Mr. Jeff Hardcastle
State Demographer
4600 Kietzke Lane
Building L, Suite 235
Reno, NV 89502

Dear Steering Committee Members:

Thank you for your letter following up on our February 26 responses to your questions about the U.S. Census Bureau's adoption of differential privacy to protect the confidentiality of respondent data for the 2020 Census.

As we stated in our prior letter, the Census Bureau places great value in the partnership and support provided by your networks, and we appreciate your collective commitment to helping the Census Bureau meet its dual mission of producing high quality statistics about the nation, while safeguarding the privacy of our respondents and the confidentiality of their data.

In your letter, you raised six additional questions about the Census Bureau's adoption of differential privacy and the implementation of the Disclosure Avoidance System, and make three recommendations for the Census Bureau's consideration. Enclosed you will find our responses to your questions and recommendations.

Thank you for your continued commitment to a successful 2020 Census.

Sincerely,

John M. Abowd, Ph.D
Associate Director and Chief Scientist
Research and Methodology

Enclosure

FSCPE Question #1 – "We have heard repeatedly that the decennial census has error in it for several reasons. John Abowd has emphasized this and is hiring staff to examine and quantify the sources and their contribution to that error. The underlying principle for differential privacy is that the census is accurate enough that an accurate identification of an individual can be made. Given the current delays in 2020 Operations with the current crisis and issues with address listing, etc. how can moving forward with differential privacy be justified?"

For decades, the Census Bureau has been diligent at assessing and reporting on sources of error in census counts. These sources of error include operational error, coverage error, and measurement error. Sources of error are routinely reported as part of the Census Coverage Measurement Program. In spite of these errors and the additional error introduced to protect privacy through data swapping, our internal evaluations have determined that the 2010 Census data were still accurate enough to enable confirmed re-identifications for 52 million individuals using only a portion of the published data. Our internal assessment was later confirmed by an independent, external group of scientists and data experts convened by the JASON advisory group. The JASON report on the Census Bureau's decision to adopt differential privacy for the 2020 Census states, *"In the view of JASON, Census has convincingly demonstrated the existence of a vulnerability that census respondents can be re-identified through the process of reconstructing microdata from the decennial census tabular data and linking that data to databases containing similar information that can identify the respondent."* The report goes on to state that *"in view of the demonstrated vulnerability, it is clear that the usual approaches to disclosure avoidance such as swapping, top and bottom coding, etc. are inadequate."* With the JASON's findings confirming our own internal assessments, the Census Bureau stands by our decision that the only way to meet our statutory obligations under Title 13 to protect respondent privacy is to modernize our disclosure avoidance methods through the application of differential privacy for the 2020 Census.

FSCPE Question #2 – "Does the proposed Disclosure Avoidance policy for the 2020 Census represent a new interpretation of Title 13? If so, why now?"

The Census Bureau's decision to adopt differential privacy for the 2020 Census does not reflect a change in our interpretation of Title 13. Rather, it reflects growing empirical evidence, confirmed by outside experts and our own internal researchers that the privacy risks associated with publishing large amounts of highly granular tabulations have increased substantially over the last decade.

FSCPE Question #3 – "Similarly, some DP literature talks about it as a response to a potential not actual threat. Has there been research that assesses the threat level and types of risk to the general public that DP is meant to prevent other than examples like Netflix?"

When the Census Bureau published the 2010 Census Data Products, the disclosure avoidance methods employed for their release were sufficient to protect respondent privacy at that moment in time. Within a few years, optimization algorithms had improved sufficiently to significantly increase the risk of re-identification. Recognizing that the Census Bureau cannot

rely on assumptions that the residual risk of re-identification from the application of disclosure avoidance methods will always remain constant, the Census Bureau determined that we needed to use techniques that do not rely on assumptions about what technology a would-be adversary would be able to leverage against us. Differential privacy establishes a future-proof upper bound on the leakage of private information in published data. While this privacy guarantee does represent a worst-case scenario, and thus does not necessarily reflect the actual risk of re-identification at any given moment in time, our experience has demonstrated that there is no way to predict how that actual risk will increase over time. Our conclusions on both the risk and the need for formally private solutions have been validated by external researchers, including in the JASON report referenced above.

FSCPE Question #4 – "Why, at such a late stage, is the DSEPC convinced they can create useable data with minimal input from stakeholders (given the timeframe) when their efforts to date have not provided quality data?"

The Census Bureau is engaging extensively with our stakeholders to ensure that the 2020 Census Data Products will be fit-for-use for the priority uses of census data, consistent with our obligations under Title 13. In addition to our ongoing stakeholder engagement, we are working with expert working groups organized by the Committee on National Statistics and by our National Advisory Committee and the Census Scientific Advisory Committee.

FSCPE Question #5 – "Can we be assured that the published data, including the second group of products to be released, will be internally consistent, for example that household population in the population- based tables is consistent the household population in the housing-based tables?"

All Group I (PL94-171 redistricting file, Demographic Profiles, and Demographic and Housing Characteristics file) person-level tables (P-tables) will be internally and hierarchically consistent, as will all the Group I household-level tables (H-tables). There are currently no plans to ensure consistency between the Group I P- and H- tables. Consistency between the Group I and Group II (detailed race, detailed ethnicity, tribal data, and person-household joins) data products will be established through constraints that the Group II P-tables must be less than or equal to their Group I P1 equivalent, and the Group II H-tables must be less than or equal to their Group I H1 equivalent.

FSCPE Question #6 "We appreciate the recent release of the metrics. However, it is still unclear the path forward for engaging stakeholder in a dialogue. It is still unclear what the DAS implementation plan is and so what our role is, what other groups are part of the outreach effort, and what are the deliverables and due dates. We need this information to fully inform our elected officials and impacted agencies."

The operational schedule under which the DAS team is currently working has the Census Bureau's Data Stewardship Executive Policymaking Committee (DSEP) making final decisions in September 2020 about the overall algorithm design and the final list of which data elements

will be held invariant. DSEP will then set the final privacy-loss budget and its allocation across data products, tables, and geographic levels in March 2021. Stakeholder feedback to inform those decisions will be invaluable to DSEP's decision-making. To that end, we are currently actively engaging with our stakeholders through various channels, including working groups organized by our advisory committees, by the Committee on National Statistics, and by FSCPE, as well as ongoing engagement with American Indian and Alaska Native tribal leaders and data users, among others. In particular, we are asking these groups to provide feedback through this summer about how we are assessing accuracy and identifying the priority use cases of census data to help inform the September 2020 DSEP decisions. Similarly, feedback that we receive from these groups on minimum acceptable thresholds for accuracy to support various priority use cases, and suggestions for communication materials and supporting guidance on fitness-for-use will help inform the March 2021 DSEP decisions.

FSCPE Recommendation #1 – Provide a clear, consistent and timely communications plan for keeping the full range of stakeholders informed about this. In the letter, Dr. Abowd provides a general overview of outreach efforts. However, we continue to find that the squeakiest wheels are getting the Bureau's attention. This appears to be the case for privacy advocates as well as those of us concerned about data accuracy.

Improvements to the DAS are occurring on a continuing basis, and we recently have expanded our traditional communications channels and enhanced our ability to get timely information out to the diverse data user community by establishing an [email newsletter](#). Data users can subscribe to get prompt notifications regarding important DAS developments. We also post all of our updates on our [DAS Updates webpage](#). Data users that want to provide feedback can do so by emailing our DAS and Data Products teams at 2020DAS@census.gov.

FSCPE Recommendation #2 – As soon as possible, we need information on how the Bureau will provide information on the noise infused products. We need information on whether not DP can be treated as a random error and the ranges of that error. We need to be able to inform our state and local administrative and legislative bodies on changes that will materially impact their operations.

We thank you for this recommendation. The Census Bureau is committed to providing fitness-for-use guidance for our 2020 data products. What form that guidance will take is still under consideration. Similarly, any guidance on error ranges or similar metrics will necessarily need to wait until finalization of the DAS algorithm design and its parameters (i.e., the final privacy-loss budget and its allocation). In the interim, we welcome any feedback that our data user community would like to provide regarding what form this guidance should take and what it should contain. Please send this feedback to Michael.B.Hawes@census.gov and 2020DAS@census.gov.

FSCPE Recommendation #3 – Most importantly, we believe that any further iterations of demonstration data run through a revised DAS are necessary for review. We need to be able to rerun our earlier evaluations against any revised data. Researchers from the CNSTAT

meeting also reached this conclusion. The Demonstration data product has proven to be both useful and necessary to ensuring both the quality and utility of the final release product. As the implementation of the DAS system will impact the data for everyone in the country the need for through, rigorous, independent review of the data is obvious. This review must be a full review of the data not simply based on metrics.

The accuracy metrics that we have developed are intended to allow our data users to assess our ongoing improvements to the DAS algorithm and their impact on fitness-for-use in a variety of ways. That said, we recognize that for some important uses of census data there is no substitute for actually examining the underlying data. You recommend that the Census Bureau release additional demonstration data products to support in depth analysis of the data's fitness-for-use. Unfortunately, the tabulation, documentation, and quality control processes that the Census Bureau employs for public releases of data products are enormously time and labor intensive. With the 2020 Census now underway, we are unable to support additional releases at the present time. That said, in order to support these detailed assessments without overburdening our tabulation and data products teams, the Census Bureau is committing to release the differentially private, but untabulated Privacy-Protected Microdata File (PPMF) produced by each successive iteration of the DAS algorithm for which we release new Detailed Summary Metrics. While these PPMFs will not be in the standard table structures associated with the PL94-171 or DHC data products, it would be an easy matter for some of our public data users to tabulate them accordingly. The PPMFs provided will be exactly the same data used to prepare the Detailed Summary Metrics. They are also exactly the same data the Census Bureau would have tabulated into new demonstration data products. We trust that this solution will meet your needs.