

**From:** Wendy Underhill  
**Sent:** Tuesday, December 22, 2020 10:37 AM  
**To:** James Whitehorne (CENSUS/ADDC FED) <[James.Whitehorne@census.gov](mailto:James.Whitehorne@census.gov)>  
**Cc:** Susan Frederick <[susan.frederick@ncsl.org](mailto:susan.frederick@ncsl.org)>  
**Subject:** NCSL's continuing concerns

Hello James.

NCSL wrote to the bureau on [May 26, 2020](#) expressing our concern about both differential privacy and then-likely delays in the release of data. Uncertainty around delays is still a concern, but I'm writing today on behalf of NCSL to bring forward again NCSL's concern about the use of differential privacy and the impact it may have on data used for redistricting.

We realize that the bureau is balancing many priorities, not the least of which is the protection of respondent data. Another priority is an accurate count throughout the nation.

With the decision to use differential privacy as the bureau's method of disclosure avoidance, and the overlay of post-enumeration processing to mitigate the inherent issues that differential privacy brings, uncertainty about the quality of census data is much higher this decade compared to a decade ago.

Recognizing the significant policy implications for data users relating to the adoption of differential privacy, the bureau produced data demonstration sets at the end of 2019, again in May 2020 and more recently in November 2021. The bureau has worked to overcome some of the anomalies that showed up in the 2019 release, such as census blocks that had no housing units but still had population assigned to them and a "shift" of population from urban areas to smaller areas.

The November 2021 demonstration dataset continues to show total population by racial categories reported at census block and census tract levels that are 10, 20, and even 50 percent different than the 2010 released data. These swings may balance out, but that is cold comfort for anyone who is using census blocks to create equal population electoral districts based on census blocks. This has an impact whether they be congressional, legislative or local districts, although the smaller the unit, the more significant any variation from the as-enumerated number will be. For instance, in Maine, each state House district has less than 9,000 people, and many municipalities in every state have fewer than 500 residents.

Similar differences are still present for smaller racial groups as well. In the demonstration data set, larger race or ethnicity groups have reported totals that are close to the 2010 reported data, but populations of small racial groups show significant distortion. Redistricters do not need dozens of racial categories to do their work, so combining (or binning or collapsing) multi-race groupings would work for these purposes. We recognize that other end users have different needs; the bureau must balance these needs.

The use of differential privacy on prison populations—while holding the number of prisons invariant—will cause reporting that will likely be different from the reported population available directly from the prison administrators.

While we appreciate the bureau's efforts to remedy the issues with differential privacy, data deficiencies introduced by differential privacy remain. We are hearing, as you are I'm sure, that litigation over the quality of the data is likely.

The expectation among at least some states and stakeholders has been that the privacy loss budget used for the demonstration datasets might represent the privacy loss budget the bureau will use . If that is *not* correct, then the user community would benefit (and the outcome of the 2020 decennial census would benefit) from an additional demonstration dataset, based on a more realistic privacy loss budget, or perhaps a series of increasing privacy loss budgets. Users will then have a better understanding of the impact they can expect.

At a minimum, the user community would benefit from better and more frequent bureau communications about the purpose of the demonstration data sets. NCSL has learned that these were not created with a realistic privacy loss budget, but with one set to allow users to see where issues would arise—but not the magnitude of those issues. This may result in an incorrect user assumption that the demonstration data sets represent what will be shared in the P.L. 94-171 file.

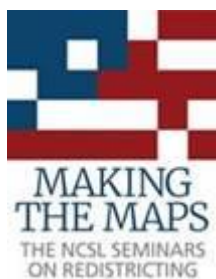
In addition, for months the bureau has told us that a decision about when the privacy loss budget will be determined is coming soon. To date, the bureau has not informed us of when to expect that information or how to provide input on this decision. We realize there are many uncertainties for the bureau, but please know that there are potentially adverse consequences to the states. We respectfully request that the bureau be more responsive to state concerns about the timeline for decisionmaking and for data releases.

As deadlines draw near, time is of the essence to address the remaining important issues states encounter with differential privacy. Please feel free to forward this email to Director Dillingham or others as you deem appropriate. If further explanation of the issues is needed, please ask.

I will close by thanking you for your work throughout 2020 to keep legislators and everyone in the redistricting community apprised of census news and decisions. Please continue to share the bureau's evolving plans with both [Susan.Frederick@ncsl.org](mailto:Susan.Frederick@ncsl.org) and me ([Wendy.Underhill@ncsl.org](mailto:Wendy.Underhill@ncsl.org)).

Best wishes,

Wendy Underhill  
Director – Elections & Redistricting Program  
National Conference of State Legislatures  
303-856-1379 (o) | 303-802-6673 (c)



January 6-8, 2021 online seminar! [Register now.](#)

**[Redistricting is right around the corner!](#)**

NCSL's redistricting seminars will prepare you for this once-a-decade task.