

THE CHALLENGE OF DIFFERENTIAL PRIVACY CONFIDENTIALITY VS. USABILITY

Background

Under Title 13, Section 9, the Census Bureau is prohibited from publishing any information that would reveal personally identifiable information about respondents. With the rise of powerful computer algorithms that can reconstruct individual-level records from tabulated data and the proliferation of third-party data sources that can be linked to those records, the risk of compromising respondents' privacy, especially when publishing highly specific statistics, has increased significantly. As a result, the [Census Bureau introduced a new method of disclosure avoidance](#) called "differential privacy" to ensure compliance with its Title 13 obligations. The Bureau did not issue any Federal Register notice to request public comments on this change. John Abowd, the Bureau's chief scientist, stated to the Federal-State Cooperative for Population Estimates (FSCPE) Steering Committee that there are no other acceptable methods of privacy protection at this time.

What is Differential Privacy?

Differential privacy is a method of "injecting noise" into raw data to prevent the ability to reconstruct individual-level records. The Bureau's plan is to give each state its total population "as enumerated" (i.e., without noise), but at all other levels of geography—from congressional districts down to census blocks—to "inject noise," which will result in changes to the raw data. The smaller the geographic area, the "noisier" the data will be, which will inhibit the accuracy of many sorts of analyses at the smaller geographic levels.

For small geographic areas, the disclosure avoidance system reduces concentrations of people with similar characteristics and moves them elsewhere, lowering the highs and raising the lows. This effect is by characteristic, not persons per se. The system puts everyone and their characteristics into a data blender, mixing up those characteristics and reassembling the characteristics into synthetic people for noise-infused distribution. The result is terrific for ensuring protection of individual identities, but it will have a significant impact for many data analyses used in policy- and decision-making processes based on those characteristics as the resulting data are only partially reflective of reality. It will also impact the ability to enforce both the federal and state voting rights acts.

Differential Privacy Leads to Notable Distortions and Errors

The Washington state Office of Financial Management (OFM) reviewed two iterations of 2010 demonstration data produced by the Census Bureau, and it found notable distortions and errors that would impede effective use of the data for population estimates and redistricting, as well as decision-making and analyses conducted by other data users like businesses, policymakers, and academics – for the next 10 years. A few examples of these anomalies are outlined below:

- The Disclosure Avoidance System (DAS) displaced nearly 18% of Washington's population at the census block level making the data unreliable for small areas. The Census Bureau contends that, when the data are aggregated to a reasonable size, the noise in the small cells will cancel out. It is hard to understand how that is possible when one looks at population characteristics of individual census blocks. One extreme example is the census block containing the Washington Correction Center for

Women (Purdy). In the original 2010 data, the population was 99% female; in the demonstration product after DAS is applied, it is only 12% female.

- Census blocks with a small number of housing units have higher than expected populations in the demonstration data, while blocks with more than 20 housing units appear to have lower than expected populations. In terms of household population, census blocks with only one housing unit had collectively 64,195 more people after applying the DAS. There were also 15,253 people in census blocks that had housing but no population in the original 2010 data. Together, these numbers represent 79,448 people, which is nearly the equivalent of OFM's estimated 2020 population estimate for Chelan County (79,660). These differences will adversely impact all kinds of analyses that rely on census data from political redistricting to planning for urban growth.
- In Washington, there are 401 census blocks where **all** of the population is over 85 years of age and 3,353 census blocks where **all** of the population is 14 years old or younger (i.e., 16,775 youth apparently lived in households without adults present).

The Census Bureau's Disclosure Avoidance System (DAS) Still under Construction

The Census Bureau has already decided to use Differential Privacy in their DAS on 2020 data. Its Data Stewardship Executive Policy (DSEP) Committee will approve final design specifications for the algorithm and the final list of data elements that will be reported "as enumerated" in September 2020. Key decisions about how much noise to inject into the data and which non-statutory data items/tables will ultimately be released need to be made as well. The timing of the decisions about noise and tables is unclear because the Bureau has never released a DAS implementation plan. Decisions about the timing of data delivery – and even the end of the enumeration process – are still up in the air, but they will have to be made before the release of redistricting data.

The Bureau continues to accept feedback on the DAS system and the demonstration products it has produced. It plans to release a third set of 2010 demonstration data and summary metrics for comment and review in September. Still, it is very concerning that the system is NOT yet ready for use. It is equally concerning that key decisions about which data elements will be reported "as enumerated" and how much noise will be infused into the data still need to be made. OFM is convinced that the Census Bureau needs to give itself more time to develop its Disclosure Avoidance System before it is deployed or it risks seriously undermining the quality of the data and its usability across all sectors of society.

What Has Been Done to Address the Problem?

OFM has expressed the state's concerns through its participation in the Census Bureau's Federal-State Cooperative for Population Estimates (FSCPE) and State Data Centers (SDC) programs, as well as through its participation in the National Academy of Science's Committee on Statistics (CNSTAT) and its metrics workgroup. Based on its review of the Bureau's demonstration products, OFM submitted two analyses pointing out the weaknesses in the system (see [first analysis](#) and the [second analysis](#) with [supporting charts](#)). Other states, including Alaska, Colorado, Maine, New York, Utah, and Virginia, have also tested the data and pointed out the flaws of the system. The Bureau also continues to dialogue on differential privacy with American Indian and Alaska Natives through its Tribal Listening Sessions. Members of the Bureau's National Advisory Committee Working Group and Census Scientific Advisory Committee Working Group are another group of advisors that may also weigh in. Finally, the [National Council on State Legislatures](#) is watching the issue closely and wrote to the [House Committee on Oversight and Reform](#) and the [Senate Committee on Homeland Security and Governmental Affairs](#) in May 2020.

For more information, contact: Mike Mohrman, State Demographer, OFM mike.mohrman@ofm.wa.gov

Updated: September 15, 2020