National Academies of Sciences, Engineering, and Medicine

2020 Census Data Products: Workshop on Demographic and Housing Characteristics Files

Director's Remarks as Prepared for Delivery

June 21, 2022

- Good morning, everyone.
- I'm tempted to say, "And so we meet again."
- I see so many familiar faces and some new ones, too…it's heartwarming.
- Seems like lately I've been speaking at a National Academies event monthly.
- And that's all good because these events provide important stakeholder feedback.
- So I look forward to our discussions today and learning about the use cases.
- The topic of the day is our Demographic and Housing Characteristics File, or DHC.
- This product includes data on a number of demographics such as age, sex, race-ethnicity, household characteristics and so forth.
- And as you're aware, we're using a new disclosure avoidance methodology involving differential privacy that – it's fair to say – doesn't enjoy a consensus among our stakeholders.
- It's a methodology that most would agree is much more advanced than the data swapping methods used in the last decennial.
- And I think we all agree that differential privacy is far more protective against disclosure than previous methods.
- It does come at a cost of diminishing some of the granularity that data users have enjoyed.
- It's also very complex and computationally demanding.
- It takes time and user input to find the right balance of geographic accuracy and demographic detail while still preserving confidentiality.
- The result is a DHC production schedule that has been extended by many months.
- Now, we know that differential privacy operates by adding noise to census tabulation cells, and it does so in a measured and a transparent way.
- The approach addresses our 21st century confidentiality threats, which over time will only grow in sophistication.
- And because of our obligation to protect census data, we believe this approach gets ahead of these threats – not just for today but for the coming years.
- It can meet the dual objectives of disclosure avoidance and data granularity and accuracy that meet most data user needs.

- Our subject matter experts are present and eager to discuss this and other issues over the course of the workshop.
- But there's something else I'd like to do before we begin.
- Many of those in attendance are seeing me in my new role for the first time.
- So as *your* director of the U.S. Census Bureau, I also want to talk generally about my goals.
- As I've said now on many occasions, we are committed to producing data that reflect an accurate portrait of our nation.
- And we do that by continually striving for excellence, which can be challenging in our current environment.
- But even in the face of challenges, I believe that we can find new ways to collect data that are more accurate, timely, relevant and most of all have utility, as in being fit for use.
- As many of you know, I am a leader who champions diversity, equity and inclusion. I call these principles DEI.
- I've lived them over a 40-year career, even before anyone called it DEI.
- I've seen firsthand that by embracing DEI principles, an organization identifies innovations and advances excellence.
- An important part of our DEI effort at the Bureau includes seeking out and listening to diverse voices.
- That certainly includes listening to our stakeholders, which of course includes you.
- Stakeholder feedback is critical to understanding the needs of our data users. And with respect to this workshop, your feedback is essential to how we protect our data from the very real threat of disclosure, but in a way that also addresses your needs.
- Data relevance and utility must be accounted for in the application of disclosure avoidance. That's why we need your feedback.
- This brings us to the issue of transparency and its role in communicating data quality.
- The Census Bureau will continue to be transparent about the quality and reliability of the statistics we publish and the data we release.
- As you know, no statistical product is ever perfect; all are subject to various degrees of uncertainty and error.
- We apply extensive quality checks and reviews to reduce errors so results are fit for their intended uses, such as apportionment and legislative redistricting with 2020 Census data.
- Thus far, we've released several 2020 Census quality assessments and some are ongoing.
- They include reports on operational quality metrics, demographic analysis, and the Post-Enumeration Survey.
- We also hosted three independent studies by groups of outside experts including an ongoing assessment by CNSTAT.
- This wealth of information facilitates a better understanding of the quality of the data that were produced.
- And make no mistake – knowing the strengths and limitations of any dataset actually increases the value of the data, because you know where to be careful with your analysis and how to better interpret results.
- Now let's return to the reason we are here today: the Demographic and Housing Characteristics File.

- The DHC is in the queue as the next 2020 Census data product that we'll be releasing in May 2023.
- We need your feedback to better calibrate and test our disclosure avoidance settings for these products.
- To that end, we plan to release another round of DHC demonstration data for public feedback in August.
- And over the course of the summer, we'll continue to evaluate and incorporate feedback from data users.
- It was precisely this method of **feedback and recalibration** that led to many improvements in the DAS algorithm to increase accuracy in our redistricting data product.
  You'll hear more about how we've used your feedback later this morning.
- Lastly, I'll leave you with this quick update:
- We believe we are close to achieving our goal of implementing a new disclosure avoidance methodology for the **more detailed 2020 Census data products** in a way that maintains their fitness for use, while protecting them from disclosure.
- And as I mentioned, we're conducting ongoing engagement with partners and stakeholders alike about our research.
- Such engagement helps identify data needs that otherwise might be missed if it were not for their feedback.
- We're so grateful for having data user and stakeholder communities that are willing to invest effort to help build data products that are useful to everyone in our nation.
- Again, thank you for having us today, and thank you for your input.