



UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

2020 CENSUS PROGRAM MEMORANDUM SERIES: 2019.25

Date: October 31, 2019

MEMORANDUM FOR: The Record

From: Albert E. Fontenot, Jr. **(signed October 31, 2019)**
Associate Director for Decennial Census Programs

Subject: 2010 Demonstration Data Products - Design Parameters and Global Privacy-Loss Budget

Contact: Cynthia Hollingsworth
Decennial Census Management Division
301-763-3655
cynthia.davis.hollingsworth@census.gov

Robert Sienkiewicz
Center for Enterprise Dissemination
301-763-1234
Robert.sienkiewicz@census.gov

This memorandum documents requirements for the design of the 2010 Demonstration Data Products that the 2020 Census Program has received from the Data Stewardship Executive Policy Committee (DSEP) regarding how to demonstrate to stakeholders our planned method for protecting the information we collect from the American public during the 2020 Census. DSEP has instructed us to apply differentially private disclosure avoidance methods to all of the statistics we produce from these data, starting with data collected during the 2018 End-to-End (2018 E2E) Census Test. DSEP will manage the amount of privacy loss associated with the 2020 Census data products, called the “privacy-loss budget,” and make policy decisions regarding the design of the 2020 Disclosure Avoidance System (DAS). I will communicate these policy decisions and their associated requirements as we receive them from DSEP through the 2020 Census Memorandum Series.

The purpose of this public release is to supply external stakeholders with a large number of tables from the 2010 Census using the confidentiality protections proposed for the 2020 Census, thereby allowing users to compare the data and better understand our new disclosure avoidance methods. It is imperative that the Census Bureau show users that the 2020 DAS will produce high quality data, and to give those users an accurate demonstration of the system’s capabilities so that they can determine whether the products will meet their needs.

At its August 1, 2019 meeting, the Data Stewardship Executive Policy Committee (DSEP) approved the release of a set of data products that will demonstrate the quality of the 2020 DAS by running the algorithms on the 2010 Census Edited File (CEF). The 2010 Demonstration Data Products include the demonstration version of the P.L. 94-171 and selected tables from the proposed 2020 Demographic and Housing Characteristics Summary File for all states, Puerto Rico, and the District of Columbia.

On September 12, 2019, DSEP was asked to make a final decision on the design parameters for the version of the DAS that would be used to create these 2010 demonstration data products. When making these decisions, DSEP reviewed many supporting materials and illustrations of the accuracy privacy-loss tradeoffs of several values of the privacy-loss budget (the technical parameter “epsilon” or “ ϵ ”), as well as analysis provided by the 2020 DAS team and the Demographic Programs Directorate. Possible values of the privacy-loss budget represent privacy/accuracy trade-offs along the spectrum between perfect privacy/low accuracy ($\epsilon = 0$), to perfect accuracy/low privacy ($\epsilon = \infty$). All of these decisions apply exclusively to the DAS used to produce the 2010 demonstration data products and do not extend to the version of the DAS that will be used for the 2020 Census itself. DSEP will decide on the DAS design parameters for the 2020 Census publications separately and decide on a value of epsilon at a later date.

DSEP made its decisions with regard to the design of the 2010 demonstration data products under the following assumptions:

1. Consistent with our legal obligation under Title 13 U.S.C. to protect respondent confidentiality, it is ultimately the Census Bureau’s goal to produce data using the lowest value of epsilon (smallest privacy-loss budget) that will support the legitimate needs of our users.
2. The public cannot make a “true” comparison between the accuracy of these demonstration data products and the data that we originally produced from the 2010 Census. This is because in addition to the standard measures of error, all of the 2010 Census publications had statistical disclosure limitation (SDL) applied to them. Since these data were produced under legacy SDL methods, we must keep the degree to which that SDL modified the data confidential to protect the confidentiality of the underlying microdata.
3. The value of epsilon used to create these demonstration data products will not generate data with the same level of accuracy if applied to the 2020 Census publications. This is because:
 - a. Between now and the release of data from the 2020 Census, we will continue to improve the performance of our DAS algorithms and the science of differential privacy.
 - b. This privacy-loss budget will only be used to produce a limited set of data products. Many additional publications will draw on the global privacy-loss budget for the 2020 Census.
 - c. DSEP will need to make decisions on how to distribute the privacy-loss budget for the 2020 Census publications based on evolving requirements.

With those assumptions in mind, DSEP approved the following design parameters for the DAS used to produce the 2010 demonstration data products. To reiterate, these figures will be produced from data collected during the 2010 Census:

- The total population will be reported as enumerated (invariant) at the state level.
- Disclosure avoidance will be applied to the voting age population at all levels of geography.
- The count of total housing units (not population) will be reported as enumerated (invariant) at the block level.

- Disclosure avoidance will be applied to the count of occupied housing units at all levels of geography.
- The count of group quarters (not population) by the seven category types used in P.L. 94-171 Table P5 will be used as enumerated (invariant) at the block level.
- The global privacy-loss budget for these publications will be an epsilon of six ($\epsilon = 6.0$). Of that six:
 - Four ($\epsilon = 4.0$) will be allocated to the microdata detail file that supports population tables produced in the 2010 Demonstration Data Product.
 - Two ($\epsilon = 2.0$) will be allocated to the microdata detail file that supports the housing and household tables in the 2010 Demonstration Data Product.

The 2020 Census Memorandum Series

The 2020 Census Memorandum Series documents significant decisions, actions, and accomplishments of the 2020 Census Program for the purpose of informing stakeholders, coordinating interdivisional efforts, and documenting important historical changes.

A memorandum generally will be added to this series for any decision or documentation that meets the following criteria:

1. A major program-level decision that will affect the overall design or have significant effect on the 2020 Census operations or systems.
2. A major policy decision or change that will affect the overall design or significantly impact 2020 Census operations or systems.
3. A report that documents the research and testing for 2020 Census operations or systems.

Visit 2020 Census on [Census.gov](https://www.census.gov) to access the Memorandum Series, the 2020 Census Operational Plan, and other information about preparations for the 2020 Census.