that.

Now let's welcome - by the way I should note that Nancy Potok the Deputy Director has joined us. Now let's welcome back, welcome Hector Benitez-Solivan from the Department of Commerce who will give us a briefing on the Federal Advisory Committee Act.

Hector Benitez-Solivan: Hi everybody my name is Hector Benitez-Solivan and I'm with the Office of the Assistant General Counsel for Administration of the Department of Commerce I guess like the parent department of the Census Bureau.

And today I'm just going to give a very brief presentation a brief overview of the Federal Advisory Committee Act, basically what the act is, how it is that it applies to you and everything that you do as committee members.

And if you have any questions like throughout feel free to raise your hand and I will try to answer questions if I can. If not, you know, we can talk later or we can go through Tommy if you have any questions and he can email them to me and hopefully we'll get all those questions answered for you.

So as I said I'm in the general law division in the Office of the Assistant General Counsel for Administration. And in addition to FACA what we provided to the Census Bureau on FACA we also provide advice on other administrative law issues.

Basically what I like to call the law that keep the government running it's all the behind the scenes stuff like appropriations, freedom of information act requests, we deal mostly with the appeals mostly.

Agreements between the Census Bureau and other federal agencies or other just organizations outside of census when they want to do work with census on something. Again a whole host of issues, Privacy Act.

So those are the things that we advise on. As you see on the slide we have our Web site and that number up there is our main line and Tommy if he doesn't know who to call on

some things or anybody at census if they don't know who to call they can call the number and ask for the officer of the day and then we'll try to answer those questions for them.

And on the Web site we have links to everything that we're going to, that I'm going to talk about today with regards to FACA. We have FAQ's on everything that we advise on in the office.

Unfortunately some of those things aren't live yet because we just transferred our old Web site to like a new format and we're still working out the kinks. But anyway so in addition to us there's also the ethics line programs division and they advise departmental employees on any ethics issues, travel gifts, acceptance of gifts, solicitations all that kind of stuff.

And the employment and labor law division and they advise, they represent the department in any labor disputes, any employment actions things on that nature.

So just a little bit of background on FACA if you all don't know this already. It was enacted in 1972 and it's one of those laws, it's a public access law what are also referred to as blue sky laws basically openness in government.

It was revised in '76 with the passage of the government and Sunshine Act. Again more openness and transparency in government. And it basically governs around 1,000 committees within the Federal Government.

There are roughly like 1,000 committees throughout the entire Federal Government 60 plus of which are within the Department of Commerce alone. And I know the slide says that census has three but we actually have two.

You guys and which one is it, race ethnic and other populations committee. The other one I was guess is no longer in existence.

John Thompson: PSAC.

Hector Benitez-Solivan: PSAC. they are still around well then okay, okay.

John Thompson: Which is a Federal advisory committee.

Hector Benitez-Solivan: Okay, I'm sorry I was mistaken that there are three. And the GSA the General Services Administration administers FACA government wide. So if there are any questions that our office can't answer then we can go to GSA and hopefully they can help them answer for us and with us.

The purposes of FACA. So as you can see up there on the slide there are three general, very general purposes that FACA exists for. One is to ensure that all of the advice that the committee's all the different advisor committee's to the government provide is objective and that deliberations are accessible to the public.

And as I've noticed with all like the technology and the camera's and the microphones and Sara told me about the youth stream and all that stuff you can't be any more accessible than what we're doing right now.

It also formalizes the process for establishing and operating and basically terminating advisory groups not that anybody wants to terminate you guys but it establishes a process for that.

And it guards against the wasteful spending on duplicative groups or any groups that may not be necessary for the government to do its work anymore.

Groups to which FACA applies. The FACA obviously only applies to certain types of groups. Here you have a very general description of the type of groups that FACA does apply to.

So basically any group whose membership includes at least one non-federal individual and which has been established or utilized by the government to obtain consensus advice is a group under FACA.

Now, as you see up there on the slide one of the keys to this is the utilized part. What does utilized mean? What we're talking about when we talk about utilization of a committee or utilization of a group is that it's actually managed or controlled by the agency that is, you know, that is the recipient of the advice.

And here we have four points, four things to consider when trying to, when going through whether a group is being utilized or managed and controlled by an agency.

One being controlling the membership or composition of the group, management or control of the agenda is another factor in seeing if an agency utilizes or manages a group, calling of meetings. If the agency is the one calling the meetings of the group then that's another factor towards the control of the group.

And participating in developing a group's final product, participation in line any of the recommendations or taking recommendations that the group provides that's another factor in determining whether the group is being controlled by an agency or utilized by an agency.

One other point is the obtaining of consensus advice. If the whole point of, you know, getting a group together calling a group together is to seek advice from the group as a group as we do with all the committee's here at census then that is a group under FACA.

And we'll talk about situations and groups that FACA doesn't apply to, which is the next slide. So groups that FACA does not apply to. One of them is groups of all federal employees.

If it's all just, you know, Jeannie and Tommy and Lisa or me getting together and discussing things and we're all federal employees that's not a group under FACA. FACA doesn't apply to our group and what it is that we're doing.

Individual advice. So you can gather groups to provide advice to the government but if you're seeking individual advice from all the group members and not advice from the group as a group all of you getting together and providing group advice.

If that's not what you're looking for if you're just looking for, you know, if Barbara comes to me and provides advice or, you know, Stephen or Horaco individually not as a group then FACA doesn't apply in that instance because you're all just providing your own individual views, your individual advice to the government.

Exchanges of facts of information if we're just, you know, getting together and exchanging facts, information not making any actual recommendations that's not a FACA group.

Some groups are statutorily exempt from FACA. Those obviously aren't groups under FACA. Bi-national or multi-national commissions. I don't know if census is a member of any bi-national or multi-national commissions but basically these groups are, you know, the International Trade Administration has several of them.

And, you know, it's where ITA and like for one instance ITA and their South African counterpart performed this bi-national commission to discuss trade matters between the United States and South Africa. Those types of group's inter-governmental commissions aren't groups under FACA.

Groups consisting of all government members not just federal employees but any government members whether it's state or local, tribal, foreign, regional if they're getting together to discuss cross jurisdictional issues.

So matters that are of import or interest to the Federal Government, to the State Government to the Tribal Governments whoever it is that's getting together.

If there are issues that cross all that affect all of them that interests all of them then those aren't groups under FACA but they have to be talking about cross jurisdictional issues.

And then finally the groups that do not advise an agency. That's what might be a little confusing but what that really means is, you know, if census is basically like a facilitator and the National Institutes of Standards and Technology another branch of the Department of Commerce.

They have a program where they get together, they facilitate talks and meetings between members of cyber, in the cyber security industry. They're just facilitating the industry members don't actually advise NIST on anything cyber security wise.

It's just a program that exists to get industry to talk to each other and develop best practices within the industry but NIST has nothing to do with it other than facilitating the talks. That's what we mean about that last point.

And there are statutory and regulatory requirements under FACA that we have to meet as, you know, as an agency in the Census Bureau. So there are regulations and requirements on establishing committees.

They should be established either by legislation or presidential directive or formal agency decision establishing them. They should all the committees if they're going to be a FACA committee should also have a charter.

And those charters are - excuse me, in place for two years at a time. So either when the charter is being established or when there is a renewal charter our office actually gets involved in clearing off reviewing those charters and clearing off on them.

Plans for balance membership. All committees have to have that and you can balance the membership through any balancing criteria that you establish whether it's a balance of

industry representation views, different members of academia like here different, you know, geographically diverse.

Diversed by gender, diversified by race any sort of criteria that you establish for the particular committee would have to be met to meet that balance membership requirement.

Open meetings, you know, I talked about, you know, we can't get it more open than we are right now with the access over the Internet and just, you know, announcing another point there announcing meetings at least 15 days on the Federal Register to the public.

Making meetings accessible to the public and open to the public as much as possible because FACA is an open government law meetings are presumed to be open.

Records retention, we have to retain records committee records for the life of the committee. Documents in minutes just like the openness and making the meetings open to the public but documents at the meetings, committee documents, minutes taken during any meetings must also be made available to the public.

And Tommy and the DFO he can, approves and calls meetings and obviously is attending meetings, adjourns them and helps set the agenda. And there are annual reporting requirements the GSA this is all stuff, you know, all the boring stuff that we at census get to do.

So yes open meetings the presumption is that all FACA Committee Meetings are open to the public but there are ways, there are circumstances in which meetings can either be fully or partially closed.

And if we're going to do that you have to meet certain exceptions, exemptions under FACA in order, you know, depending on the type of information that you're going to be discussing at the meeting.

If it falls under one of those exemptions then you can close either the entire meetings or part of the meetings under FACA. But the presumption is that they should be open and open to all. Yes over here.

What?

Barbara Anderson: Barbara Anderson, if you could say a little bit about what the nature is of the exemptions under which part of a meeting can be closed.

Hector Benitez-Solivan: So that - thank you for the question that's actually on the next slide. This is a list of - not it's a great segue. So this is a list of all the exceptions that apply in order to partially or fully close meetings.

One of them, you know, if you're going to be discussing any classified information then you can, you know, close that part of the meeting where you're discussing classified information.

Trade secrets if you're going to be discussing any confidential business information of any companies out there then you could close, you know, parts of those meetings.

Personal privacy information, law enforcement I don't know that any of this stuff actually applies to what, you know, this committee actually discusses. But these ten are the different categories under which you can close parts of meetings or, you know, close entire meetings if the entire meeting is going to, talking, you know, discussing any of these.

Barbara did that help? Yes, maybe?

Barbara Anderson: Well in some ways. So say hypothetically if the committee wanted to talk about the nature of its relationship to the Census Bureau where it might not want the Census Bureau employees to be there to hear what they said while they were discussing different views.

That would not be a basis for closing that part of the meeting.

Hector Benitez-Solivan: It would not.

Barbara Anderson: Okay, just thought I'd ask.

Hector Benitez-Solivan: Well and also the point, the main focus of open and closed meetings though too is open and closed to the public. What we're really worried about is, you know, members of the public accessing what is being discussed at the meetings, being at the meetings.

That's the real focus of the openness and of the FACA is the public access to the meetings and what's being - and the same way that, you know, that there's an openness requirement for the meetings themselves there's also a presumption that documents used at the meetings used by, you know, by the committee, minutes of the meetings are also, should be available to the public.

The public should have access to them. You can withhold information in those documents only if the information falls under a FOIA exemption. So if any information let's say is again classified or, you know, you're again discussing, you know, maybe any possible business confidential information.

That information so if there's any discussion of law enforcement or like personal privacy interests all of that information could be withheld in those documents.

Again I don't know how applicable it is to this committee but it could be withheld because that type of information can be withheld under FOIA. It doesn't necessarily mean that it has to but it can. Otherwise everything should be accessible to the public.

These are just some potential consequences of not complying with FACA and, you know, in order just to avoid any of these like, you know, getting sued and having claimant seek

an injunction in order to open a meeting that may have been closed or compelling disclosure of documents that had been originally withheld or just, you know, preventing.

They can seek an injunction to prevent census from utilizing any recommendations that this committee or any committee any census committee provides. These are just some examples of things that can happen when we don't comply with FACA.

So things to avoid in order not to violate FACA and there are examples under all of these categories. But anyway so activities exceeding the scope of the charter. We don't want to engage in any activities that exceed the scope of a particular committee's charter because, you know, then we might be violating FACA.

You know, kind of like, you know, going on the road or conducting outreach activities in order to publicize a report or like a committee's activities. That might not be, you know, that might not be within the scope of a committee's charter.

And an example that we've dealt with it was outside the scope of the committee's charter. So, you know, we want to avoid doing that, you know, just so we won't have anybody questioning, you know, the work that the committees are doing.

Committee making government decisions, you know, the committee's advise and their backer committee's advise they don't actually make decisions. The census or the Federal Government makes the decisions they take the actions not the committee's.

So in one example, you know, there was one committee that wanted to issue grants all by itself. Well they can't do that they don't have the authority to do that or draft legislation for, you know, for Congress to consider.

Again a wonderful idea, you know, but it might have been that their hearts might have been in the right place but they don't have the authority to do that only the agency can.

Grass roots lobbying, committee's shouldn't be engaged in grass roots lobbying to Congress, you know, lobbying for support of a law that they thought was advantageous to the agency stay away from that.

Publicity and propaganda again we had one example where one committee suggested, you know, doing like a road trip going on the road to publicize a report that they had come out with, you can't do that or we at least want to avoid doing that or geared towards the agencies like unusual timing of appointments.

You know, we don't want - if somebody's appointment still has some time to it but then, you know like a decision is made and it's well let's make new appointments even though there was, you know, there's like a year left on somebody's term.

You want to avoid doing that as well again it just questions the integrity of the work that the committee is doing and the integrity of the work that the agency is doing as well.

And disclosing internal government materials. Listen up census employees. Disclosing internal government materials to the committee. It's like of course, you know, we can, you know, we can provide documents to committee members to the committee's.

But we want to be wary because at least in the FACA context I mean in the FOIA context if there's anything in those internal documents that can be, that would have been withheld or that, you know, an agency doesn't want released to the public we can waive our ability to assert any FOIA exemptions on that information because any documents provided to a committee have to be available and open to the public.

So this is our role, my office's role I shouldn't say my office, our office's role the general law division in when it comes to FACA. So again we provide advice to census officials and actually we're the attorneys for the entire department when it comes to FACA.

So anybody and any departmental employee having FACA questions we provide advice on especially like when establishing committee's or with renewals or anything like that.

Like I had mentioned before we review and clear. We do offer official clearance on new and renewal charters. So every two years once your charter comes up and we get, I get to renew them for now.

Our office does rotate committee's and attorney's among the committee's because there are only 14 of us and there are 60 plus committees. But hopefully I get to stick around with CSAC for a while.

We review membership appointments and reappointments like membership letters if they're signed by the Secretary. If those letters are signed by the Secretary then we get to review them and we do clear off on them as well.

We review NOD's. So NOD's I should have mentioned that before. They're Notices of Determination. So if any part of a meeting is closed or if a full meeting is closed census has to issue an NOD informing the public that any part of a meeting will be closed.

And that has to be done at least 15 days before the meeting just like announcing on the Federal Register of a meeting date. So we review those NOD's and clear off on them as well.

And we assist defending the department in litigation although that's mainly like a DOJ function we act as an advisor. And we provide ongoing legal advice to the DFO's or anybody else at census or anybody else in the department who may have a question.

So again like if anybody here in the committee has any questions, you know, like you can come through Tommy and he can forward the questions to us or anybody at census can forward the questions through Tommy and we'll help, hopefully answer those questions.

And that is it for me today. So again there's our main line number, there's our office, our Web address if you have any questions on anything feel free to give a call again through Tommy or email and - yes.

John Thompson: I'd like to ask a question. Other committee's that we've had have asked this question so I'm just asking it so...

Hector Benitez-Solivan: Yes.

John Thompson: ...that is can a subset of the committee say half the committee get together to talk about the relationship with the Census Bureau like Barbara was saying?

Hector Benitez-Solivan: Can a subset of a committee - ask again I'm sorry.

John Thompson: Yes so Barbara said they can't have an executive session, essentially was what Barbara was asking for. The committee can't have an executive session unless they have one of those exemptions...

Hector Benitez-Solivan: Right.

John Thompson: ...but we don't have those exemptions here at census because we don't discuss anything like that.

Hector Benitez-Solivan: Yes like I was saying it hasn't been, it's been forever since a census was like - and the committee meetings have been closed or anything.

John Thompson: Right but so could have the committee I'm using half, could half the committee meet and have equivalent of an executive session outside of FACA, outside of notifying the public?

Hector Benitez-Solivan: Well I mean we want to be careful because if they're doing it in their capacity as committee members then I would advise against that. I, yes if they're doing it in their capacity as committee members I would advise against it.

John Thompson: And I just wanted to make sure just get that out there so everybody knows it that...

Hector Benitez-Solivan: Yes, yes, yes.

John Thompson: ...you have to be careful even if you're just (unintelligible).

Hector Benitez-Solivan: Anybody else? Yes.

Barbara Anderson: This committee has working groups that needs to, that need to talk to each other to figure out what they think. And they also need to talk to parts of the Census Bureau to be informed and not to make recommendations but to try out ideas to see if we thought this what you think of it or what are we overlooking. That's not a problem is it?

Hector Benitez-Solivan: No I mean you can, I mean you can have working groups I mean...

Barbara Anderson: And there's no way all those meetings can be public I mean you'd go nuts.

Hector Benitez-Solivan: ...right, no I mean you can have, yes so yes you would go crazy. You can have working groups and working groups because the point of the working group is like you said ideas and hash out may be things that then the full committee would recommend to census.

Like those aren't, those aren't I mean they are meetings of committee members but they're not meetings under FACA because they're kind of like, they're just so internal.

Barbara Anderson: They're not making recommendations to the Census Bureau...

Hector Benitez-Solivan: They're not making recommendations exactly.

Barbara Anderson: ...they're presenting things for consideration by the committee and those meetings are public.

Hector Benitez-Solivan: Yes, yes.

Barbara Anderson: Thank you.

Hector Benitez-Solivan: Yes you're welcome.

Tommy Wright: So thank you very much Hector and everyone. Thank you very much.

Hector Benitez-Solivan: Thank you, thank you for having me.

Tommy Wright: You're welcome. Next we will hear a presentation from Rebecca Blash the Center of Enterprise Dissemination Services and Consumer Innovation.

Rebecca Blash: Good afternoon everyone. So my name is Rebecca Blash and I'm heading up a new effort here at the U.S. Census Bureau that's really changed the paradigm of how we disseminate all our public use statistical data for the public called CEDSCI, which is the Center for Enterprise Dissemination Services and Consumer Innovation.

And CEDSCI - I apologize I just forgot how use this. CEDSCI really the vision of CEDSCI is to open up our data for the public for innovation. It's to allow the public to innovate more with the data.

It's to help users solve for specific problems that they are looking at from researcher's to developer's developing applications, visualizations to entrepreneurs trying to expand or open a business, to public policy makers to solve for policy issues within their communities.

So it's based on the federal digital, aligning with the federal digital strategy, which is really all the statistical agencies have been tasked by the Administration to open up the data.

To not disseminate in ways to the public that the agency feels that the public maybe needs to see it but how the public wants to use the data. And with emerging technologies

and the rapid sophistication of our users we really have to think about how we can provide the data and information.

Even the content that we put out, press releases, PDF's in a way that the public can access them much easier, find it easier and work with the data much easier and in more innovative ways.

And the CEDSCI initiative itself came out of the work of a task force that Tom Mesenbourg stood up about two years ago that was cross directed that asked us to primarily look at how we disseminate our data for the public.

How the public needs to use our data, how can we improve the public interaction with our data and meet their needs in ways that are less restrictive than potentially can report, you know, static table's things like that.

But it was also we also were looking at ways to improve the efficiency of how we disseminate our data. By streamlining a lot of the activities that the bureau undertakes across the various directorates that collect the many, many censuses and surveys that we put out.

So it was really a twofold approach but it really was again with that user centric perspective in mind. So I'm sorry I keep forgetting that I'm not doing this sorry about that.

So this is about the promoting of data aligning with the federal digital strategy that I just mentioned and the work of the task force. So these were really the goals of the task force, you know, as our guiding principles as we worked on this.

And as a result the task force turned in a concept of operations to our operating committee about a year ago spring, a year ago. And after, you know, review and presentation and vetting with the census bureau the operating committee decided to go

ahead and move toward implementation of the concept of operations the task force came up with and I was asked to head up that effort which became known as CEDSCI.

So the goals of CEDSCI really are again to foster and maintain a customer focus dissemination environment yet be cost effective. To view our dissemination as a business function with a technology component.

Everything is technology now days from the Web site to the dissemination systems to API's application program, interfaces but we have to ensure that the business needs of both the internal stakeholders at the Census Bureau as well as our external stakeholders are met through the technology that we're developing.

So there is a large technological component of this project but again user centric engaging with stakeholders early and often collecting feedback and engaging and I'll get into a slide that kind of shows how we break out what we call the components of the CEDSCI project.

The intended benefits of CEDSCI really, you know, we look at cost savings of course. We're trying to get away from a siloed approach to dissemination. Traditionally the Census Bureau over the years as we've gone from taper to the Internet and more streamlined ways to release data evolved across the directorates, the economic directorate to the decennial and the demographic directorates.

They all collect censuses and surveys and look within their directorate how to put the data out for the public to be useful to the public. But moving now with the modernization of technology and the sophistication of users wanting to both access the data more easily but also potentially combine it in ways that we may not think of and be innovative with that data we really need to move to an enterprise approach.

Again for a cost savings as well that all of our, we are one organization putting out massive amounts of data. How can we take an enterprise approach to putting out this data collectively across the board?

So more efficient effective work environment for the people here at the Census Bureau. We hope that by reducing redundancies in some of the activities through services and tools to either tabulate the data, disseminate the data, stream the data to what we're calling a service layer to put it out for the public will free up those people.

Those employees and subject matter experts to be more creative, be more innovative, engage with the users, think more thoroughly how they can potentially develop new products or new ways to serve the data to the public. So we really want to encourage innovation both within the census bureau as well as for our external users.

In terms of managing this transition to CEDSCI we were stood up the CEDSCI team itself mid to late summer. So we're really getting organized we've got proof of concepts off the ground and underway now and we're looking at how can we reuse, redeploy, reengineer some of the existing systems.

So the purpose of CEDSCI is not to stop the old way of disseminating and develop from scratch a new way of disseminating. It's really a transition how we do transition to a more streamlined effective way of disseminating and really change the paradigm of how we're disseminating data.

So a lot of the work that the task force had done in developing the and now that the CEDSCI team is picking up is examining existing systems and identifying existing methodologies and ways of for example creating metadata.

And looking at our metadata models to see what can we leverage within the Census Bureau not just technically but also the institutional knowledge and expertise that the folks working on dissemination have put in, you know, have garnered over the years, many years of working here and really leverage those talents and the systems and identify some of the ways we can reuse some of the systems.

So we are not starting from scratch and in fact we've already established what we're calling our metadata model that will then curate a separate metadata repository. So again along the federal digital strategy and others is opening the data is separating the metadata from the databases themselves.

That frees up the data views in many more ways and repurpose, a big cost savings and efficiency that we plan to achieve and believe we will is that by separating the metadata you're no longer taking discrete copies of all of the data that the census bureau disseminates.

And formatting it and structuring it and curating the metadata for one discrete tool, one discrete purposes. And so now you've got these multiple copies of potentially the same data for different uses rather by having a separate metadata repository and a service layer you can create tools, visualizations, applications.

Even smart text enhanced search that all go to the metadata repository to pull back the data from the source. So essentially what we're saying is there is sourcing the data once, repurposing it many times which will lead to great efficiencies and streamlined access to the data.

And forgive me I keep forgetting was I not on that slide when I talked? My apologies. So again the reengineer that the program errors will be delivering like I said to a single system of the services layer that we're calling it through a separate metadata repository. And this will eliminate that siloed approach as I talked about in the past reference and really, you know, bring us to the, you know, forward facing future facing as one enterprise for dissemination of our data.

In addition CEDSCI will coordinate very closely with the geography division. This isn't just about statistical data and dissemination it's really about how the public wants to see our data.

And of course a lot of that is through visualizations and maps and location, you know, people want to know where they are on this earth if you will and how the information around them relates to them.

So we're working very closely with geography division. We have a leadership team in place a cross directorate and the geography division is very engaged with us on identifying more standardized geography, how to handle geographies from the hierarchies or summary levels to geographies over time.

Kind of tackling all of those issues and in fact we have an effort underway now looking at how econ the econ director defines geography for the surveys they collect, which is different from the demographic data and how the places and geographic places.

So how can we look at standardizing across those? So if a user wants to look at economic data in the context of demographic data it's a new place for them. So that's one example of how we're engaging around the research efforts for this project.

And also as I mentioned separating the data from the presentation and providing common environments and shared services across the enterprise.

So this is how we're organized. CEDSCI really is four main areas. The sort of green box around it for those of you that can see it. It is really the business process management, you know, aligning with our survey lifecycle design process and complying with, you know, any OMB regulations around how we put a project forward.

But the blue boxes really are the four key components, enabling technology platform, which is what I refer to as the service layer. That's really the technology it will be modular, it's service oriented architecture.

We're trying to look at as much open source products as we can for this and it will provide a common infrastructure. In addition I believe that you've been briefed earlier on the CEDCaP project, which is really the enterprise collection.

This is really the sister project to CEDCaP for dissemination. So how can we leverage some of the existing infrastructure back in shared services and tools to streamline the collection process to dissemination process?

The other is metadata standardization and we really call this content harmonization. It's looking at all of the metadata across all of the census and surveys we collect and all the different types of data that we collect.

From public use micro data that we disseminate to aggregate to longitudinal to time series. And the goal with this is to start I guess I would say the 80-20 rules with a separate metadata repository.

As I said we have the model in place that can curate and house all of the different types of metadata for our statistical data that we put out. As well as even in what we call unstructured content so tags for news press releases and PDF's and things like that.

That will then enable and enhance search. So if a user comes to the Web site and types in a topic that they're interested in, the search can go through the metadata repository looking at the metadata that we've collected and the tags and pull back potentially the dataset about that topic as well as other things they may be interested in.

Maybe it's just a report, it's a press release if it's about, if the topic also involves the geography, everything to do with that particular geographic choice. So by having a separate metadata repository tools like that can be built that can pull the data back for the user.

In addition it allows you to curate metadata and never have to I guess for lack of a better word you don't take any down. Because what we're doing with this metadata repository is once the metadata for a release is curated it stays.

Subsequent releases we just want to collect delta's, changes, did something stop, start or does it mean something different now. And then that gets added to the metadata repository.

So every time there's a new release of an ACS or even the decennial every ten years the entire metadata file that goes with that dataset does not have to be delivered to this metadata repository rather only changes.

And also the beauty of this is if a variable has stayed the same in a dataset we simply put in the new release year and that will allow people to do comparisons over time if nothing changed.

And business rules are built into that and we're already testing that and we have some things in place that allows that to happen. So we do have the model stood up to do such a thing.

Next we have what we call EIS Enterprise Information Services and we say information instead of data because again it's about all of the information and content the Census Bureau disseminates.

We disseminate more than just our statistical data, public use statistical data. So it's looking at everything from the economic indicators to all of the tools that currently exist and we're doing an assessment of those tools to identify redundancies and ways potentially that a tool could be combined with another and so multiple tools don't have to be maintained.

We're also looking at new innovative ideas around tools and products come on board this also falls under this EIS and how the user interface interacts with that. So again a lot of the EIS component is based on that user centric feedback that we're working on.

And then finally customer engagement management which is again the user centric piece and we're looking at, you know, stakeholder relations and ways to curate and get customer feedback.

And a big component of that besides the personal interaction is something called SEM and I'm not sure that you've been briefed on that. Jeannie, yes okay a little bit.

So this dashboard where we're streamlining all of the inputs to our customer feedback can then inform the development of the technology and the capabilities that we're putting in place to ensure that we're meeting customer needs and anticipating those customer needs.

And again the platform is also being designed so it can be extended, you know, we are anticipating, you know, who knows what five years from now, ten years from now in terms of emerging technologies.

And ensuring that because it's componentized it can be extended to meet those needs. And, you know, an example I always give is 15 years ago an API wasn't in our lexicon not it is.

And because of infrastructure we're actually leveraging as a baseline here at the Census Bureau we were able to develop that API and put it out and it's one we've done pretty well sort of leading several statistical agencies in our API development.

And I guess so the next one, slide is really this is sort of our future state. This is really showing an information flow primarily. So you can see on the left those are the datasets.

And as they come through the system they go through tabulation and preparation for either to become PUM, Public Use Micro Data or tabulated files. And then the metadata is separated from that content.

So that as you go through the product preparation you can access that same data source and produce multiple products out of that data for dissemination tools or visualizations.

And then you'll see the common services, common tools, common infrastructure. This is really where we want to leverage across the board, across program areas to provide these kinds of services.

So that each program area or director or division doesn't have to develop their own tools to handle their data and get it out the door as well as link back and leverage work that CEDCaP is doing on the collection side for some of these services.

So that's what I have. I know we're kind of running over time for CEDSCI but just to know that, you know, we have a lot of capabilities we're looking at. We're really forward facing and we plan to have a prototype in place by the end of summer, fall that can move into a beta.

So we're going to have a lot of customer and stakeholder engagement in testing that data and it will, we plan to have a cyclical process. So it won't just be a one-time sort of comment period.

We want to put things out, get that feedback in, make corrections, put it out again and just have a continuing ongoing engagement with our stakeholder community.

Guillermina (Willie) Jasso: Thank you so much. Committee comments, questions?

Roberto Rigobon: So I had a question I thought a reaction to something that you said about the innovation both internally and externally. And I don't know if you have thought about how to do that.

But let me suggest competitions and tournaments are just a fantastic way to get every single college student to design whatever you want. Very cheaply.

And then you bring them here and you give them a certificate that they actually designed the visualization to release data households geographically and things like that.

I mean at law school we have hundreds of ideas about how to do innovation. The only one that actually I have seen that works is truly a tournament.

So I mean that, you know, being the census it would be a - you can create something gradually shaped that is like an award and you give them a particular project.

So this is the consumer designing something for themselves and you have the data organized and then the interface you can do it through that way. And we love beating Stanford all the time so, no joking.

Rebecca Blash: Thanks Roberto the middle of that competition.

Roberto Rigobon: Yes you was not listening so you didn't pay attention to it. But I thought that, you know, this would be kind of a recommendation where you just create a certificate, you know, the outsource get the name on this software and that's kind of the award.

And it's just a better way also to have a constant innovation of visualization tools that is going to be really hard for you to keep up with that.

Rebecca Blash: Right correct.

Roberto Rigobon: And therefore but they are starting that right away. So...

Rebecca Blash: That's a great idea thank you. And some of the things as you know we've looked at really more around our API versus, you know, other ways to innovate as you suggest.

Our participating in the national civic day of hacking things like that, you know, and going to meet ups but that's really more the developer community but we also want to engage with all of our stakeholders around ideas for product visualization. So thank you.

Guillermina (Willie) Jasso: Yes Sunshine.

Brian McGrath: I'm sorry could I just add one, would it be okay one quick point. There's a concept within the government called challenge.gov I'm sure you're familiar with it.

We're, we are likely to be using that avenue to marshal the wisdom of the crowd around Internet self-response.

Guillermina (Willie) Jasso: Sunshine.

Sunshine Hillygus: Sunshine Hillygus. Looks like I don't work....so I'll just this is a stockholder request in the field we don't use census nearly as much as we could or should. In fact I didn't even realize some of the things related that political scientists would be interested in that the census is already producing some civic engagement questions and CPS and some other places.

And so part of I think, you know, thinking about how to get new users, new stakeholders involved might involve some outreach to groups that could be using and I don't know if you have thought about not just in terms of ease of access but expanding usage as well.

And then just one other thing is again as a political scientist in terms of geography Congressional district and media market would be two geographies especially of interest.

Rebecca Blash: Okay thank you. As far as the outreach and engagement that we're looking at indeed we are working another component that wasn't so much up on the slides is of course a communication strategy and marketing it for outreach to the public.

So that would involve not just the and to your point the metrics that are coming in because that's just what people are looking for but who are we not reaching.

So it is engagement in conferences and we have a lot of, you know, affiliation groups like (Apu) and (Copas) and others that we want to reach out to. And then as we - and we're working with our state data centers and there's census information centers.

I just presented to them on Tuesday and we're already following with them on ways that we can get more involved on their whole member, you know, through Webinars and WebEx's.

So any ideas around, you know, how to do that outreach have been great. Especially to specific audiences, you know, we know some but we don't know everybody. So any kind of feedback on, you know, the sort of targeting and then we could also look at how we present this and customize it for that particular audience to really say what is in this, you know, what's in this for you kind of thing how can we serve you better. So that's one thank you.

And as far as geographies we do have what is it the CD 113 there's an application actually we have on our Web site now to look, it's looking at population variables.

So we are always looking for new ways to put the data out as information not just data on an FTP site or even through an API and obviously not everyone's a developer.

So what kinds of ways can we be more innovative in turning the data into information, useful information for the user? So we'll definitely keep those things in mind and we welcome ideas and suggestions of course.

Guillermina (Willie) Jasso: More comments, questions? Yes Jack Dangermond.

Jack Dangermond: Hi, I noticed Brian left the room but this is as much for Brian as it is for you Rebecca. There is this emerging notion of public private partnership with the cloud computing companies.

And it's built around this notion that they will give you free storage and this is still experimental. It's free storage in the cloud. Brian this is also for you but there's this emerging notion in commerce and part of this commerce open data committee or whatever it's called is going to talk about this notion.

That in a public private setting somebody like let's say Amazon or Microsoft or Google would provide free storage of your information but they would charge for the user manipulating it.

I mean it's just I think things are changing like storage is becoming the price of, you know, nine track tape in the cloud. And if that's really the case whether the cloud companies provide it as they are now with some of the remote sensing data like said and so on.

And they charge for people accessing and manipulating it this is a new kind of model in a way. And it means that private sector companies could stand up manipulators or visualizing software tools in the cloud around this free data storage and they could exploit it.

So that in some ways it builds on what Roberto was talking about of startups or kids playing around it. Let's imagine startups are using government data as an infrastructure or a platform and the cloud just happens to be the place where it's standing.

And I think there is an opportunity not only for innovation but also for just normal business where how does this work financially? I think it's just going to fall into place certainly by the time the 2020 Census is here there will be all kinds of places I'll store your data for free and I'll tell everybody else for you about the data being there for free.

And people with software tools have put their software on my cloud and then boom, you know, it takes off. So it's this odd kind of combination that I think it's worth exploring and commerce will be doing some of that exploration anyways.

But where it may not require investments in infrastructure for the dissemination side yes for the processing side and leading up to that but the green box in your last thing the actual release of products could be not only released in that space but also exploited more effectively than say writing your own even API's or software tools to do it. It's just an idea.

Guillermina (Willie) Jasso: More discussion, yes Steve Ruggles.

Steve Ruggles: I was wondering about the metadata the product/presentation metadata. We spend a lot of money developing metadata for Census Bureau products and it sure would be nice if we didn't have to do that. So I was just wondering, you know, what sort of a format or structure or standard are you planning on providing?

Rebecca Blash: Well, for the presentation metadata specifically, we're looking at both how a tool or, say, a table creator what access the data and then structure it on the - I don't know if you want to call it the front or the back end. I guess the front end as the user sees it so it's building in business rules for how it's presented back to the user through the tool or the product. Backend infrastructure is just the database itself of the statistical metadata and the tag so we haven't on two sides. So on the presentation side...

Steve Ruggles: So will you be releasing, then, what the content metadata, what's in your in your diagram? I mean is that going to be made public? I mean, it doesn't sound like what you're thinking of is the product presentation metadata will be much use to us, right?

Rebecca Blash: Yes, I'm not sure on that side. On the content, yes, through our ATI, we're looking at developing a metadata service, almost an ATI for the metadata, to then have it be publicly accessible so that they can get straight to the metadata and not have to necessarily go through the data and all of the data dictionaries and all that stuff and doing all that structure and working about - it will be metadata as a service to the public like an API to the metadata - content metadata itself.

And we're already looking into that and doing the research and I think some of it may even be proof of concept as far as being developed right now around that. We'll keep you posted.

Tommy Wright: Thank you very much, Rebecca, and everyone. Now we will welcome Ron Jarmin and Bill Bostic and they will talk about big data.

Ron Jarmin: All right. So Laurel and Hardy will take over now. So - well, thanks for having us back to update you on what's going on, on big data, a term that I think the vast majority of people either don't know what it means or don't like to use it.

Maybe I'm in both of those camps. I'm not sure. So last time we talked about sort of the Census Bureau moving out on this front and there were sort of four broad areas that we needed to sort of concentrate our research agenda around.

And those were to refresh your memory that the logical, computational, sort of policy-legal type issues, stakeholder engagement. And I think we still feel that if we fail to do any of those, that we probably won't be successful in this area.

So since we last talked, we've been chugging along. And so just to update you on some of the things that we're doing, so in the research and methodology directorates we're going to stand up this new center and it's kind of already doing some things.

So when I - you know, we had to send the slides and really early, you know, at least by my standards for giving talks, and so we didn't have a name. And I still don't know we have a name but we have a name that so far at least the majority of people who've heard it didn't get violently ill when they did and so let's try this out - the Center for Optimized Data Applications, or CODA.

So if you hate that, let me know and we'll go back to the drawing board. But anyway, what we intended the center to do is to be sort of a hub for Census Bureau efforts on big

data and related topics and that would include leading projects, so the center would be in charge of moving out and there is things.

But also as Bill will talk about here little later, sort of working with affiliated projects that are managed in the other directorates, so there's some stuff going on in E-Con that we're sort of collaborating on.

In the last sort of picture is that we're looking for chief, so if you know summary that's interested in going to the Census Bureau and being the chief of CODA, let us know, and we'd like to talk to them.

In the big data center or CODA or whatever we're going to call it, you know, what are some of the projects that we have sort of going on already? The first one is one going to talk to you about a little bit of detail here in the second.

We're calling the innovation - the measurement initiative and we'll get back to that.

We're also sponsoring some workshops partnering with MIT. One of them is going to be on trying to rethink sort of the data infrastructure that underlies statistics on commodity flows.

Right now, we use a survey to do that. And we like to think more broadly about that. But also big data and privacy and then big data in adaptive survey design which we might be tweaking that last one, but it's not nailed down just yet.

We've also been sponsoring big data class. We've had two of them, joint with the Patent and Trademark Office; The University of Chicago was the educational institution we partnered with. I think we're going to try to move that to the joint program on survey methodology if that's possible.

And then lastly, we've also sort of built a small sandbox to test some of the software tools that would want to use in this area. Right, so, now to talk in a little bit more detail about this project, the innovation measurement initiative, which is sort of - could be viewed as a

prototype of thinking about how we want - might sort of modernize some of our data collection activities around organizations.

So this is a collaborative project that we're working with folks from the University of Michigan, Ohio State Chicago. And the basic idea is, if you're familiar with the star metrics project that Julia Lane started when she was at NSF, we have - we've been able to get sort of that type of data on all the activities associated with federally funded research grants at a number of Big Ten universities.

So this is data on all of the staff that are paid off of the grants, whether they be faculty, post-docs, grad students, undergrads, directors, you name it. All of the vendors that they do business with, and so it's a big chunk of, you know, sort of transaction level data that we've been able to get from these universities.

And so from that we want to produce some statistics that are consistent with the Bureau's economic and social measurement goals but also directly relevant to the data providers.

So for those of you from academia, trust me, your folks in your ledge affairs offices are drooling over the stuff. So just some background - just come you know, so kind of where we involved in this?

I guess that would view this partly as a project of opportunity. It was something that came across the trends and looked like something that we could actually start working on and make some tangible progress on.

So, you know, it does give us an opportunity to improve measurement of a small but important sector of the economy. You know, again, you know, federally funded research is, you know, compared to GDP, is a pretty small number.

But, you know, we can address some gaps and measurement and look at sort of the process of innovation. But I think, important for us, if we're thinking about going

forward and trying to negotiate more frequent deliveries of more information from businesses in the economy, our users want more timely data and more detailed data.

The only way we're going to get that is to get more source data. If we're going to be - if we're going to try to do that, we're probably going to have to work harder to deliver data products that we might actually want to consume as opposed to a lot of the stuff that we do right now.

And so then, you know, to do prototypes and sort of learn whether we can scale this and extended to other sectors of the economy because, you know, the resource load is taking right now is probably feasible to think that we would do this for the whole economy.

So some sort of innovative aspects of this, so it's a collaboration with this new institute at the University of Michigan that I'll talk about in just a second. But like I said, it's an experiment on getting off that type of data from a particular sector of the economy.

So we're getting, you know, the data that we're getting are very much complementary to the sort of data that we collect from organizations and firms right now - that on workers and their input and that sort of thing.

But we're getting it at a much more granular level of detail. So, you know it's how do we make sense of all that? And then, you know, I think it also is nice that it makes extensive use of some of the skills that our classes have been teaching and our students have been acquiring.

And so when - it's hard to see that up on the board but when - the goals, when we build this, will have an infrastructure, sort of a large, sort of longitudinal database that tracks basically all of the inputs and outputs of university-based research from, you know, when it was at the university into the economy. And I'll sort of describe some of that here in a second.

But to talk about this institute just for a minute, so we got some funding from Sloan and Kaufman to stand up this institute at Michigan. And it's going to be the go-between, between the universities in the Census Bureau.

So the data will flow from the universities to IRIS, the Institute for Research on Innovation and Science. There'll be some basic manipulation of the data there and some products that come out of IRIS.

And then the data would flow to the Census to be linked with the Census data assets and get other products produced, right. So when we're thinking from, you know, the benefit of this kind of arrangement for the Census Bureau, is that, you see in the bottom there, one MOU rather than N.

I think that's - there're some folks from policy in here. You know, MOUs are a real pain in the neck around here. And if you - you know, so we think eventually we might have upwards of 100 universities in this is data set, you know, to do one MOU as opposed to 100 is a colossal improvement.

In fact, it wouldn't be feasible if we were going to do 100 MOUs. So I think that this new arrangement is very beneficial. All right, so, let me just, you know, I want to quickly give you a sense of what we could do with the data, so you that the linking of the university data to the Census Bureau data is both feasible and seems to give some sensible patterns.

But we're not going to really go into any real detail here. But anyway, so these are some of the things that we would like to look at doing this project. So we can look at the worker characteristics of organizations that people who were on university research projects, where they work now, so whether they are still in academia, whether they are in a private sector, whether they work in the government.

We can look at the job placements of students, the characteristics of the vendors that supply areas - that their services to these things. A particularly interesting thing, especially from my own research agenda, I think here is that we can actually identify

firms that start up as a result of - or maybe not as a result but that are associated with folks that were on these grants.

We can look at patents we can look at trades. So here just real quickly is a - we looked at job placement of folks that left the University who were paid on these grants and we can see where they go.

So sort of a majority of them are in private industries. So these are - like a center Big Ten universities so several of them have medical schools and the other ones seem to have engineering schools, so you know, this sort of, you know, the track towards industry makes sense given the concentrations.

Interestingly, you know, so one of the things that the universities are interested in is how many of the folks stayed within, you know, stayed local, see to look at the percentage of the folks at leave and take jobs that are, say, within the same state or within 50 miles of the university.

So we're able to answer those sorts of questions by linking their data to our data. This sort of shows - you can look into by the funding source. You know, I'm not to spend too much time on this but it just, you know, gives you a sense of the kinds of cuts you could do.

And so you can look at, you know, whether the grant was from NIH or NSF. There are other federal things in there too but we just didn't want to have it exhaustive. You can look at the industries that, for the folks that go in the private sector, you can look at what industries they go to.

So here you can see that some concentrations - so manufacturing is sort of over represented versus the whole - you know, this is the LBD. This is basically what the national average look like.

So, you know, we've got more folks going into manufacturing, more folks going into professional and scientific technical services and more folks going into healthcare.

So the sort of pattern seems to make some sense given with the nature of the workforce is coming out of these universities and the folks that were funded on research grants. And you can look at it by more detail here as well.

And then, you know, just, you know, so that folks can come you know, so we look both at the industries that are overrepresented, so you think architectural services, computer systems.

But then the ones that are underrepresented, so it's good to know that our federally funded research grants aren't paying our next generation of scientists and engineers to go to work in limited service eating places which is fast food restaurants.

So they're very underrepresented in the - although some are, so I don't know. That would seem like a pretty big failure. And then we're also able to look - so this is the start up, so we can look at two types of business formation behavior by these folks.

I'm not showing you basically, you know, which is by far the most common, so a lot of faculties do some consulting work, and so are part of what we call the non-employer business universe here at the Census Bureau.

But this is for folks that start businesses that actually have paid employees. You know, so this is - we're able to identify this. Some of this work, I think, some sort of a lower bound - these are people that we can actually identify through applications for employer identification numbers.

We could see that they actually started the business in that business has paid employees. So just, you know, this just kind of shows that we're able to take these data from these universities, match it to our data and do some interesting things.

But, again, this is, you know, very much at the beginning of the project. This is mostly a proof of concept. But now to talk about some of the affiliated projects, so you know, these are things that we're doing, so we're working with 20-20 on a few things that I think Lisa actually talked to more detail around this morning. But we're also doing a lot of stuff around retail statistics. So I'll hand over the clicker to Bill.

Bill Bostic: Okay, good afternoon. We get a lot of demand from stakeholders about more geographic data. Most of our monthly indicators and annual surveys, we provide national snapshots of the economy, retail being a very important sector.

And we thought about what other large big data sets are available, and so we targeted retail and created this retail big data project. And so the goal is actually to supplement using big data sets that produce more geographic data with the retail sector.

Pretty much, we give a lot of retail information of the economic census that's every five years. So we really want to do it far more frequent during the interim years of the Census.

So our current focus was on the estimates of sales at this time. So we're not exploring other estimates for data items such as inventory or purchases or expenses.

And we're also going to look to test kind of the big data environment to supplement survey responses obtained directly from retail companies. And so we want to explore this because certainly a response rates have declined over the years especially in our voluntary surveys in recent years.

So I'll tell you about kind of the acquisitions process that we went through. We had a lot of discussion with a third-party electronic payment processor just to get some information on the process of dealing with credit cards and debit cards, et cetera.

And so we went out and posted a request for information. As a matter of fact, we posted several requests for information hoping to get other entities that, within this business, to act to provide some information to us about their capabilities of meeting our data needs.

Unfortunately for us, we didn't - we had two inquiries but no actual formal responses to these requests. So we thought okay, let's put out a request proposal, let them know that we're really serious about purchasing data, and it's all about the money.

Well, we did get a bite or two from MTD. They actually are the scanning data people or at least one of the companies that scan data. And they actually - we had an agreement in September and they provided data for us from the automotive parts and jewelry and watches.

And the reason why we chose those two industry groups is because they align nicely with the North American industry classification system for our monthly retail and are annual retail programs.

So they provided as monthly data for 2012, 2013 and 2014, and we received the last of the information in February of 2015. So we put together a retail big data theme, and that included staff from the economic directorate as well as the research and methodology directorates.

And what we wanted to do was to leverage this team for other efforts in the future but we give them a short goal term to evaluate the data obtained from MPD and to determine its usefulness in meeting, supplementing our retail statistics with more frequent geographic level efforts.

So I'll give you some information about MPD. It's formally that National Purchase Diaries. It's a private company founded in 1967 and provides market research and focuses on consumer packaged goods.

The cornerstone of MPD's market research capability - because they have agreements with retail companies to obtain aggregated transactions data on a weekly basis.

Apparently MPD has agreements with approximately 900 retailers worldwide covering approximately 150,000 retail locations and \$400 billion in annual sales.

So there agreements are generally with the large retailers, so they do not cover small businesses. Certainly they provide weekly feeds on store identification and the location, the items and product codes that they scan, dollar volume of sales, units sold.

They calculate the average price and certainly they flag distinguishing online from in-store sales. And they focus on non-food and non-drug categories. They provided us the names, and in some instances, addresses of the retail companies and indicated which companies are included in each product category.

MPD gets a complete fee from these retail companies but they only have an agreement to use some of the data. So we asked them what data do they not use? So just to get a real good handle on what they're capturing from these weekly fees.

They focus on products such as auto parts, beauty, electronics, toys, video games, apparel, footwear and office supplies. And there are two other market firms - Nielsen and IRI. They cover the food and drug categories per agreement with large retailers that they have.

And between these three market research firms, they cover about 95% of the very large retailers. So the retail the team, they developed a plan for evaluating the data. And so the team will develop an understanding of MPD's processing of data to the extent possible, produce summary and descriptive statistics for each data set, as well as identify the potential areas in coding and recoding of the data.

And two examples are ensuring the address of the retail store is correctly coded to metro areas, state or census region or a division and making sure missing or rare data values for some of the geographic areas or product categories are valid and are not errors in processing.

And we're going to compare MPD data with Census Bureau data to obtain kind of a rough assessment of coverage of the retail universe as well as determine if MPD data

tracked well enough with the Census Bureau data that serve as a potentially informative predictor in a model-based estimator.

This will be done by various industries by geographic levels for aggregate totals, period to period changes such as current month to prior month or current to prior quarters.

We're going to use Census Bureau data from our monthly retail survey, our annual retail survey and the 2012 economic census and possibly with our business register for these comparisons.

So the current status - to date, the retail big data team, they've held two meetings with MPD staff to better understand the MPD data that we received. We've corresponded with MPD about data processing and questionable or missing data values.

We compared total and period to period changes of MPD data to sales estimates from our Census Bureau surveys. We've provided addresses - MPD provided addresses to us to help us identify the jewelry and watch retailers in the Census Bureau databases.

We started extracting the 2012 economic census data for companies included the MPD totals. And so the next steps are for the team to get the data to complete additional comparisons using these data.

And then we will look to summarize a draft report, and this draft report should be ready in late May, early June. Now, some other possibilities for big data and producing more detailed and timely retail statistics - so we're looking to explore the feasibility of obtaining data feeds directly from retail companies either through agreements with companies themselves or through a third party source such as MPD and Nielsen or IRI.

We think this could reduce a reporting burden on companies and it would allow us to obtain more detailed data more frequently. A third party could also help us with obtaining the data in a standardized format because we think that's important to look down that path.

We'd like to test actually this idea with some large retailers in the 2017 economic census. Another opportunity that we are pursuing is obtaining credit card transaction data by retail location from the credit card companies, third-party payment processors or requiring credit card transactions from issuing banks.

BEA, the Census and the Economic Statistics Administration, ESA, we're currently having conversations with MasterCard. It seems pretty favorable of obtaining the data.

Now I'd be remiss if I didn't acknowledge some of the risk associated with each of these alternative data sources. Some of the major risks include data security, data quality, lack of format standards, cost and inherent bias in the data.

The federal statistical agencies will need to determine the best way to mitigate these risks so that we can further explore the roles of big data in producing official statistics. So we'll stop there and address any questions and comments (unintelligible).

Guillermina (Willie) Jasso: Thank you very much. Let's turn now to Noel Cressie, who is our committee discussant.

Noel Cressie: Thank you, Willie. Thank you, Bill and Ron. I wasn't sure if I was going to get the microphone that doesn't work but it seems to be working so far. We had to see up and that screen. It's a bit easier in that screen. We also have handouts if you need to get a hold of them.

I thought I would take this discussion around a few highlights that have come out of Ron and Bill's excellent presentation, and that's the second slide please. I'd like to talk a bit about the goals and the research agenda that's shaped around those goals, the new center at the Census Bureau and the new institute as well.

Can I have the next slide please? Oh, thank you. It looks like I have control. Well, I mean, how deeply do we want to go into big data? How multi-varied do we want to be?

How many - do we want to track every person in the United States and all the features associated with them?

And clearly not from any point of view do we want to do that. And so almost by definition, we're dealing with aggregated populations and sub-populations. I'd like to continue to emphasize this word aggregation and, in particular, selective aggregation.

Governments need timely information of an aggregated nature. Businesses do as well. They need to react to those and come up with good decisions for the marketplace, governments, the sub-populations for which there dealing or compromise between various sub-populations.

Mandates need to be met and services needs to be provided and planning decisions need to be made. There's absolutely no argument about that. The point is we now have an embarrassment of riches. We have many, many data sources upon which we can draw.

However, the goal is still, and has been -- big data hasn't changed that -- we want to estimate economic and social characteristics in the presence of uncertainty.

It's just that big data has given us a chance to do more than we've done in the past. It may or may not reduce that uncertainty and that's another theme which I'm going to head on, so yes, the fog of big data.

Size matters but so does noise and messiness of individuals, and messiness of variables. They are all features that we've come across as researchers and it's that messiness that - and noise that causes us great concern.

If we have a lot more noise or a lot more messiness or a lot more of everything, it doesn't take away from what we're trying to achieve as a goal. There are statistical design principles they can be applied whether the data be big or small, but generally speaking, if you think about the population as being the big data and this principles and stratification clustering and randomness, the randomization helps us get at the signal.

I'd like to add to that list aggregate - and they keep coming back to this notion of selective aggregation. I don't agree hear enough about it when it comes to big data. And I'll make that case a little later in the discussion.

There are also computational principles as well. The notion of data sets or a data set residing in one place is no longer something we can count on. Data archives are typically distributed. We also need to worry about where were going to do the analytics, socioeconomic analytics.

And often the distribution of these very large data sets means that they cannot be passed across the Web. It's not sensible at all to move data around. There's a cost associated with it.

And so those analytics need to be performed (unintelligible) and you get into the notion of sufficiency of various statistics or near sufficiency. So these are important principles or issues or methodological things that I think a data center might be looking at.

Moore's Law perhaps might fail us in this case of big data and we need to find ways to keep Moore's Law going that in 18 months we can do twice as much as we did eight months ago and that perhaps parallelization, again, another computational principle, that we could apply as we try to work up our big data centers.

So big can also mean many, and the notion of big data being this humongous mess can be sort of divided and conquered by working down into a number of smaller data sets although still quite large perhaps.

And that would like to bring up the notion of confidentiality. Again, aggregation is often used to deal with it but here, you know, it may have many in terms of (multi-variants) and those interactions may actually compromise confidentiality.

So if one particular data set as very poor confidentiality, that maybe the Trojan horse which would allow you to get into larger data sets and compromise its confidentiality.

So one has to treat confidentiality in terms of the whole and not as a piecemeal. I like this quite a lot, from Josie Stiglitz. What we measure affects what we do. If we have the wrong metrics we will strive for the wrong things.

And since I read it a couple of years ago, I've applied it to my own research and I think anytime we're trying to strive or we have goals, I think we need to be very careful about what we measure. It may lead us down the wrong path to freedom measure the right things.

We have been very worried as statisticians and people dealing with data about bias and variance and mean squared error and cost as well. But we've been mostly concerned about these notions of bias and variance.

We're very worried about cost now but I'd like to put to you that it's more like an upfront cost and that bias and variance, if we don't worry about them, if we treat them essentially as a constraint, we may have a long-term costs associated with them.

So we should worry, I think, about whether or metrics have changed, and this, as a consequence, what are we actually striving for? So let's look at the traditional metrics of bias and variance. I used to be at the Census Bureau about 30 years ago. I was an ASA census fellow.

I understand a little bit about its history, some of the great people who have walked the hallways of the Census Bureau. Those always have been torn down and replaced with newer hallways.

But those of us who remember the old buildings in Suitland, have a certain amount of affection for, but happy to see the Census Bureau now in really good accommodations.

During those older days, it was really sample-based bias and variance that seemed to dominate what many people were talking about at the Census Bureau, and when I came as sort of a model based guy, I did feel out of place.

I'm happy to see that the terrain is changing somewhat. Happy, but not willing to give up notions of sample-based bias and variance. But there are so many things that we need to do in our data that can only be done in a model-based framework.

And so happy to see that some of the old views, that model-based is not to be used because of its - because we can't trust that are we can't quantify it, but that it's now being replaced and we're simply asking the hard questions and giving the answers based on statistical modeling.

Let's see, where am I? Yes so I got a little bit ahead of myself. There's one small typo on the slide, the very first - that should be just Sigma squared not Sigma squared divided by N. I'll get that fixed and get it to Sara after the advisory committee has finished.

And anyhow, the point that I'm trying to make here is that if we do the aggregation and we think that aggregation is driving down variances, we should be very careful about worrying about biased.

And they could be sample-based or model based biased. And so the notion that big is going to save us might actually be like a fool's errand. Big might simply mean that, as we aggregate and aggregate, we end up with just the same biases before that will not drive down our notion of precision or accuracy. Sorry, more (unintelligible) accuracy.

So this is the slide it was speaking to just a little bit earlier about sample-based and model-based. But now let me move on to notions of accuracy which is usually - means squared error, and to point out to you that bias and variance are in that relationship together.

So if you think that large N is going to get you a better mean squared error, then have another think because it won't be true. If you haven't control of both, on I decent bias and variance, you do not have control of the mean squared error.

Okay so let's move on to - with those cautionary notes and somewhat optimistic viewpoint about what can and cannot be done, let's move on to the new center. I'm very excited about this.

When the slides were sent to me in the slide you saw, there was - just a small comment that they were working on a name, so I threw in my two cents' worth, so that can be ignored and treated with equal misgivings as the name that Ron mentioned.

Anyhow, I just thought I'd spend 20 minutes coming up with seeds. And I know it's not bad but I like yours better, Ron. That's right. We'll have an acronym face-off or something.

So I think the center is going to be a real innovation, provided it gets moving in the right direction. And that really comes back to this notion of goals. They need to be formulated with priority areas within those goals, and then working on problems within those areas.

And that really comes back to the statistical principles and the computational principles that I was talking about. So that requires some more thoughts and it obviously does need a chief which- will formulate those.

It also has to be funded properly, and by funded, I just mean serious FTEs. If we start slicing up people's time into -- I don't know -- 10% or something (around) the center, and then we put ten people on for 10% and we say, oh, is that a full-time FTE?

It's very hard for people to find 10% of their time, so if we're going to put people onto it, they should be given serious - a serious fraction of their FTEs or, indeed, people hired to serve with it.

Okay, so the IMI is a nice way to get some outside influence into the center and they'll be working with the University of Michigan's IRIS Institute to work on big data.

But I would like to caution, it's good to start but I find the start to be more about, like, can we rather than why would we do this or does this fit into some of these principles that I was talking about?

And one should ask, what have we learned? If it's just that we can analyze a certain problem, so be it, but what does it contribute to some of these principles, the statistical design principles and computational principles?

And so I do believe that the center is still looking for a chief and it's also looking for a focus that will be driven by a chief. I like the retail big data project that we heard from Bill. Its goals are clear, to trying to supplement existing surveys to obtain small area estimates were frequently.

And the obvious question is how small and how frequent? But nonetheless, that is a very loadable goal to go down. I do believe this would be a great place to show that inference should and could be model-based it's really ask hard questions like where the roadblocks?

When is that actually being squeezed and I can get through it? What are the problems scalable and when are they not? I think these fundamental basic questions are distinct from did we - how do we achieve it?

Let's try to break it and see if we can fix it. And that's the sort of methodological development that's going to be needed for the big data. So my final slide is that this is a great initiative. It's the - the appointment of the chief should be done tomorrow if not before.

Projects need to be chosen strategically and I think I've made a case that following principles of statistical and computational design need to be incorporated into that to provide focus.

That needs to be staffed with a serious number of FTEs. CSAC - us - we have our place in this and we have a nascent working group which we'll start discussing some issues little bit later during the day.

But we should not ignore the fact that the data will avoid us from discussing uncertainties. On the contrary, we'll be discussing uncertainties. They will not go away. It's quantification is certainly needed whether the data be big or small and the size of the data leads to increasing complexity or discussion.

So thank you very much, both Ron and Bill, for a very stimulating paper. I know I found it stimulating to talk about. So I think you, Willie, and they leave it open to the floor. Thank you.

Guillermina (Willie) Jasso: Thank you so much, Noel. The floor is now open for committee discussion. Let's begin on my left with Dan, continue to Roberto. Raise your hand. Continue with Barbara.

Dan Atkins: Okay, Dan Atkins, University of Michigan. So thank you for that presentation. I have two things that want to say. The first is, I would encourage further exploration of the name of the center.

For me, at least, out of the science fourth paradigm perspective that I come, it's about exploration with data, not necessarily optimization. I'd suggest something like the cen- if you want to cast a very broad net, try something like the Center for Data Exploration and use the X at the end. You can call it CDEX.

I also think CODA, you know, is the end and you're just at the beginning, right? And so, for what it's worth. I'm very familiar with Julie Lane's star metrics work and now the

work at Michigan, the U Metrics work, we should also acknowledge Owen Jason Smith who's the prime mover there.

I think this is actually an example of a very strategic project that I'm glad to see that the Census Bureau is involved in. It grew out of the science of science initiative that Jack Marburger took when he was a science advisor and Julia had a lot to do with shaving that.

I was at a meeting in Ann Arbor last week where we had John Holden and Fran Cordoba and a bunch of other science policy people from PCAS, and there was this kind of a star presentation at that meeting - was a U Metrics project.

And it's viewed as something that would potentially be transformed there with respect to informing science policy. So I'm delighted to see that the Census Bureau is supporting that.

Guillermina (Willie) Jasso: Thank you. Roberto Rigobon.

Roberto Rigobon: So I am on the camp that doesn't like the word Big Data. And given that we're in the Census, may I suggest a name that I heard here in the Census? Rob Groves, some time ago, said he didn't like the word big data. He called it organic versus a design data.

And truly I think that that's a much better way also signal and highlight what are the changes that you have. I mean the design data has great properties that go exactly to what Noel is saying.

They are being designed to make sure that prices are small, that representativeness is guaranteed, that we over-sample things that are hard to sample, et cetera.

So I mean the characteristics of design that is that you have a particular question you want to answer and you design it to try to make sure that those biases are kind of reduced, eliminated.

The problem with organic that is exactly the opposite, it's that because the lags that design, we are bound to make mistakes with the bias. So the challenge that you have is that want to use the properties of organic data which means they are cheap, easy, fast, that captures things that we cannot capture with design data, et cetera, that individuals generated without knowing so there's less disruption and less - you know, less disruption actually from the agent.

However, you want to use those characteristics but you want to convert the organic data into something that is actually design data. I mean, that's a challenge to do the transformation.

So I would suggest that you take the leadership and use your own name. I mean, I don't under- this is the only place that I heard it in the Census and it was about growth so maybe you heard it somewhere else but you can take the lead.

Let me talk a little bit about the retail sector. I know a little bit about retail sector and collecting this data. And actually the MPD data have always been quite worrisome to me.

Let me just give you a little - so when you hear 900 retailers, 150,000 stores and \$400 billion, that sounds like - wow, that sounds big. No, well, Walmart, according to their financial statements, the annual sales in the US are \$280 billion. Amazon is \$90 billion.

So that's two stores, no? Now next, retail sales in the US are \$4-1/2 trillion. That which of these 900 retailers - I mean, they clearly don't have Walmart or Amazon or Newegg or Best Buy.

Actually, they probably have not a single one that matters. These actually are selling half a billion dollars, on average, each of them. And very sympathetic to the projects, but you know, please, please, be very careful with this data in general.

It just - we - it's so overwhelmingly big. This is the issue - that we assign them characteristics and properties that are not necessarily the correct ones. The likelihood that

they are making a mistake or that they're - I mean, this is voluntary participation, so the question is which products is Walmart giving them, all the transactions. Then it should be \$280 billion.

So they're not getting all the transactions. So my point is, and I'm - it's the same public information so there's nothing private here, okay, so it's not like this is all from the financial statements.

So in some sense, my point is - and this goes exactly to Noel's point and it's exactly the same. I'm just saying this a little bit differently, is that the fact that this is big means that we're going to estimate incredibly well whatever bias we have.

And so that's - I mean, I think that's what Noel is saying. Is that correct? So whatever mistake is in the data, we're going to estimate it perfectly, okay? So there is no - so we have to actually put more weight on the techniques, more weight on the models because we know by construction that the data will not be there.

So I don't want this to be discouraging whatsoever. Is that okay? None whatsoever. I want to encourage this. I think this is very valuable. This is what I do for a living every single day, okay, and I actually have access to all three data sets.

And I can tell you there are better data sets. Let me put it that way, in a polite way. If I were more Latino, I could tell you exactly what I think about it. But anyway, so, my point is, they are very valuable, they are very valuable sources and they can answer very important questions.

But usually what happens with this is that you have to start from putting yourself on the other side, thinking that it's actually a very bad data set. So let me think what part of that we can extract because you're not collecting it.

And as you know, my research - actually I ended up collecting my own data. So it should be a good thing now that I decided to collect my own data what I had access to this, because I probably thought that this cannot answer my question.

Guillermina (Willie) Jasso: Thanks. Barbara Anderson.

Barbara Anderson: I found the presentation and Noel's comments very interesting. And there's a great deal of interest in big data, and I'll call it that since that's what everybody calls it. And I think it's good that the Census Bureau is thinking about it.

But I also will underline what Roberto was saying, even though he walked out, I could still agree with him about Noel's comments about bias and selectivity.

Now, there's some very good work in the area of big data and Jason Owen Smith, who's a colleague of mine, has done excellent work on patents and drug stuff and all. He's very careful and often criticism of big data, which I'm not telling you anything you don't know, is it's a mile wide and a quarter of an inch deep.

And I think - and that would underline again the importance of what Noel said. I also would criticize both of the proposed names for the center because they are kind of imperialistic.

There's a wide range of work in social science including in demography, extremely interested for decades, if not virtually centuries, in data quality and good quality data in the names proposed, both of them, I think are entirely non-descriptive.

It sounds like you're taking over the world of everybody that cares about data quality and modeling or doing a good job with anything. And I think you're going to encounter a great deal of well-founded resentment from people if you go with either of these sets of names.

I would disagree with Roberto, who's still out of the room, and suggest you call it kind of what you're doing which is something like Center for Big Data Research and Applications because people may not like the name but at least they'll understand what the world the center is actually about which is fine.

And I think if you can raise the quality of work -- of general work -- in this area and the amount of scrutiny in terms of the statistical issues, I think that you all will be making a major contribution.

Guillermina (Willie) Jasso: Jack Dangermond.

Jack Dangermond: Yes, I would really second that point of make it very practical and what it is in the sense that what you're after is not big data but big understanding. I keep thinking I've heard enough about big data. I'd like to get the big understanding because big understanding can drive actions.

Then in terms of priorities, what is the big data project? I think it is a research project and it's a research center. So what product is going to - project is going to deliver the most value for the Bureau?

And I would say looking at the various censuses, the operational aspects of them, people walking around is going to create a lot of data - spatial data that can be modeled and analyzed to create big insights about the different ways that you're taking the census operations.

And I like actually your name a lot. It's descriptive and it is a research project. I mean, big data is a research project in the Bureau, right? And you're searching for - I think you even said sort of opportunistically, what - where's the goal? Where can I create big insights?

And again, I'll repeat myself and say hunt for operational efficiencies in the Bureau, so Toyota, a big shipping - there's a lot of transit research that's now coming out there really is creating great efficiencies in the big data space and I'd encourage you to look at that.

I am not sure, Noel, if little data and big data are the same thing. In terms of research methods - methodologies in the way that you deal with statistical inference or methods for manipulating that, the way - I'm beginning to feel like big data is just bigger data but you still approach it in the same way with respect to the spatial aggregation or visualization or analytics.

Like, if I had an accelerator machine to be able to deal with big data to make it run out 100 times faster, problem sets that I'd have that lasted five minutes now but would take 500 hours if I had bigger data sets.

I would actually approach the research side or the statistical side or the visualization side or the analytics side in the same way. I mean I'm not quite sure that I wouldn't.

Okay, just as you said, so there's no such thing as big data. It's just - as a category, in other words I'm hunting for a big understanding, whether it's a small data set were a big data set.

I'm able to process a very large data sets because I have a bigger machine are I have Hadoop that effectively hides the concept of being able to do faster - yes, you get the idea. In other words, don't treat big data as a separate thing from a methodology perspective.

I'll just say treated just like - everything that applies in the past, applies in the future. That's what I'm trying to get across, I'd say. There is actually nothing different. So, you know, the way that I'm approaching big data is I'm building a big data accelerator behind a (GIS).

So I still do my GIS but the fact is it runs 100 times faster. That's the basic concept. I still use aggregation. I still use the same statistical tools. It's just the bigger thing behind it.

So the theory, the thinking, the methods process remains same. The other big thing I got very excited about, by the way, was when you talked about spatializing business data. The world needs that, whether it's only a 10% version of it or a 70%, I don't really care.

But spa-I think you would do a huge amount of good by staying on this track, Bill, and you mentioned this before, of spatial aggregation of the data but then getting it out by something other than the whole country.

I would just say that could be the biggest gift for 2015. I'd like to see it in 2015 - that could do, and I'd say if you are processing what project to work on, that should probably be one of your highest ones because being a little, it'll have huge implications, both in the research community and also in the business community.

Peter Glynn: Maybe I'll take a slightly concurring view here with regard to the word big data. I actually think that it is distinctive relative to historical statistics in the sense that these days it's often the case that we're gathering so much more data with somebody more potential explanatory variables that can be used.

And one of the big issues, in some parts of the big data environment, is that people can try to build models that can potentially have millions of explanatory variables in them and that really leads to very, very different environment relative to conventional statistical model building.

So, you know, I think in the context of, say, census, you know, when you think about the basic problem that the census has in terms of counting 300 million people, that's not particularly a big data issue these days.

In the raw amount of data, 300 million people, that's not big in the sense of comparing to, let's say, the types of data sets that cosmologists gather in which they are gathering

enormous amounts of data based on looking at gathering radio telescope signals from the entire universe, that kind of thing, trying to parse through all of that data.

And that's truly, you know, massive data sets that we're looking at. But it is the case, I assume, with some of the things that the census is looking at that potentially the number of explanatory variables could be very, very large. That leads to high dimensional statistics and there are distinctly different things that I think people are exploring these days in that context that may be potentially relevant.

Guillermina (Willie) Jasso: ...respond, and then we'll take a break and then we'll have our next big data session and continue along the same lines and with the same themes. Oh, okay, go ahead.

Nancy Potok: So I just wanted to clarify something that he said to make sure that I understood what he said when you were talking about using large data sets to inform our operations - yes, Jack. No, Jack. So were you talking about sort of using the data that we're collecting through our systems for our, like, interviewing and things like that - that kind of thing?

Because we are doing that and we've set up a special analytics unit in our field director it that is specifically looking at the operational data from the surveys.

And we're looking at it of course through the censuses and through the systems that we put together, both to manage in real-time but also to inform kind of long-term trends to increase our productivity and our efficiency and look at sort of some of the relationships between the cost of our operations in the quality of the data.

And what we ideally be doing is these - both of these units are being set up, one focused on our own production, and the other sort of more focused on what can we learn that informs us some of the products that we're putting out, that anything that we're learning in terms of analyzing the data and the quality of the data, how we measure the quality, the biases, all of those things that would be sharing (that) between those two areas in any case if there's relevant information.

Jack Dangermond: Yes, that's good, Nancy. I'm just thinking if there's a center and there're resources and people are playing with Hadoop and Spark in these sorts of methods, sharing those or dividing them but doing both of them, I just see there're huge payoffs in the operational efficiency that - so I didn't know about that. That's really good.

Probably some of you have seen the taxi cab data in New York in big data. Have you seen some of that information? My guys were just playing around with it a few weeks ago and they made some discoveries. Probably other people have made the same discoveries.

But one of the discoveries was that all the taxicabs pick up a lot of riders around - what was it, Grand Central Station about 9:00 in the morning and they dump them all off at the UN about 9:30 or that period.

And so this is looking at the last three years, playing around with the data. Why not just build a bus system that goes from Grand Central Station down to the UN? Particularly if I was the UN I would say, "I'm going to fund the ta- you know, I'm going to fund a little bus that goes back and forth because I'm paying all these taxi bills."

That's what I call great insight, and it took a few hours of playing around with the data. I'm using that as an operational example because I can simply see tracking all of these enumerators and coming up with some insights about that kind of space/time operational information. And Jackie must be doing stuff like that at UPS, the same kind of thing, but you're not to talk about it.

Jack Levis: I think it's funny, you know, the definitions of the data. And, by the way, I've always felt big data is the how, it's not the what. I always say the same thing - I care more about big insights and big impact and, you know, but it's about the insight and insights that doesn't lead to a better decision is trivia.

So you - there needs to be some exploration and who knows what you're going to find? And we do a lot of that. And 16 million packages a day in the scheme of things is really rather small.

You know, so - but we do a lot and we find a lot of interesting patterns from descriptive analytics and we do some forecasting and then vehicle maintenance. Think of all the data - there're, you know, 220 sensors on one of our vehicles. That's a lot of data that were looking through to see we can predict that a vehicle is going to fail before it fails.

Guillermina (Willie) Jasso: Thanks. Ron and Bill.

Ron Jarmin: So I don't think we want to respond, you know, in detail here but just to thank everyone. And so - you know, and we are thinking about these things as - you know, the reason we're doing this is because we're trying to improve economic and social measurements that we provide to the public.

And so folks want more detailed data. They want more timely. If we were to attempt to use our traditional methods to achieve those goals, it would cost us a fortune and pose a huge burden on the public.

So clearly, some other form of bringing data into the building needs to be done if we're going to give the customers, you know, sort of what they're demanding.

So, like, geographically detailed business data, you're not going to get it with the sample size of the monthly annual survey's you're going to have to do something else. But that brings up to the point of, you know, if you're looking at bias and these things - we don't control these, but do we use, you know, surveys or other, you know, sort of design data collections to try to combat the bias that we might find in some of these things?

So there's a way of optimizing across the two data collection methods that will solve that problem. But that, you know, the - you know, we didn't talk about all the just sort of

privacy issues, so one of the strategic things that we are doing is doing a lot of our efforts on business data right now and not household data.

Because also with the New York taxi cab data, I don't know if you read the articles about how they figured out, you know, which celebrities don't tip taxi drivers very well, which celebrities were at strip clubs and this, that and the other thing, you know, because there is other information that you can link to those things.

And so, you know, I think there is this - you know, there's this desire to try to figure out everything and our challenge will be to stay focused on what the interesting measurement problems that Census Bureau should be focused on, so.

Bill Bostic: I echo Ron's sentiment on the comments. We are aware that MPD doesn't have all the big players. But this is really about, as you all mentioned, research. It's about playing with the data and learning some insights.

We're aware that there are other market research companies that have retail data, so this is how - kind of a first that it playing with scanning data. Certainly we are interested in playing with credit card transactions to get some insight and allow, you know, our staff to learn and grow. So this is the very first important step for us. So we certainly appreciate the comments and we do have, you know, a lot of issues to address going forth.

Guillermina (Willie) Jasso: Thank you very, very much. We now have a 15 minute break.

((Crosstalk))

Woman: Let's reconvene. Let's reconvene, CSAC. I had been under the impression that as we were running late, we would be able to and late. But it turns out that things have been set in motion. The bus is coming. Who knows what might happen if we're not out there, so this may have to be an abbreviated session.

But don't worry because anything that is left unsaid we will continue with tomorrow. That's a wonderful - okay, the bottom line is, it's happening in real time and we don't know yet but you will know when you need to know.

More on the bus. Okay, this is the session on the working group on big data in the working group's plans. I'm going to say a couple of things generically about working groups and then I have asked Peter Glynn, who kindly agreed to lead off the discussion.

I'm sorry. No. Did I ask you? No, I asked Ken Simonson. Yes, I am so sorry. Okay, the key thing, all of you have the standard operating procedures, this document on the standard operating procedures, for CSAC.

Inside the document is a section on working groups. And, of course, you have plenty of time on your own to read all of this. The one thing I will say about working groups right now, because it's very relevant to us, is that the key thing about a working group is that it provides advice to CSAC, not to the Census Bureau.

CSAC then deliberates and makes, if it chooses to do so, recommendations to the Census Bureau. So with that as a background, that's all I will say, and now I will turn it over to Ken to start this. This is only the beginning of the discussion.

Ken Simonson: Exactly right. This is very much the beginning, and I think we really had the beginning about an hour ago with the very helpful comments that so many of you made. And I hope you'll keep this coming to the working group or to the entire committee.

The schedule calls for working groups to meet monthly. I'm not sure that we will try to hit that precisely but we would certainly welcome comments at any time.

For my part, I set up my own personal advisory group of business economists because I think that I have a different set of people that I normally touch from the rest of you and I will be asking them to provide reaction to the things that we hear from the census people and from other members of this committee.

Obviously not items that are meant to be confidential and, for my part, things that I hear from them, I'm going to be asking them each time is it okay to share with this group what do they want me just to retain that for my own information?

But I think we have a lot to do in order to figure out how can census fit into making use of these data sources and how can they overcome the massive institutional restraints on a government agency using sources that have this degree of bias and other limitations that Noel laid out so well.

So I'm hoping that we will have a 21 member working group in terms of getting input from everybody on the committee who has interest in some aspect of this and who has outside contacts who can also contribute to our understanding and that of the Census Bureau.

The terms of an end product for a time for the working group, I'll be looking to Barbara to see when she would like us to come up with either an interim or a final product and put ourselves out of business.

But I think we'll need to explore that as we go forward in terms of what we think would be most useful for helping census get its own efforts to stand up and keep going forward. And really that's all I have to say at this point.

Guillermina (Willie) Jasso: One key point that you made, Ken, is that we really take our mandate from CSAC. The working group takes its mandate from CSAC. So now, and in the future continuing, CSAC members, tell us what you want from us, where you want guidance, where you want new ideas.

I will start off right now with only two things. And one is take - be sure to take with you this sheet which is the preliminary statement on the working group on big data.

You see the members so far, and importantly, the SMEs, the Subject Matter Experts, that is to say, Census Bureau staffers, Ron and Bill. And then I only have one thing more to say and will open it up.

And that is I've been going over and over in my own mind how to distinguish between what big data can do for census and the census mission and what big data can do for science.

And the two overlap but they are not the same. And it will be our understanding of how these two missions overlap that will help guide us in the work and the deliverables that we choose. Okay, the floor is open.

Ken Simonson: Well, let me jump back in for a second. One of the initiatives that I'm slightly involved in in my role as a board member and former president of the National Association for Business Economics, they'll be having a day and a half conference in Boston in June on big data, and in particular, how businesses contribute to it and use it.

Roberto will be one of the speakers they are, but be glad to get you more information on that conference and also some of the information that comes out of it. If you want to see the agenda, go to NABE.com - for the National Association for Business Economics.com.

Guillermina (Willie) Jasso: Thank you. Comments, questions. All right, to break the ice, during the break people were...

((Crosstalk))

Dan Atkins: Did you have...

Woman: Excuse me.

Dan Atkins: Are you asking for comments on your question about the role of big data? Well, there's the role of big data in the service of the mission of the census. There's the general big data in the service of science.

And then there's this kind of middle ground of data resources of the census being used in conjunction with other data resources in service of science or maybe both. Okay, so there's an intersection of the two that's the only thing I would add.

Guillermina (Willie) Jasso: That's very good, the v... (Unintelligible) of the camel format. Then you have seven characters, the magical, mystical number. And with that, I'll turn to Barbara.

Barbara Anderson: Well, it seems to me that one logical thing for this working group to do, and especially since Bob and Ron our members is, among other things, with this new not yet existing center, help them figure out what would be sensible for their plans and for their work plans to be in to be an expert group to consult.

And help them figure out where they're going. Like, I think what they said was fine, but they clearly were asking for input and this working group seems like the perfect group to do it.

Guillermina (Willie) Jasso: More comments.

Ron Jarmin: I think we have questions. I mean, I think what Barbara just said is exactly the sounding board for our own thoughts and another source we don't want this to be a not invented here kind of phenomena we want external engagement that informs the works that we are going to be doing.

Guillermina (Willie) Jasso: Noel Cressie.

Noel Cressie: So let me ask Ron and Bob what they - Bill, sorry. Excuse me - but they learned from their respective studies that you might call case studies. I gather that Ron, you did the educational one and Bill, you did the retail one and you did them just to get going, right?

I mean, obviously you had some things in place. But what did you learn in terms of what the next set of studies might be or were you filling in, like, a table of things that you were trying?

That's the sort of thing that I think we, as the working group, could start having checklists of concepts that we're trying to impose in terms of our principles of - our computing principles and our statistical design principles. So can you perhaps tell us where you - what check - what list you checked off when you did your respective studies?

Bill Bostic: So for the retail project, we just started in February because - it was February of '15 will make up the last data set for 2014. And so the team has been operating for the last couple of months and what I gave was a status, a kind of an update where they are expecting a report to come at the end of May, early June.

So we're kind of in the middle and so we did kind of have to wait to see what they learn from evaluating the data set. We're in the process of trying to get some credit card information. You think we're close to getting some from MasterCard. And we have to play with that.

So we're in the learning process to see what insights that we gain from this project and the feasibility, looking at the quality of the data, for any data product that we are able to produce.

Certainly we have to defend it in the way of thinking of the methodology that we use, et cetera. So Ron might be a little further along in his project but we're kind of still in the early stages on the retail project.

Ron Jarmin: So I guess we viewed the university project is maybe a microcosm of many of the issues that we would encounter more broadly. We haven't - I don't think we've really come across any computational issues just get but we've certainly been grappling with the representativeness of the data.

But, you know, what is it that - you know, in two ways. You know, there's a lot of linkage that goes on here, so not only are we extracting information from a selected subset of the universe that we would like to study, but the linkage issues, you know, so there's a bias from the selection of the university, then there's a bias because of the subset of those records that can actually successfully linked to the Census Bureau data.

So, you know, those are issues that we've dealt with in the past and I think we feel confident that, you know, we know what to do with those. But the things that are more novel here are, you know, I think that this is the first time we've actually sat with somebody who we've asked give us data and asked them the question, you know, "What would you like to see us do with this data?"

That's a novel thing. And then to the extent that, if that had to be a model going forward, or at least a model for a subset of our data providers, the likes the largest companies or

something like that in order to get automated feeds from them, that we would have to, you know - there'd be some quid pro quo.

What would that look like? And how would we scale that effort so that it wasn't something that took, you know, half the FTE of the Census Bureau to accomplish? So I think these are some of the, you know, maybe the not so statistical issues that were, you know, more on the policy side that we've been grappling with so far on this project.

But as we do scale it up and the amount of data coming in the door starts looking more like big data than it does right now, then I think these computational issues will start to rise especially if we're thinking about it in the context of the business sector.

Unless they were getting retail that in on a weekly basis from a you know, from a large subset of the retail universe to do these, you know, timely detailed statistics that everybody would like to see, you know, how do we cope with that?

How do we make sure that when the data feeds from some large retailer, gets interrupted when we - that, you know, right, alarm bells go off and come you know, we don't put out data that looks screwy just because of some (goods).

So, you know, these are sort of - I think we're taking these issues as they come right now. But I think all these projects will eventually, as they mature, show - give us all of the different aspects that we're probably worried about.

Noel Cressie: Can I just follow-up? I mean, you talked about getting data in, Ron, but what about if the data residing in different places on different computers, they get into the notion of distributed archives and the cost of moving data and how you might handle that.

I just wondered if you had explored that aspect, if it's more like Bill's project, it's more about that and yours is more about getting data in. And if so, what have you learned from that?

I always think if you can break it then you understand it. And it sounds like you haven't - you know, haven't stress tested it enough yet. I recognize that it's early days but the notion of starting small or smaller than trying to break it, but then sort of pushing it to its limits, has some real benefits.

Looking at different - the data residing in different places I think is important. Learning from the two projects you've started, what methodological problems haven't you been able to solve do you think they're going to come up in other problems of a like nature?

I just think from what you know how to do and the linkage areas that you've got good experience with. The idea of bias, not actually just sitting there but you might try to model the bias and give it a hierarchical structure and put some - because it's uncertain, give it some uncertainty.

And then he sensually apply some form of base rule to get a hold of what you're trying to get in the presence of bias. Instead of being stuck with it, you try to actually model it.

Other notions that I deal with in my remote sensing work is having a strong calibration and validation component to my big data sets which involves - I don't like the word ground truth because, you know, the truth is something you can never get perfectly but they notion that you've got some really high quality data which you can anchor your analyses based on big, imperfect, noisy data, missing and all that sort of thing.

These are principles which, you know, a lot of people working in scientific endeavors try to harness and apply. And then in my discussion, those notions of false positive- well, I didn't mention it but, you know, we have some pretty nice measures of how we do false positives, false negatives, the calibration validation data gives you.

I mean, we can talk about all those sorts of things and guide in general but the fact that you actually are starting to do things, to me, it's a great blessing for us of the working group because it's not meant to be critical but it's just meant to see, well, now you've got your sleeves rolled up and you're actually doing something.

Let's see what it is that we can do well and what it is that we can't do well and maybe that provides us a forum. So that's - perhaps the idea is to hear more about some of those two projects and for us to suggest other projects, if you like, look at the opposite side of the coin.

Ron Jarmin: So I think that would be perfect. I mean, partly that's why we started these things. Some of it is, you know, you know that you'll discover the things that you need to be looking at when the problem presents itself.

So rather than sitting around talking about them, we thought it was important to get at least some stuff underway. So we would really appreciate help with sort of the critical review of these things so that we know that when products to start rolling out, that they have some veracity to them.

Bill Bostic: I - you know, I think that even insight on other potential data sets. Jack mentioned that, under taxicab the data sets - because a lot of - I mean we're aware of kind of retail with the scanning data in credit card but, you know, we ask information for hotel rooms.

Well, there's information out there about hotel rooms. That's information that we could take off and not collect on our collection instruments. So even some insights of other potential data sites that we should examine and investigate would be helpful to us in this process.

I think looking at a data set, and we're still talking about leveraging administrative records, and so the whole notion of compiling statistics from various data sources and how - you know, methodologically how do we handle that? Those are challenges that we will have to address because we feel that is the direction that we will be going in.

Jack Dangermond: I have a general concept - you started it, Noel, with this notion of do you actually bring all the data into one big data environment or do you deal with this notion of

distributed nodes with addressed end points, they you do some kind of a virtual integration and processing with?

That architecture is still emerging and there're both games going on in the technology space as you probably know. So, yes, looking at models for how people approach this is interesting.

One of the things -- I'll return to my field, spatial -- one of the things that we're finding is that spatial is a really interesting way to integrate things. So you have distributed data sets and how do you bring them together using common items or common relate keys?

Geography has its interesting dimension of being quite common to almost everything. So it's a way to build an architecture. And if you are going to go down the route of modeling multiple architectures like that, that would certainly be one thing to explore. I had another one but I forgot what was.

Ron Jarmin: So I think we be interested in meeting with the committee and doing more of an academic seminar style presentation of some of these things than we typically do in an advisory committee setting.

Jack Dangermond: I have one guy that's actually probably one of the lead people in the world. His name is Monsiour Raad - R-A-A-D - and I would volunteer him for a day in such a session because he's working probably with 20 customers, users, and the big data environment. And he has just a whole - he's developing a whole vocabulary for how to approach this from an architectural perspective.

He's working with all the open - he's working with all the platforms so you can give you some sort of high-powered advice. He gives workshops once in a while in DC here so piggybacking on them.

Tommy Wright: Willie's looking at me because I was trying to chime in a little bit earlier. Tommy Wright. Along with Ron and Bill -- not Bob -- there's an effort behind the scenes kind of

to draw a Venn diagram actually and I think I rediscovered that you have great difficulty if you try a Venn diagram with four circles.

It is very difficult to do. So let me just make a comment in response to - it's difficult to do so we went to the use of Ellipse, so the idea is imagine - and thinking about space and time, imagine a small - imagine a geographical area with -- well, excuse me -- what's in the universal set?

What's in the universal sets are retail sales. So that's the element that's in the sale, and so there are four sources that are on our minds. One, of course, is the Census Bureau and one, of course, is maybe perhaps credit cards.

And another is the people who control the machines that you swipe. And then the fourth category are people like MPD, for example. And so the thought is we use - we're using the word exploratory research so this is a key word.

We want to know what are the - what's the size of these 22 regions relative to each other in terms of overlap. Where the gaps? What's the overlap? And because we want, perhaps, these companies to become partners with us, we use the word supplementary as Bill used in his talk.

But we're using complementary on the slide that we're (joining), is how can we complement - how can we have a complementary role perhaps with these partners that we might want to think about what MasterCard gain out of a partnership like this?

But I think it's impossible to do a Venn diagram with four circles. You better use than Ellipse. So it's very - as Bill said, it's very - that may be a focus, what are the sizes of these gaps - exploratory, to get some data, try to mention compare, very exploratory in nature at this phase - a part of that. And that's a part of the retail sales sort of conversation.

Noel Cressie: No, I like to do categorical rather than vin diagrams and I think you could look at sales then and with sales, you're filling in...

((Crosstalk))

Noel Cressie: What sales and completing as you complete your research agenda? Some sales won't be relevant at all.

Tommy Wright: Right.

Noel Cressie: And other sales will be. And I guess I was just wondering which sales are being checked off by Bill and Ron's studies and what sales are remaining.

Tommy Wright: I understand that there has been - some people have been thinking about maybe taking out some of those spaces. You're right. Exactly. But we're doing - we're going through several drafts of this vin diagram.

Man: So maybe this was mentioned when I was away for the last half-hour but it's a little bit of Tommy's diagram but really a social network. At least I view - but say, that specific example of credit card data, our data, so on and so forth, is one of the goals is to say is there some latent truth and some latent attribute of either a region or whatever that we're trying to estimate or measure?

And how can we bring to bear on that? I'm going back to traditional values - some kind of a latent variable model where we have all these measurements that are - have very different kinds of prominence and sampling plans and quality, but can we figure out a way to put them together to estimate the it? And it's not even trivial to identify the it but I think it's step one in some sense.

Guillermina (Willie) Jasso: More discussion? Yes, Jack Dangermond.

Jack Dangermond: I remembered the second thing I was going to ask, and that is, is the Bureau going to create an American Business Survey like it does with the population survey? And if so, this whole big data exercise is really very supportive and interesting to me.

I mean, if you - if that was the vision for the mission that we really wanted to have an annual business survey which was, you know, recording every American business and the pulse so that it could feed America's input-output model so we could have really econometric modeling of economic sectors.

And, you know, we really have the weather - it's equivalent to what the weather service does with modeling. You know, the measure with satellites and then they run it through these models. And today we benefit - all of us benefit because we understand when the hurricane is coming or when the weather is, you know, rain or drought or whatever it is.

And it seems like if we look at it holistically nationally, we need something that actually protects at the micro level, and that means we need to have a systematic comprehensive survey of business that feeds something like let's just say input-output models that gives us forecasts on an ongoing basis.

So we get out of this - oh, I'm surprised or the sector drop, like, what's the real implication in all the rest of the sectors of the oil price reduction? I mean, it impacts Jack's business in ABC and that has (reverberation) on this.

And, you know, that whole - so, it seems to me that what your research is playing around with is coming up with how you actually ETL the data in or connect to different sensors, public-private, bring the information into an environment so that we can actually feed models they give us some sense of predictability of the economy. And I guess that's both the comments and also kind of a leading question. Is this what you're really up to in the Bureau?

Bill Bostic: Funny that you mentioned that, Jack. I've been floating this idea for the last year, that we should explore and research coming up with an American Business Survey similar to ACS in concept.

We have our annual surveys. They're sector-based. Operationally we use different reporting units. And my thought process was if we had a holistic survey that addressed the economy as a whole that we'd probably create a lot of efficiencies and could produce geographic data far more frequently than what we do now.

Certainly that is - we have a contract with the National Academy of Science to look at our sample revision process and to also consider an alternative, as such, to explore. But I've been floating this idea for the last year, that we need to do a lot of research and perhaps move in that direction.

Jack Dangermond: I mean, we do have resistance by business people and Congressional interest not to fill out more surveys by business people, right? But if you gave them - sort of this give-get proposition, if business - the business sector got geographic area forecasts, that same, "How can I subscribe?" And they would want us to do it just like we all want NOAA to, you know, to build a whole weather forecasting environment.

So it takes on a different wrinkle for both the Department of Commerce and Census is the instrument owner for this, but it's a great vision. And I would simply predict that in the next ten years this is going to happen and that my interpretation of your - let's just say, fooling around with big data, playing around with big data business and the subset is all about that. We just didn't really realize it. That's an assertion, by the way.

Ron Jarmin: So to build on what Bill just said, to follow up on your remarks, so the - changing how we do surveys for complementing that with administrative and other big data sources, I think what Bill and I have talked about is creating the ability that eventually the BEA would be able to re-benchmark the NIPAs on an annual basis, that we would have enough source data, frequently enough for them to be able to do that.

Jack Dangermond: I mean the point is if you don't do it who else who is

Ron Jarmin: Right.

Jack Dangermond: ...in the position to do this with America's confidence is it going to be the in the private sector no. It's not going to work right.

Ron Jarmin: And BEA would like us to be able to do that for them so that are interested of course.

Tommy Wright: I mean, right, given that the bus has left six minutes ago, we'll all just start running. Let's end. Thank you everyone for today's presentations and discussions and we'll catch the bus in the morning at 7:30 and we'll start at 8:30. Thank you very much.

END