

NWX-US DEPT OF COMMERCE

July 6, 2021

12:00 pm CT

Coordinator: Welcome everyone to today's conference call. At this time your lines have been placed on listen-only for today's conference. Once again all lines are on listen-only for today's conference call. Conference is being recorded. If you have any objections, you may disconnect at this time. I will now turn the conference over to our host, Mr. Michael Cook. Sir you may proceed.

Michael Cook: Good afternoon and thank everyone for joining us today. I am Michael Cook, Chief of the Public Information Office at the U.S. Census Bureau. This is the first briefing for the National Urban League plaintiffs in the case of National Urban League versus Gina Raimondo.

The Census Bureau has agreed in a joint stipulation in settling this lawsuit to brief plaintiffs every two months to allow an opportunity for question and answers regarding the status of forthcoming data quality metrics and assessment of released data quality metrics.

A final briefing will occur after the release of the Post-Enumeration Survey final results.

Today you will hear from Michael Thieme, Assistant Director for Decennial Census Programs, Systems and Contracts; Deb Stempowski, Assistant Director for Decennial Programs, Operation and Schedule Management; as well as Michael Hawes, our senior advisor for Data Access and Privacy in our Research and Methodology Directorate.

Today's briefing is open to the plaintiffs, their lawyers and the public - albeit only the plaintiffs and their attorneys will be able to ask questions after today's presentations.

Media and members of the public you can find more information in our online newsroom at [Census.gov](https://www.census.gov). You can access today's presentation in this briefing's electronic press kit.

Let's get right into the presentations. I'll turn it over now to our first speaker.

Michael Thieme: Good afternoon everybody. This is Michael Thieme, Assistant Director for Decennial Census Programs for Systems and Contracts. And I'm going to walk you through sort of our processing flow and talk about where we are today.

As this graphic shows, you're seeing the graphic with the chevrons at the top and the green box. That's everything that we have done so far. But I'm going to go a little more - into a little more detail about it.

As the graphic shows, there are several steps to post-collection processing prior to delivering redistricting data. Each step of the process builds on the previous one in a sequential order. And during each of these steps, we process data and then deliver the results to subject matter experts who conduct data analysis and data validation.

When their review discovers possible data issues, we resolve them and iterate the testing and review process until the data issues are resolved and validated.

The initial step to process redistricting data after the release of apportionment results back in April is the production of the census edited file -- or CEF as we call it -- you might recall that the apportionment counts were based on the previous step which we call the census unedited file, or CUF.

So the census edited file provides much more granular data required for redistricting. And the census edited file we process the detailed information about individuals living in households as opposed to the simple population counts that we need for apportionment. This includes complex processing for the race, ethnicity, sex and age information that states need for redistricting.

So I'm really happy to report that our CEF processing successfully completed as scheduled on June 23rd. So we're already a couple of weeks out.

Next slide please. Okay, thank you. So this slide shows the schedule for all the 2020 census processing steps. As you can see the next stage of processing after the CEF that I just talked about is production of what we call the disclosure avoidance application and the microdata detail file or MDF. This stage completely revolves around keeping your data private and protected.

Microdata detail file processing and review started on June 24th right after we finished the CEF and is scheduled to complete on July 17th and we're currently on track for that date.

As you know, the Census Bureau is required by law to protect the confidentiality of information provided by respondents. For the 2020 census,

we're using a methodology called differential privacy to provide the best possible privacy protection available. My colleague Michael Hawes will provide you with some additional information on the production of the MDF and the privacy protection it provides later in this briefing.

In the MDF stage, we process the nation as a whole and we align - and again we include time for review and possible rerun to ensure the system is functioning as designed and that your data are protected.

After the privacy-protected microdata detail file is complete, we're ready for tabulation. Here we tabulate the data into easily understandable and usable data tables for the public. This process accurately integrates geography, population and characteristics in a huge number of combinations. The sheer volume and detailed nature of tabulation requires a thorough review and validation that takes nearly a month. And it's critical and it's a critical step that we need to take to ensure that the accuracy of these products is out there.

Our tabulation processing and review will be conducted from July 19th through August 16th.

The final step on the road to redistricting is the redistricting release process. This step involves the creation of materials to send to the states, the loading of those materials on our Web-based systems, testing to ensure those Web sites function and display correctly and the actual delivery of the physical datasets on DVDs and flash drives.

Our first release of 2020 redistricting data is scheduled for August 16, 2021. This first release will provide the states the legacy file that - what we call the 2010 format summary files. This will allow the states to get a jumpstart on redistricting activities prior to the physical media release, which

is the DVDs and flash drives I talked about which is scheduled for completion by September 30th, 2021.

So now I'll turn it over to my colleague, Deb Stempowski, to talk about some of the other things that we're doing. Deb?

Deb Stempowski: Great. Thanks, Michael. And pleasure to be back with everyone today. I'm Deb Stempowski and I talk about data quality frequently and I look forward to giving you this brief update today.

So we've talked about quality a bit already. But a critical component of that work is the engagement of our external expert groups as part of our commitment to transparency.

Our strategy for working with external experts was designed to provide assessments on different aspects of the program and on different timelines.

First, our quick turnaround feedback on processes, procedures and metrics was provided by the JASON Group. That's an independent group of technical experts that advise the federal government. We did publish their report on their Web site - our Web site rather in February.

But just as a reminder, a couple of highlights of their recommendations and conclusions. First, the JASON supported the Census Bureau's proposed metrics as useful to evaluate data quality and they provided recommendations for additional analysis such as an assessment of data quality across geographies and for relevant demographic groups.

They called for the implementation of - improved communication strategies that explain the census in a way that builds confidence and also is done in

plain English. Our blog series, which you may have looked at over the past number of months, our associated briefings, press releases and the announcement of key dates in advance all reflect our agreement and support of this recommendation. And I hope you've seen we've shown that over the past handful of months since they did that recommendation.

They also recommended that we ensure our post-data collection processing schedule allowed adequate time for complete, accurate and the transparent processing of the data which we've already done and Michael just reviewed.

And then finally they did stress the importance of leveraging this experience as we get underway with 2030 planning.

Moving on to the second component of that expert engagement, that's what I call a real-time assessment. We have been working with a team of experts from the American Statistical Association on Quality Indicators. And when I say "real-time," I mean they're on a similar pace as the bureau is. We're looking at things kind of at the same time and they are posting their own regular updates on the ASA Web site and they are slated to release their first report in the coming weeks.

And then the third bullet there the longer-term assessment is something that actually we're working on now. When I first started doing this update, that wasn't something that we were working on each day but now we had awarded this contract and we had our kickoff meeting on May 3 to get underway with the Committee on National Statistics as we establish a consensus panel that'll conduct an evaluation of the quality of the census.

And that will run up to 24 months. So you can see it's a much longer-term assessment and that time frame is actually very similar to how the Census

Bureau does its regular assessments after a census. So we actually have a full week of initial meetings with that panel coming up later in July.

Go to the next slide. There you go. So in addition to the external assessments, the Census Bureau is going to continue its own work on data quality. As you likely know, we released demographic analysis results back in December of 2020. Quick back. There you go.

We also released for the first time operational quality metrics on the same day as apportionment and we released the second table of metrics at the state and national level at the end of May.

We will also release a set of operational metrics around the time of the redistricting release.

And then lastly, as we've always done and always assess the quality of the census, we evaluate how well the census operations were conducted. This isn't new to those of you who are longtime census followers. As part of each decennial census since 1950, the bureau's incorporated testing, evaluation and experimental program to evaluate the current census and then to facilitate planning for the next census.

Now we can go to next slide. And to close here, I wanted to redirect everyone to our quality Web page. I hope you find it easy to get to different spots on there to find different releases that we have. If you haven't visited yet, I certainly encourage you to do so. There's more than just metrics there. There's all kinds of technical documentation and FAQs. But most importantly, that's the place to go as the Census Bureau stays committed to letting you know what we know when we know it.

And so with that, I'm going to turn it over to my colleague, Michael Hawes.

Michael Hawes: Great. Thank you, Deb.

So as Michael Thieme mentioned a few moments ago, we are currently in the middle of the disclosure avoidance and production and analysis of the microdata detail final phase of production.

As you may know, the Census Bureau has modernized our privacy protections that we use for the decennial census for the 2020 census. And our new approach is based on a mathematical framework known as differential privacy which allows us to calculate and very precisely quantify the privacy risk associated with each and every statistic that we're going to be publishing.

And by quantifying that risk, it allows us to then finally calibrate the amount of statistical noise or uncertainty that we have to inject into those statistics in order to protect privacy and in order to ensure fitness for use of the resulting data.

And central to that process is the setting of the various parameters that are necessary for the system. This includes setting the overall privacy-loss budget that we'll be using for each of the data products we'll be producing as well as allocating that privacy-loss budget across the many individual tabulations that comprise those releases.

On June 9 of this year, the Census Bureau announced the key parameters that we will be using for producing the 2020 census redistricting data summary file.

Next slide please. Setting those parameters and determining that privacy-loss budget and its allocation the Census Bureau's Data Stewardship Executive Policy Committee considered a wide variety of factors. Chief amongst those was the expenses feedback that we received from our data users on a set of demonstration data using 2010 census data that had been run through our disclosure avoidance system that we released back in April.

As I said, we received a wide array of feedback on those demonstration data and there were a number of common themes that recur through much of that feedback and analysis that we received.

Some of those major themes that our Data Stewardship Executive Policy Committee considered were overall accuracy for American Indian and Alaska Native tribal areas and other geographic entities that exist off of our central tabulations plan, overall accuracy for places -- both incorporated places and census designated places -- minor civil divisions and for tract level data.

There were some recurring concerns about lingering bias in the resulting data due to privacy protections, both geographic bias -- urban versus rural, et cetera -- as well as characteristic bias. So different patterns of error for more diverse or less diverse areas.

There was feedback on the overall accuracy of statistics about race and ethnicity as well as feedback on accuracy for occupancy rates at various levels of geography.

Next slide please. So as I mentioned, our Data Stewardship Executive Policy Committee gave substantial attention to the feedback that we received and factored those viewpoints and analyses into their decision-making about setting the parameters for our disclosure avoidance system.

The privacy-loss budget that was set for the redistricting data file does reflect our efforts to ensure fitness for use and to correct issues that our data users had identified in the April 2021 demonstration data.

Some of the key parameters that reflect these changes are an increase in the overall privacy-loss budget. So for those who are familiar with the framework of differential privacy, the privacy-loss budget for the redistricting data summary file is an epsilon of 19.61. This is a significant increase towards accuracy over the privacy-loss budget that was reflected in the April 2021 demonstration data.

That privacy-loss budget is split between an epsilon of 17.14 for the persons level file and epsilon of 2.47 for the units file.

Also included in the parameters that DSEPC approved were significant improvements to the optimized geographic post-processing hierarchy that our disclosure avoidance system uses. This helps to ensure greater accuracy for American Indian and Alaska Native tribal areas.

We allocated additional privacy-loss budget to population counts at a variety of geographic level, additional privacy-loss budget for race and ethnicity statistics to improve their accuracy and additional privacy-loss budget to occupancy rates at the block group levels and above.

Next slide please. Over the coming weeks we will be releasing additional information about these parameter settings. We actually just last week released a detailed breakdown of the privacy-loss budget parameters that we used. Later this summer we'll be releasing the production code for our

disclosure avoidance system so that data users can go in and see exactly how those parameters were implemented.

And we will be releasing one final step of 2010 demonstration data using these production settings where we will run 2010 census results through the disclosure avoidance system with the exact same parameters that are being used for the 2020 census redistricting data summary file.

Taken together all of this information, these additional parameters as well as some detailed summary metrics that we brought against the 2010 data, the production code and this final set of demonstration data will allow our data users to really gauge and assess the overall accuracy and fitness for use of the 2020 census data by being able to compare 2010 results protected with the disclosure avoidance system against what we published in 2010.

Next slide please. I'll turn things back over to Michael.

Deb Stempowski: I think I'll take it, Michael.

Michael Hawes: Oh, sorry, Deb.

Deb Stempowski: That's okay. There's a lot of Michaels on this call but I - there's just this Deb who's going to bring us home and just remind you of what we'll be working on over the summer and looking towards that August 16th release that Michael Thieme talked about where we'll release the legacy format data for redistricting and then we'll make it more user-friendly working through September. Same information in both releases and then we'll also be developing the quality metrics release information with a first release or a release shortly after August 16th.

And so now I think we'll turn it over to Michael Cook for questions or to facilitate the questions.

Michael Cook: Thanks, Deb. And as you duly noted, there are a number of Michaels on the call today. But I'd like to thank Michael Hawes, Michael Thieme as well as Deb Stempowski.

But before we begin to take the questions, I'd like to remind everyone that only the plaintiffs and their attorneys will be able to ask questions today. For those who - of you who are asking questions, please announce your name and who you are representing.

And as a reminder, everyone can visit our Web site Census.gov where you can access today's presentation within this briefing's electronic press kit.

Operator we'll now begin taking questions.

Coordinator: Thank you, Mr. Cook. At this time if you would like to ask a question, please press star 1 on your touchtone phone. Please ensure that your line is unmuted. Please record your name and your organization to be introduced to ask your question. Once again, it is star 1 at this time. Please stand by for questions.

Michael Cook: And while we're waiting for people to ready themselves in the queue, just another reminder, if you're not familiar with Census.gov, when you go to Census.gov, on the top of that page you can navigate to our newsroom. This is the page that says "Information for Reporters." And when you get to that page, there is a link that'll take you to our electronic press kits.

And the title of today's presentation is the title of that electronic press kit where you'll be able to see all these slides that have been presented today as well as the news releases that have been put out for today's news conference. So just wanted to remind people how to navigate through our Web site to get to that information.

Operator do we have any callers? Any questions?

Coordinator: Thank you, Mr. Cook. Yes we do. Thomas Wolf with Brennan Center for Justice. Your line is open for your question.

Thomas Wolf: Hi, good afternoon. Could everyone hear me?

Michael Cook: Loud and clear, Thomas.

Deb Stempowski: Yes.

Thomas Wolf: All right. Thanks for hosting this session. We really appreciate it. Calling from the Brennan Center. Just a few questions. Most of them relate to Slide 4. I was wondering if folks could explain a little bit more the concept of the consensus panel with CNStat. Among the questions, what institutions or entities or people will be represented on the panel and how will this study differ from the study that the ASA panel is currently working on?

Michael Cook: Thanks for that line of questions. And to help facilitate the response, I want to ask, could we - is it possible to get back to Slide 4? Give us a moment. And then we'll have our subject matter experts walk through that response for you.

Deb Stempowski: Okay. I think that I'm the subject matter expert on that and thanks for the question. So let me do this 10,000-foot strategy and let's start with the

bottom, the consensus panel with NAS. That's a process they frequently use. It's something the Census Bureau has used extensively where a group works together. And this is making it sound like not a lot of work. It's a huge effort to come out with a report that is released by a panel of experts.

Those folks that'll be serving on this panel will be named and their names and bios will be published this week. So you can take a look at that for yourself and we can certainly direct you to where that link is.

And I'll say when we close out on the NAS panel, which is saying this is something that the Census Bureau does frequently and we've done it not only for the 2020 census, done it for other things. So this is something that's here but not unusual where the work we are doing in the second bullet with the American Statistical Association in that task force is something that we've added in after the ASA released a report in October in their report and I really would like them to speak more on their behalf than I. They're doing something that's obviously shorter term, right, than the 24 months that the CNStat panel would take.

We're providing them with data tabulations and working with the team of statistical experts, if you will, that they have that service sort of their data team and they will put a report out and that will be done on their schedule through their processes.

I think I answered all of your - all the parts of that question. I wrote a few notes. I just want to make sure I didn't leave anything out.

Thomas Wolf: Sure. I appreciate it. I was wondering if you could say just a little bit more about what's going to make the ASA report different from the CNstat report beyond just CNstat running for a longer period of time?

Deb Stempowski: I think it's premature for me to answer that very specifically right now. The contents of the ASA report and the specific topics that they will cover in that release I think is what they will announce and we've been running that work that way.

So I think probably in the next update we could compare it a little more - at a more detailed level what's in there. But I would direct you for the ASA. They did release their initial report in October on their Web site. That gives you an idea of what they felt was important at the time to look at to assess quality.

Michael Cook: Thanks for those questions, Thomas. Operator do we have our next caller?

Coordinator: Once again if you would like to ask a question, please press star 1 on your touchtone phone and record your name.

Michael Cook: And while we wait for people to ready themselves to ask questions, just one other quick reminder. If by chance you have not yet signed up or opted into getting e-mail alerts for releases related to the 2020 census, I encourage you to yet again visit the newsroom and click on "Subscriptions." And when you click on that link, you can subscribe to any number of topics that the Census Bureau releases on a daily basis, weekly basis or monthly basis.

Operator do we have our next caller?

Coordinator: Yes. It's from Ajay Saini with Lawyers' Committee for Civil Rights.

Your line is open.

Ajay Saini: Thank you and thank you all to the Census Bureau, folks on the call for having this briefing. I just had a couple of really just quick questions. The first one relates to Slide 5. I understand that you all are planning to release some operational quality metrics with the redistricting data product in August or September of this year. I was wondering if you all could provide whether you all were planning to release information, the quality metrics at geographic levels below the state level. I understand that the earlier operational quality metrics were at the state level. So that was my first question.

And then the second question was relating to the demonstration product, the differential privacy demonstration product that - the data that was referenced on Slide 10. I was wondering if there was a timeline for that if there was a sense of whether that would be released at the same time redistricting data is released or is that something that you all are anticipating releasing later on in the fall.

Michael Cook: Thanks for those line of questions, Ajay. We'll go ahead and first start with Slide 5. We're sitting on it right now. And then after we address your questions in relation to Slide 5, we'll slide on over to Slide 10 and address that. So thanks, Ajay. Now I'll let the SMEs take over.

Deb Stempowski: Okay. So in terms of our upcoming release of the quality metrics, the answer is yes, we'll be releasing summary metrics at that substate level. We know that stakeholders are interested in that detail. So that is indeed the plan.

So I think that answers short and sweet. I'll pass it over to Michael Hawes for the second part of the question.

Michael Hawes: Sure. Thank you, Deb. So the demonstration data products that we've released has - typically contain two components. The first component is an

extensive suite of accuracy measures and fitness for use measures and metrics that have been informed by the various use cases that have been provided to us.

The second component is the actual privacy-protected microdata files. So those are the privacy-protected microdata that are output from the disclosure avoidance system.

The first of those components, the detailed summary metrics reflecting the production settings we've already released. Those were released publicly last week and those are available on our Web site. You can just go to Census.gov and search "disclosure avoidance" and you can find those.

The privacy-protected microdata file, which is the much more extensive component of that, is currently slated to be released in September. I don't have an exact release date for that but it's intended to be parallel with the September 30th redistricting data product or shortly in advance of that.

Michael Cook: Thank you, Deb. Thank you, Michael. Operator do we have our next caller?

Coordinator: Yes we have a follow-up from Thomas Wolf.

Michael Cook: Hey, Thomas.

Coordinator: Your line is reopened sir.

Michael Cook: Hi, Thomas.

Thomas Wolf: I may have been muted. Can you hear me?

Michael Cook: We can hear you, Thomas, loud and clear.

Thomas Wolf: Right. That's my fault (on my headset). Going back to Slide 5, I was wondering if folks could say a little bit more about the subject matter, the assessments and the evaluations. What sort of subject matters are going to be covered in those studies?

Deb Stempowski: So the assessments are a broad suite of everything you can think of in terms of how our operations run. So where I draw a parallel to what we did in the very first release of the quality metrics, which we're showing you, I'll call it with the lens being at the US and the national and state level, what our assessments and evaluations do is take things and look at them very specifically.

For example, a detailed assessment on the non-response follow-up operation on its own, a detailed assessment of group quarters, advance contact and enumeration and so forth.

So I will call it this is very, very detailed which is why they take longer to get out the door. Those first releases coming out will be on our - both in office address canvassing operation as well as our infield address canvassing operation.

Michael Cook: Thank you, Deb.

Deb Stempowski: Okay.

Michael Cook: Operator do we have anyone else with questions?

Coordinator: Once again, if you would like to ask a question, please press star 1 and record your name at this time.

Michael Cook: And while we wait for callers to ready themselves in the queue, just to remind the public and/or rather the media who's on the line that's listening, if you have any questions about today's briefings or plan to run news stories and you need clarifications, I encourage you to reach out to the Public Information Office at pio@census.gov and we'll be ready to answer any of those questions that you might have and to address them in a timely fashion.

Operator any callers?

Coordinator: No sir. At this time I show no further questions.

Michael Cook: Thank you. Well, just a couple of more reminders before we close, you can sign up again for subscriptions in our online newsroom to receive alerts about some news to stay up to date about future briefings and the vast information that is released about our nation's people, places and economy. Again for those members of the media, I invite you to contact the Public Information Office at pio@census.gov with any questions you may have about today's briefings.

I'd like to thank everyone. And on behalf of our speakers today, Deb Stempowski, Michael Thieme and Michael Hawes, I'd like to thank everyone. And this concludes our briefing for today. Thanks everyone.

Coordinator: This does conclude today's conference call. We thank you all for participating. You may now disconnect. And have a great rest of your day.

END