

2010 Demonstration Privacy-Protected Microdata File 2020-11-16

Over the past several months, the Census Bureau has been making a number of improvements to the 2020 Census Disclosure Avoidance System (DAS) to address the concerns raised by the data user community at the December 2019 Committee on National Statistics workshop. Throughout this process, we have received numerous requests for additional tools to help evaluate this ongoing progress.

We are pleased to announce that in response to this feedback, we have devised a solution to produce updated demonstration data sets. This would not be possible without the support of the IPUMS National Historic Geographic Information System (NHGIS), members of the Committee on National Statistics' (CNSTAT) expert group.

The detailed summary metrics we released for these DAS data runs, and that we will continue to release as we make future improvements to the algorithm, allow our data users to assess these improvements and their impact on fitness-for-use in a variety of ways. That said, we recognize that our data users assess accuracy and fitness-for-use for diverse use cases in very different ways.

Privacy-Protected Microdata Files

To assist with these assessments, we are now releasing new [“Privacy-Protected Microdata Files”](#) (PPMFs), which are the underlying microdata files for the entire nation used to generate the Detailed Summary Metrics. It is important to note that while the data in the PPMFs look like individual records, all of the data are privacy-protected. The microdata records generated by the DAS ensure respondent privacy through the application of differentially private statistical noise. The microdata included in the PPMFs do not include any actual census responses. They are simply the microdata format used by the Census Bureau's production system to produce privacy-protected tables that the DAS generated.

While these PPMFs are untabulated microdata records, NHGIS is tabulating, formatting and posting data tables from the PPMFs for our data users to evaluate.

Algorithm Improvements Reflected in This Release

The version of the DAS used to generate the PPMFs includes a number of system improvements developed over the last few months to reduce distortions caused by post-processing and to improve the accuracy of population counts for legal and political entities. One of the most notable system improvements reflected in the new PPMFs is a change in how the DAS processes data for American Indian and Alaska Native (AIAN) tribal areas. Previously, the DAS



census.gov
2020census.gov
@uscensusbureau

performed all of the [post-processing](#) along the standard geographic hierarchy. This prior approach led to notable problems for key “off-spine” geographies like the AIAN tribal areas. The new PPMFs reflect system design changes to address this issue. We have implemented a new geographic hierarchy specifically for AIAN tribal areas within each state, and have added a new state-level population [invariant](#) for AIAN tribal areas. Note: this version (2020-11-16) corrects a defect in the system code discovered following the original release. We have updated the PPMFs and Detailed Summary Metrics using the corrected code base.

What data are included in the PPMFs?

The Census Bureau’s recent operational schedule changes have necessitated that we focus all of our current attentions on preparing for the production of the PL94-171 redistricting data files by the statutory deadline of March 31, 2021. Consequently, these new PPMFs are limited to the data necessary to support tabulation of tables P1-P5 and H1 of the redistricting data. Data elements necessary to support other data product tabulations are not included in this release. The full set of data elements necessary to support the Demographic Profiles and Demographic and Housing Characteristics files will be reincorporated into DAS runs at a future date.

Privacy-Loss Budget

The DAS run that generated the PPMFs v. 2020-11-16 use a comparable global privacy-loss budget as was used for the 2010 Demonstration Data Products that were released in October 2019. The person-level tables received a privacy-loss budget of $\epsilon=4$, and the housing unit-level tables received a privacy-loss budget of $\epsilon=0.5$.

For this DAS run, the privacy-loss budget (PLB) was allocated across geographic levels and queries as follows:

Allocation by Geographic Level (for both PPMF-P and PPMF-U):

Geographic Level	PLB Allocation
Nation	20%
State	16%
County	16%
Tract	16%
Block group	16%
Block	16%

Allocation by Query for PPMF-P:

Query	PLB Allocation
total	30.0%
numraces	10.0%
cenrace	10.0%
hispanic*numraces	10.0%
hispanic*cenrace	2.5%
votingage*numraces	10.0%
votingage*cenrace	2.5%
votingage*numraces*hispanic	2.5%
votingage*cenrace*hispanic	2.5%
instlevels	10.0%
gqlevels	7.5%
detailed	2.5%

All of the PLB for the PPMF-U was assigned to the H1 table query.

For More Information, See: [Developing the DAS: Demonstration Data and Progress Metrics \(census.gov\)](#)