2010 Demonstration Privacy-Protected Microdata File (2020-05-27): Frequently Asked Questions

What are Privacy-Protected Microdata Files?

The new "Privacy-Protected Microdata Files" (PPMF) supplement the Detailed Summary Metrics we've created to help data users evaluate ongoing improvements to the Disclosure Avoidance System (DAS) algorithm. Since we released the 2010 Demonstration Data Products in October 2019, we've received numerous requests to produce a comparable data product that would give data users greater flexibility to assess ongoing improvements. With the support of the Committee on National Statistics (CNSTAT), we are able to do so.

The PPMF is released as untabulated microdata records. However, members of the CNSTAT expert group are tabulating, formatting and posting these records into data tables that you might find more user-friendly. Those will be available after upcoming design sprints, soon after we post each PPMF. For those who choose to work with the untabulated PPMF, the Table Layout File is designed to guide you in that work. Our partnership with CNSTAT allows the Census Bureau staff who would otherwise perform the time-intensive tabulation, data review and release process in-house to continue their focus on other important data collection and processing work.

It is important to note that while the data in the PPMF looks like individual records, all of the data is privacy-protected. The microdata records generated by the DAS ensure respondent privacy by applying differentially private statistical noise. The microdata included in the PPMF do not include any actual census responses.

What privacy-loss budget is used for the PPMF releases, and why?

The PPMF uses the same privacy-loss budget (PLB, also known as the epsilon value) that was used for the 2010 Demonstration Data Products released in October 2019. By keeping the privacy-loss budget constant, we can isolate impacts attributable entirely to the DAS algorithm and postprocessing development and compare "apples to apples." The person-level tables received a privacy-loss budget of ε=4, and the housing unit-level tables received a privacy-loss budget of ε=2.

What DAS improvements are reflected in the 2020-05-27 PPMF and Detailed Summary Metrics?

Both the 2020-05-27 PPMF and corresponding Detailed Summary Metrics stem from the same DAS microdata run from the March 2020 algorithm design sprint. This sprint focused on "postprocessing"—changing how the DAS TopDown Algorithm (TDA) converts the noisy tabulations taken from the confidential data into non-negative integer counts.

What types of microdata are included in the PPMF?

The first PPMF release (2020-05-27) only includes person-level microdata, as it reflects the postprocessing focus of the March 2020 sprint, a topic that is relevant to population counts. Specifically, this release includes the person-level microdata supporting the P.L. 94-171 Redistricting Data Summary File data and the population tables in the proposed Demographic and Housing Characteristics file.

How often will new PPMFs be released throughout DAS development?

We are using the "Agile" development process to engineer DAS updates. "Agile" is an approach used in software development that allows teams to plan, design and test iterative solutions in a series of short intervals, or "sprints." Each DAS development sprint cycle is approximately six weeks in length and focuses on two to three executive priorities. Results and discoveries from each sprint guide the priorities of the following sprint. With the ongoing support of CNSTAT, we intend to release a new set of Detailed Summary Metrics and PPMFs after relevant DAS development sprints through the remainder of 2020.

For More Information, See: Developing the DAS: Progress Metrics and Data Runs Web Page