UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION

THE STATE OF ALABAMA, *et al.* Plaintiffs, v. UNITED STATES DEPARTMENT OF COMMERCE, *et al.*, Defendants.	No. 3:21–cv–00211–RAH–ECM–KCN

NOTICE OF SUPPLEMENTAL EVIDENCE

Pursuant to the Court's April 19, 2021 Order, Doc. 93, Defendants submit supplemental evidence in further support of their Response in Opposition to Plaintiffs' Motion for Preliminary Injunction and Petition for Writ of Mandamus, Doc. 41.

Defendants submit: Exhibit 1, the Supplemental Declaration of John M. Abowd; and Exhibit 2, the April 19, 2021, Disclosure Avoidance System Update reflecting that the "2020 Census P.L. 94-171 data" will be released as a "Legacy Format Summary File" by August 16, 2021, also available at https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html

DATED: April 26, 2021

Respectfully submitted,

BRIAN M. BOYNTON
Acting Assistant Attorney General

ALEXANDER K. HAAS
Director, Federal Programs Branch

BRAD P. ROSENBERG
Assistant Director, Federal Programs Branch

/s/ *Elliott M. Davis*
ZACHARY A. AVALLONE
ELLIOTT M. DAVIS (N.Y. Reg. No. 4596755)
JOHN ROBINSON
Trial Attorneys
Civil Division, Federal Programs Branch
U.S. Department of Justice
1100 L St. NW
Washington, DC 20005
Phone: (202) 514-5336
E-mail: elliott.m.davis@usdoj.gov

Counsel for Defendants

CERTIFICATE OF SERVICE

I hereby certify that on April 26, 2021, I filed with the Court and served on opposing counsel through the CM/ECF system the foregoing document.

DATED: April 26, 2021

/s/ Elliott M. Davis
ELLIOTT M. DAVIS
(N.Y. Reg. No. 4596755)
Trial Attorney
Civil Division, Federal Programs Branch
U.S. Department of Justice
1100 L St. NW
Phone: (202) 514-4336
Fax: (202) 616-8470
E-mail: elliott.m.davis@usdoj.gov

Counsel for Defendants

EXHIBIT 1

IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION

STATE OF ALABAMA, *et al.*,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF COMMERCE, *et al.*,

Defendants.

Case No. 3:21-CV-211-RAH-ECM-KCN

SUPPLEMENTAL DECLARATION OF JOHN M. ABOWD

I, John M. Abowd, make this supplemental Declaration pursuant to 28 U.S.C. § 1746, and declare that under penalty of perjury the following is true and correct to the best of my knowledge. I am submitting this Declaration to supplement the Declaration I submitted in this case on April 13, 2021. In this supplemental Declaration, I clarify and respond to allegations and claims made by plaintiffs and their declarants.

REVERTING TO SWAPPING WILL FURTHER DELAY THE RELEASE OF REDISTRICTING DATA.

1. In my prior Declaration, I stated that there would be substantial additional delays to the release of the redistricting data were the Court to require the Census Bureau to revert to using swapping for the 2020 Census (Abowd Decl., ¶¶ 84–86). Plaintiffs counter, "there's good reason to think the 2010 methods could be applied more quickly than still-in-development differential privacy." (Reply, p. 21). This is categorically false. Given the scale, complexity and critical importance of the decennial census, the Census Bureau has developed and consistently applied rigorous standard operating procedures to ensure the integrity and reliability of census data processing. Reverting to swapping for the 2020 Census would require numerous analysis, policy, auditing, deployment, system testing and quality assurance steps before swapping could be used in the 2020 Census production workflow.
2. Some steps that would be required include that the Census Bureau's Disclosure Review Board (DRB) would need to assess the disclosure risks and develop the proposed swapping algorithm, parameters and swap rates to be used. Next, the DRB would need to assess and document the residual risk of disclosure for this methodology. The Data Stewardship Executive Policy (DSEP) Committee would then need to review and approve the DRB's proposal and residual risk assessment. Once approved, technical staff would need to program the swapping algorithms for use in the 2020 Census computing environment and their work would need to be audited to ensure that the software accurately implements the selected swapping rules, parameters and rates. Once

audited, the swapping software could be deployed to the 2020 Census computing environment, where it would have to undergo extensive, mandatory system integration testing before it could be used in production. Even if all these steps were expedited, this whole process could take 24-28 weeks[1] and would inevitably delay further the release of the redistricting data product.[2]

3. The 2020 Disclosure Avoidance System using the TopDown Algorithm (TDA) is fully operational and has already completed all necessary auditing and system integration testing currently required for 2020 Census Information Technology systems. All that remains is the final Operational Readiness Review on May 20, 2021 and the final setting and allocation of the privacy-loss budget by DSEP in June, incorporating data user feedback from the April 2021 demonstration data. Under any scenario, using the 2020 DAS will enable the Census Bureau to release the redistricting data sooner than would be possible if we were required to revert to swapping.

---

[1] The estimate is based on the time required to completely refactor the code base for swapping, port the refactored code to the production environment for the 2020 Census, repeat the Test Readiness, Production Readiness and Operational Readiness Reviews with the same protocols used for the current DAS, then resume the production sequence to produce a clean, certified Microdata Detail File.

[2] Just like in prior years, the disclosure avoidance needs to be applied prior to the release of the redistricting data, or any data product other than apportionment. The federal government and the broader statistical disclosure limitation field have long acknowledged the necessity of considering all releases of related data when making decisions regarding disclosure risk. Office of Federal Statistical Policy and Standards (1978) Statistical Policy Working Paper #2 "Report on Statistical Disclosure and Disclosure Avoidance Techniques" p. 14, available at https://nces.ed.gov/FCSM/pdf/spwp2.pdf#:~:text=Policy%20and%20Standards%20Statisti-cal%20Policy%20Working%20Paper%202,Economist%20Office%20of%20Federal%20Statistical%20Policy%20and%20Standards. See also: Cox (1976) and Fellegi (1972).

THE CENSUS BUREAU'S TOPDOWN ALGORITHM CAN BE TUNED TO MAKE POPULATION COUNTS EFFECTIVELY INVARIANT

4. The privacy accounting framework of differential privacy and the hundreds of finely tunable parameters of the Census Bureau's 2020 TDA are extremely nimble and precise. For example, the Census Bureau could meet numerical accuracy targets for block-level population counts through reallocation of privacy-loss budgets using the tuning parameters. Allocating a sufficiently high privacy-loss budget for population counts at the block level could result in nearly all block population counts being reported exactly as enumerated. Note, however, that the Census Bureau has already leveraged the flexibility and precision of the TDA to meet the accuracy targets we established for the redistricting and Voting Rights Act use cases. As I explained in my prior Declaration, keeping the block-level population variant is important because even the slightest uncertainty in each block-level population provides exponential protection to the data set as a whole. We could reallocate privacy-loss budget to meet accuracy requirements within the current algorithm and schedule, but that would be at the expense of accuracy for other characteristics.[3]

THE POTENTIAL CONSEQUENCES OF HOLDING BLOCK-LEVEL POPULATION INVARIANT.

5. Holding block-level populations invariant would present a number of challenges and would be difficult to implement within our existing Disclosure Avoidance System using TDA.[4] First, implementing this invariant would risk further delaying the Redistricting Data. A new invariant would also put at risk the fitness-for-use of the

[3] Reallocating privacy-loss budget for block-level population accuracy implies that other data such as voting age, race, ethnicity, sex, and age will be less accurate.

[4] The ability of DAS to find a feasible Microdata Detail File in the presence of an additional invariant, e.g., block-level population totals, depends upon proving that a technical condition in mixed integer linear programming remains true in the presence of the new invariant. That condition has not been verified for the production version of the DAS

remaining 2020 Census data because it would remove *all* confidentiality protection on a key identifier used in re-identification attacks—the census block. If the Census Bureau removes that uncertainty by forcing a block-level population invariant, stronger disclosure avoidance (more noise) would have to be used for other variables like sex, age, race, and ethnicity.

6. Requiring the Census Bureau to hold block-level population invariant could further delay the release of the Redistricting Data. The 2020 Disclosure Avoidance System using TDA can hold certain tabulations invariant. Current invariants are already programmed into the algorithm including total population counts at the state level, the number of housing units at the block level, and the number and major type of group quarters facilities at the block level.
7. The 2020 DAS TopDown Algorithm performs a series of complex optimizations at each geographic level, from the nation down to individual census blocks. Imposing constraints on these optimizations through invariants limits the number of possible solutions to these optimization steps and runs the risk of the algorithm either taking longer than expected to complete the optimization or crashing entirely.[5]
8. The Census Bureau has extensively tested the stability and performance of the 2020 DAS with the current list of invariants. Unless similar testing and analysis is done on the potential impact that including this additional invariant might have, we cannot guarantee that the DAS will be able to complete its production run of the 2020 Census

---

planned for the P.L. 94-171 Redistricting Data Summary File in the presence of a block total population invariant.

[5] The operational vulnerability caused by invariants is not unique to our implementation, nor to differential privacy as a whole; swapping algorithms like those used in 2010 could face similar "unsolvable" situations if those algorithms are unable to find households in the target geographies that match on the key swapping characteristics (i.e., the invariants).

redistricting data in the period of time currently allotted for disclosure avoidance processing in the production schedule.

9. To add an invariant for total population counts at the block level to the algorithm, we would need to modify the algorithm—in other words, we need to complete a mathematical analysis of the full equation system that produces the Microdata Detail File to ensure that a mathematical solution to the system exists and can be found in finite time using the state-of-the-art commercial optimization software embedded in the DAS. The production code base is scheduled for finalization in its Operational Readiness Review (ORR) on May 20, 2021.[6]

10. There are many data processing steps that need to occur between the release of the apportionment data and the release of the redistricting data product. The application of disclosure avoidance is one small, and relatively short, step in that process. Under our current production schedule, targeting release of the redistricting data in the legacy file format by August 16, 2021, each sequential data processing stage has a tightly constrained duration and there is no margin for schedule slippage. If any processing step takes longer than it is allocated, the release date would likely be delayed. The introduction of a new, untested invariant poses just such a risk.

11. Including a block-level population count invariant would also impact data quality for the remaining 2020 Census data. In my prior Declaration, I explained how including even minimal amounts of uncertainty in block level population counts greatly reduces the ability of an attacker to perform a reconstruction-abetted re-identification attack (Abowd Decl., ¶42). Were the Census Bureau to impose a block-level invariant on population counts, we would necessarily need to apply additional privacy protections

[6] After the Operational Readiness Review, the privacy-loss budget can still be adjusted and reassigned, but the optimization problems the DAS solves are locked.

to tabulations of other characteristics (sex, age, race, and ethnicity) to meet our obligations under 13 U.S. Code §§ 8(b) and 9. These additional protections, either in the form of table suppression or reduced privacy-loss budget would have deleterious consequences for redistricting and Voting Rights Act enforcement, as well as other statutory uses of decennial Census data, including the Population Estimates program and federal funding allocations. If the goal is to achieve down-to-the-person accuracy on block-level populations, then the correct way to accomplish this is to change the allocation of privacy-loss for that variable.

PLAINTIFFS MISCONSTRUE HOW THE TDA PROCESSES TRIBAL AREAS.

12. The 2020 Census DAS TopDown Algorithm operates along a geographic hierarchy; this is what ensures that the accuracy of statistics increases as the underlying population increases. The standard hierarchy starts at the national level, then descends to state, county, tract, block group, and finishes at the individual census blocks. This hierarchy posed some challenges early in the development of the DAS because it was difficult to ensure high levels of accuracy for geographic entities that split these geographic levels (e.g., for incorporated places that contain portions of different census tracts). The Census Bureau addressed this challenge by implementing several changes to the geographic hierarchy used by the TDA to improve accuracy for all "off-spine" geographic entities, like voting districts. One change to the geographic hierarchy that we implemented in September 2020 was to create an alternate geographic processing hierarchy for federally recognized American Indian and Alaska Native (AIAN) tribal areas within each state.
13. Plaintiffs allege that the separation of the AIAN tribal areas within the geographic hierarchy demonstrates that the Census Bureau is "prioritizing the accuracy of the data for certain racial and ethnic groups over others" (Reply, p.39). This is false. The changes to the AIAN geographic hierarchy were implemented to address the distinct

legal and political status of those geographies and to reflect the government-to-government relationship we have with federally recognized tribes. County, tract, block group and block statistics for these AIAN tribal areas receive the same allocation of privacy-loss budget (level of accuracy) as their corresponding geographies in the remainder of their states.[7] In fact, the isolation of the AIAN tribal areas is very similar to the way the TDA post-processing isolates group quarters facilities at the block group level from their surrounding non-group quarters populations. Privacy-loss budget, and its corresponding impact on accuracy, is allocated by data table element. Within each such element, all demographic sub-groups receive the same allocation of privacy-loss budget. Thus, the accuracy improvements derived from privacy-loss budget allocation apply to all demographic groups equally. The TDA does not, and will not, allocate greater privacy-loss budget to any particular demographic group or subgroup over another.

THE 2000 DEPARTMENT OF JUSTICE LETTER.

14. Plaintiffs quote two Census Bureau sources that referenced a 2000 "agreement" between the Census Bureau and the U.S. Department of Justice that supposedly established block-level invariants as a legal requirement (Reply, p.10). I personally investigated the history of this supposed agreement after it was raised by staff within the Census Bureau. After diligent research, we found that the supposed "agreement" was a March 15, 2000 letter from Acting Assistant Attorney General Bill Lee, of the DOJ Civil Rights Division, to Census Bureau Director Kenneth Prewitt, merely stating that the attorneys at the Civil Rights Division did "not believe that the application of [the proposed] disclosure avoidance techniques [for the 2000 Census] will impair the

[7] As I said in my first Declaration, accuracy is measured as the absolute error in the counts, not relative percentage errors.

use of these data for enforcement of civil rights programs." The body of the letter is below and the full letter is attached as an Appendix.

Dear Dr. Prewitt:

This is in regard to the Census 2000 Redistricting Data and confirms the February 25, 2000 telephone conversation between Marshall Turner of your staff and my Deputy, Anita Hodgkiss.

At a November 29, 1999 meeting here at the Civil Rights Division, you and your staff discussed with Ms. Hodgkiss the possible need to apply certain disclosure avoidance techniques to the detailed race and ethnicity statistics included in the Census 2000 Redistricting Data files in order to meet the confidentiality requirements of Title 13, U.S.C. We have reviewed the information you provided at the November 29 meeting and we do not believe that the application of these disclosure avoidance techniques will impair the use of these data for enforcement of civil rights programs.

We greatly appreciate your kind assistance.

Sincerely,

Bill Lann Lee
Acting Assistant
Attorney General
Civil Rights Division

15. Over the subsequent years, misinterpretation and faulty recollection of the content of this letter by Census Bureau staff led to the perpetuation of an erroneous oral history on this subject. As the Census Bureau began developing the requirements for the 2020 DAS, we retrieved and reviewed the original letter. As can be seen from the face of the letter, it contains no agreement or legal analysis requiring *any* invariants, let alone block-level population.

THE CENSUS BUREAU'S CONCERN ABOUT LOCATION PROTECTION IS NOT NEW

16. Plaintiffs assert that the Census Bureau's decision not to hold population counts invariant at the block level, in order to enhance respondents' location protection, reflects

a "new interpretation" by the Census Bureau regarding its obligations to protect confidentiality (Reply, p. 29). To the contrary, the Census Bureau has long recognized the growing disclosure risk of releasing highly accurate data for small geographic units. That was precisely the rationale behind the table suppression methodologies used prior to 1990 and the transition to swapping for the 1990, 2000 and 2010 Censuses.

17. Accurate and precise location information significantly increases the risk of re-identification (making it easier to match individuals to addresses contained in external data) because of the prevalence of persons in the population who have unique values for the combination of census block, age (in years) and sex. The Census Bureau has long employed disclosure avoidance methods to reduce this risk. Geographic aggregation (reporting only at higher levels of geography) was the primary protection afforded by table suppression in 1970 and 1980 and continued to be used through 2010 for the 2010 Census Public-Use Microdata Sample. Swapping, as used for the 1990-2010 Censuses, sought to further counter this growing risk by attempting to protect the individuals considered most vulnerable when reporting highly accurate block-level statistics. Between 1990 and 2010, the swapping methodologies and swap rates used evolved in an attempt to keep pace with the growing risks of releasing highly accurate information at the block level. But, as I referenced in my prior Declaration, the Census Bureau recognized even prior to the publication of 2010 Census data that the risks of publishing highly accurate block-level statistics were continuing to increase, and would need to be further evaluated in the context of 2020 Census planning (Abowd Decl. ¶37, fn33). Our adoption of differential privacy and our removal of the invariant on population counts at the block level are a highly effective mechanism for countering this growing threat of re-identification, while continuing to produce high quality statistics about the nation.

DR. STEVEN RUGGLES IS NOT AN EXPERT IN DIFFERENTIAL PRIVACY

18. I am familiar with the work of Dr. Ruggles. He is one of the world's leading experts on and proponents of the use of household microdata to advance social science with demographic modeling. I have collaborated with him on multiple large-scale projects, including the dissemination of the demonstration data associated with the TopDown Algorithm. In my opinion, however, he does not have significant experience with the capabilities of modern statistical disclosure limitation based on the principles of differential privacy to render an expert opinion about this matter, specifically on the subject of the relation between disclosure limitation methods and the underlying mathematical theory of differential privacy for privacy-loss accounting. While Dr. Ruggles has published some articles on differential privacy, all are merely critiques of the Census Bureau's 2020 Disclosure Avoidance System that contain the same types of errors as the errors in the report submitted in this case. In addition, Dr. Ruggles conflates early iterations of the DAS released for demonstration purposes with its capabilities in production form.

19. One simple error permeates his report: Dr. Ruggles confuses the concepts of privacy-loss accounting using differential privacy with the specific disclosure limitation technique the Census Bureau plans to use, the TopDown Algorithm.

20. Differential privacy is not an algorithm or a disclosure limitation technique—it is an accounting method for evaluating and comparing risks from different disclosure limitation techniques. One way to think of differential privacy is like the accounting methods used by businesses to track expenditures and identify waste of resources. Differential privacy quantifies the privacy loss from making certain data public, and

it quantifies the privacy protection provided by applying different statistical disclosure limitation methods. An organization using differential privacy can compare different disclosure limitation methods and detect and fix vulnerabilities that could lead to significant privacy loss. The specific disclosure avoidance technique that the Census Bureau plans to use is called the TopDown Algorithm—it is not called "differential privacy." Differential privacy is used to measure the disclosure risk after the Census Bureau applies its TopDown Algorithm (TDA).

21. Differential privacy can be used with a variety of statistical disclosure limitation techniques. Traditional, and very old, statistical disclosure limitation techniques such as randomized response (Warner 1965) and noise infusion (Evans et al. 1998) are often used with differential privacy accounting. But modern research has designed more efficient new techniques that improve accuracy for the same amount of privacy loss. For example, Google and Apple have used randomized response (a technique invented in 1965) combined with differential privacy accounting and accuracy improvements to collect mobile device usage information (citations in my original Declaration), while protecting user identities and activities on their phones.

22. Rather than being an "entirely new approach," the TopDown Algorithm is an improvement over traditional methods based on new ideas that result from scientific research. To create TDA, the Census Bureau used differential privacy methods to improve the efficiency of noise infusion compared to its traditional data swapping approach.

23. Disclosure risk assessments, such as those cited by Dr. Ruggles, are known to underestimate true disclosure risk—these assessments can only measure the success of the specific privacy attack that they consider. As a result, if these assessments estimate high disclosure risk, then the true disclosure risk is high, but if these assessments estimate low disclosure risk, then no conclusions can be drawn. These attack-specific methods also ignore a myriad of other feasible privacy attacks. Methods focused on

a specific attack strategy are not capable of measuring disclosure risk more generally, let alone the disclosure risk possible after continued developments in computer hardware as well as improvements in the algorithms used in these privacy attacks. For this reason, such attack-specific methods cannot be used as the sole measures of disclosure risk.

24. Differential privacy-loss accounting is necessary because older disclosure limitation methods are less reliable at estimating disclosure risk. Often they severely underestimate disclosure risk, as shown by a successful reconstruction attack on Aircloak's Diffix system by Cohen and Nissim in (2020, 2018) that used the same linear programming methods as the Census Bureau used to perform its simulated reconstruction-abetted re-identification attack on the 2010 Census. To re-iterate, prior to differential privacy, there was no satisfactory method for tracking privacy loss. Prior methods were based on assumptions about the attacker's information and technology but could severely underestimate the risk if those assumptions were wrong.

25. Dr. Ruggles' statement "[i]t has long been recognized, however, that there is no direct relationship between the level of ε and the risk of disclosing identities" is incorrect and his reference to McClure and Reiter (2012) misconstrues their result. The epsilon parameter in differential privacy, when accurately measured, is directly related to disclosure risk (Wasserman and Zhou, 2010); specifically it limits the statistical power of all possible tests for whether a particular individual's data record (or portions thereof) was used to produce a collection of statistics versus the record of another, arbitrary individual. This is exactly the same identity disclosure definition used by McClure and Reiter.[8] The latter show, contrary to Dr. Ruggles' claim, that some attack models do not succeed even when the differential privacy parameter ε is large. This means

[8] Technically, both methods set up the statistical re-identification hypothesis such that the likelihood ratio, the contribution of the data to the attacker's inference about re-identification, is the same.

the data may be safe from that particular attack, not that re-identification and epsilon are unrelated. Wasserman and Zhou show that *any* identity attack model is limited by ε because it constrains the optimal test statistic for an identity disclosure whereas McClure and Reiter focus on very specific attacks. The body of work supporting differential privacy shows that the larger ε is, the more likely some identity attack model will succeed, but even large values of ε can effectively protect against specific attack models, when ε is allocated strategically as demonstrated by McClure and Reiter.

DR. RUGGLES' DISCLOSURE RISK ASSESSMENT IS FLAWED AND UNDERESTIMATES THE ACTUAL RISK FROM A RECONSTRUCTION-ABETTED RE-IDENTIFICATION ATTACK.

26. Dr. Ruggles has mischaracterized the risk from reconstructed microdata for the entire population of the 2010 Census. His own risk assessment of the Census Bureau's 2010 data release is flawed, even using the standards of the statistical disclosure limitation literature that pre-dated the invention of privacy-loss accounting methods like differential privacy.

27. Since the influential work of Duncan and Lambert (1989) the risk of identity disclosure for a microdata record has been measured by the probability that the record is a population unique on key variables that can be used for record linkage to external data. As I defined in my first Declaration, population uniques have a combination of key characteristics that occurs exactly once in the entire population. The most basic set of key variables is location, sex and age. A more extensive set is location, sex, age, race and ethnicity.

28. Skinner and Shlomo (2012) use population census data from the United Kingdom to demonstrate how to estimate the risk that a record in a sample corresponds to a population unique in the census and, therefore, requires active disclosure limitation. In all disclosure limitation systems designed since Fellegi (1972) invented the discipline, records containing population uniques on key variables are the highest risk records for re-identification and receive direct disclosure avoidance protection: suppression,

coarsening categories to eliminate uniqueness, noise infusion or some combination of these. Skinner and Shlomo had to predict the probability that a sample record was a population unique because, depending on the sampling rate, records that are unique on the variables in the sample may have many duplicates in the population. They used the UK census to validate their prediction model.

29. In the case of the reconstructed 2010 Census microdata, we know the probability that a record is unique—no estimation is necessary. I presented some summary statistics on the prevalence of population uniques in the 2010 Census in my first Declaration. The location identifier is the census block code. The other two key identifiers are sex and age (in years). As I noted in my first Declaration, in the overall population, 44% of all persons are population uniques on these three variables, making them vulnerable to a classic record linkage attack identical to the one modeled by Duncan and Lambert and by Skinner and Shlomo resulting in a re-identification, when the attacker knows the name of the person associated with the location, sex and age. This is exactly the definition of a re-identification used in the McClure and Reiter paper cited by Dr. Ruggles and in the Wasserman and Zhou paper cited above. This risk assessment is derived from conventional statistical disclosure limitation methods, not differential privacy accounting.

30. Table 1 elaborates on the analysis from my first Declaration. It is based on the actual 2010 Census, not simulated data like those Dr. Ruggles uses. It uses the exact distribution of block populations found in the official Census data and the actual responses on the 2010 Census. Table 1 shows the distribution of the population by the size of the block where the person resides. Only 2.61% of the population lives in blocks with 1 to 9 persons. This is significant because these very small blocks are the ones most likely to be protected by the 2010 Census swapping method. 21.89% of the population live in blocks with 10 to 49 residents, and 22.37% live in blocks with 50 to 99 persons. Fully 46.88% of the population lives in a block with fewer than 100 residents. The column

labeled "Percent of (block, sex, age) Uniques in Bin" shows the percentage of the residents of the block who are unique in their census block, sex and age (in years) values. This percentage ranges from almost everyone (95.06%) in the least populous blocks to very few (1.12%) in the most populous blocks. There are no simulated or reconstructed data used in this table. These are characteristics of the 2010 Census resident population as they appear in the 2010 Census Edited File (CEF). [9]

31. The existence of documented population uniques, even one—not to mention 135 million—triggers mandatory active disclosure limitation, as documented in McKenna (2019b). If presented with a proposed public-use microdata file containing the variables: census block, sex, age (in years), race (OMB-designated coding), and ethnicity (OMB-designated coding) in 1990, 2000, 2010, or 2020, the Census Bureau Disclosure Review Board (or its predecessor) would have insisted on aggregation of the census block codes into more populous geographic areas and would have imposed minimum population sizes (at least 100,000) and minimum population thresholds for the race and ethnicity coding. It would also have insisted on sampling, as documented in McKenna (2019a).

[9] In the swapped version of the 2010 CEF, called the Hundred-percent Detail File, which was actually used for the Summary File 1 tabulations, 43.95% of the persons are population uniques using block, sex and age, almost identical to the 43.87% rate in the CEF.

Table 1
Distribution of Population and Population Uniques by Block Population Size

Block Population Bin	Number of Blocks in Bin	2010 Census Population in Bin	Cumulative Population	Percent of Population in Bin	Cumulative Percent of Population	Population Uniques (block, sex, age) in Bin	Percent of (block, sex, age) Uniques in Bin
TOTAL	11,078,297	308,745,538				135,432,888	43.87%
0	4,871,270	0	0	0.00%	0.00%		
1-9	1,823,665	8,069,681	8,069,681	2.61%	2.61%	7,670,927	95.06%
10-49	2,671,753	67,597,683	75,667,364	21.89%	24.51%	53,435,603	79.05%
50-99	994,513	69,073,496	144,740,860	22.37%	46.88%	40,561,372	58.72%
100-249	540,455	80,020,916	224,761,776	25.92%	72.80%	27,258,556	34.06%
250-499	126,344	42,911,477	267,673,253	13.90%	86.70%	5,297,867	12.35%
500-999	40,492	27,028,992	294,702,245	8.75%	95.45%	1,051,924	3.89%
1000+	9,805	14,043,293	308,745,538	4.55%	100.00%	156,639	1.12%

DRB clearance number CBDRB-FY21-DSEP-003.

32. This table shows that whether the reconstructed 2010 Census microdata are extremely accurate, as the Census Bureau has documented, or whether "[a] much-vaunted database reconstruction technique does not perform significantly better than a crude random number generator combined with a simple assignment rule for race and ethnicity," as Dr. Ruggles (p. 8) claims, asks the wrong question. The reconstructed data are subject to Census Bureau Disclosure Review Board regulation because they contain known population unique identifiers (the combination of census block, sex and age in years). They were produced using tabulations from a confidential Census Bureau data file—the swapped version of the CEF. And they are in record-level format with one record for every person enumerated in the 2010 Census. In their present form, they would not have been certified for release in 2011, when the other 2010 Census data products were released, nor were they certified for release in 2019, when the Census Bureau performed the full reconstruction—even though any person anywhere in the world can perform the same reconstruction because the tables *were* approved for release. The reconstructed 2010 Census data present a clear and present disclosure

risk based on the in-place standards of the Census Bureau, which predate differential privacy by several decades. They also present a clear and present disclosure risk using the traditional methods of assessing such risks, as initiated by Duncan and Lambert, refined by Skinner and Shlomo, and analyzed by the methods used in McClure and Reiter. Indeed, Dr. Ruggles' own institute, IPUMS, acknowledges that national statistical offices, like the U.S. Census Bureau, supply the microdata samples and apply disclosure limitation procedures to those data including, for recent data, limitation of the geographic detail in such microdata files even when they are samples rather than the universe.

33. The traditional standard for applying disclosure limitation methods to microdata is based on the *existence of known unique identifier combinations* in the tabulation variables—census block, sex and age in years, in this case—*not their efficacy in abetting re-identification*. Statistical agencies are expected to document the uniqueness of the identifier—that is done in my previous Declaration and in Table 1—and to continually assess the adequacy of the proposed disclosure limitation methods. Such assessments often involve re-identification studies. Such studies inform the strength of the traditional disclosure limitations applied.

34. Dr. Ruggles claims such studies are not useful because "[i]t would be impossible to positively identify the characteristics of any particular individual using the database reconstruction without access to non-public internal census information" (p. 9). The statement is false because an external agent can also conduct fieldwork or reference multiple commercially available data sources. But even more fundamentally, Dr. Ruggles' statement is irrelevant because it is the agency's duty to protect the confidentiality of the microdata and therefore it must, just as in cybersecurity, assume that attackers are clever enough to gather information that confirms the efficacy of their attacks.

35. Table 2 shows that the reconstruction-abetted re-identification attack simulated by the Census Bureau has very high precision precisely in the blocks that are most vulnerable to such an attack, whether one uses the best-case or worst-case analysis. In blocks with populations between 1 and 9 persons, the re-identification attack has a precision of 72.24% when using commercial data available in 2010.[10] Almost all of Dr. Ruggles' precision comes from the most populous blocks, whereas his precision plummets in sparsely populated blocks. In these sparsely populated blocks, the re-identification attack is much more precise than Ruggles' model. The exact block population was public information following the release of the 2010 Census data (as it may be in 2020 if plaintiffs succeed here). That means an attacker has a clean, public predictor of the success of the re-identification attack. Fieldwork in sparsely populated blocks can confirm this precision, as can sophisticated Bayesian methods like entity resolution without field work (Steorts, Hall and Fienberg 2016). If the attacker has better quality name, address, sex and age data than were available in 2010, certainly a plausible assumption, then the worst-case analysis for blocks with populations of 1 to 9 is precision of 96.98%--more precise than the 95% confidence interval test often used in statistics. Again, this can be confirmed by fieldwork or Bayesian entity resolution. The situation is only a little better for the 68 million people who live in blocks with populations of 10 to 49. The precision of the 2010-era commercial data is 53.61%--correct more than half the time, and the precision with high-quality external data is 91.68%. Although the best-case precisions for block populations of 50 or more are less than one-half, the worst-case precision, even in the most populous blocks, is always greater than one-half—*an attacker with high quality external data is always more likely to be correct than wrong.* As I reported in my first Declaration, with high-quality data, the attacker

[10] Precision is the rate at which putative re-identifications are confirmed. A precision of zero indicates the putative re-identification is never correct. A precision of 100% indicates that it is always correct.

is correct on average three times out of four regardless of the number of persons who live in the block.

Table 2
Disclosure Risk Assessment of Population Uniques by Block Population Size

Block Population Bin	Putative Re-identifications (Source: Commercial Data)	Confirmed Re-identifications (Source: Commercial Data)	Precision (Source: Commercial Data)	Putative Re-identifications (Source: CEF)	Confirmed Re-identifications (Source: CEF)	Precision (Source: CEF)
TOTAL	137,709,807	52,038,366	37.79%	238,175,305	178,958,726	75.14%
0						
1-9	1,921,418	1,387,962	72.24%	4,220,571	4,093,151	96.98%
10-49	25,148,298	13,481,700	53.61%	47,352,910	43,415,168	91.68%
50-99	30,567,157	12,781,790	41.82%	51,846,547	42,515,756	82.00%
100-249	38,306,957	13,225,998	34.53%	63,258,561	45,807,270	72.41%
250-499	21,789,931	6,408,814	29.41%	35,454,412	22,902,054	64.60%
500-999	13,803,283	3,460,118	25.07%	23,280,718	13,514,134	58.05%
1000+	6,172,763	1,291,984	20.93%	12,761,586	6,711,193	52.59%

DRB clearance number CBDRB-FY21-DSEP-003.

36. The Data Stewardship Executive Policy Committee (DSEP) determined that the simulated attack success rates in Table 2 were unacceptable for the 2020 Census. Decennial census data protected by the 2010 disclosure avoidance software is no longer safe to release. Dr. Ruggles takes issue with this conclusion for the following three reasons.

37. First, he asserts "[t]he reconstructed data are usually incorrect" (p. 10). That is not the standard. The reconstructed data are always correct for block and voting age and have at most one error—not in block or voting age—for 78% of the population. The disclosure avoidance problem is that the reconstructed data contain a known unique key (census block, sex and age in years), and therefore would be subject to the microdata disclosure avoidance rules—which mandate coarsening of the geographic identifier to areas with populations of at least 100,000.

38. Second, he asserts "[t]he reconstructed data usually do not match even the block, age and sex of anyone identified in outside commercial sources" (p. 10). That assertion

uses the wrong standard as well. What matters is the precision of the re-identification, not the absolute rate. That precision is predictably very large precisely for the population the swapping system was supposed to protect—those in sparsely populated blocks like those with a population of 1-9 people (72.24% confirmed in Commercial Data) and 10-49 people (53.61% confirmed in Commercial Data).

39. Finally, he asserts "[i]n the minority of cases where a hypothetical reconstructed individual does match the block, age and sex of someone in the commercial data, it usually turns out that the person identified in the commercial data was not actually enumerated on that block in the census" (p.10). The most favorable interpretation of his assertion is that it is based on the average precision in the best case (38%), but even under the best case, the precision is greater than half (the attacker is usually right) for the 76 million people who live on blocks with populations less than 50 people. But the Census Bureau does not calibrate its disclosure avoidance systems based on the best case because that would be irresponsible. Instead, the Census Bureau has historically relied on conservative analyses, closer to worst-case than best case, to calibrate disclosure avoidance for public-use microdata files (McKenna 2019a).

40. By relying on a simplistic and flawed analysis, Dr. Ruggles and the plaintiffs claim that reconstruction-abetted re-identification is impossible. That is wrong and accepting that view would put the privacy of millions of Americans at risk. As I showed in my prior Declaration and supplemented in this Declaration the risk to the American public from these types of attacks will certainly grow over the coming years. The risk was confirmed by non-political, career experts who serve on the Census Bureau's Data Stewardship Executive Policy Committee. And DSEP decided—based on data—that using state-of-the-art differential privacy to implement the TopDown Algorithm was the best way to protect against those real-world threats.

I declare under penalty of perjury that the foregoing is true and correct.

DATED and SIGNED: JOHN ABOWD Digitally signed by JOHN ABOWD Date: 2021.04.26 15:47:57 -04'00'

_____

John M. Abowd

Chief Scientist and Associate Director for Research and Methodology

United States Bureau of the Census

References

Cohen, A., and K. Nissim. 2020. "Linear Program Reconstruction in Practice." Journal of Privacy and Confidentiality 10 (1). https://doi.org/10.29012/jpc.711. Conference version. 2018. Theory and Practice of Differential Privacy [1810.05692] Linear Program Reconstruction in Practice (arxiv.org).

Cox, L. H. 1976. Statistical Disclosure in Publication Hierarchies. Report No. 14 of the research project Confidentiality in Surveys. Department of Statistics, University of Stockholm, Stockholm.

Duncan, G., and D. Lambert. 1989. "The Risk of Disclosure for Microdata." Journal of Business and Economic Statistics, 7(2):207-217. doi:10.2307/1391438

Evans, T., L. Zayatz, J. Slanta. 1998. "Using Noise for Disclosure Limitation of Establishment Tabular Data." Journal of Official Statistics, 14(4): 537. 551 Using Noise for Disclosure Limitation of Establishment Tabular Data (scb.se).

Fellegi, I. P. 1972. "On the question of statistical confidentiality," Journal of the American Statistical Association, 67:7-18.

McClure, D. and J Reiter. 2012. "Differential Privacy and Statistical Disclsure Risk Measures: An Investigation with Binary Synthetic Data." Transactions on Data Privacy, 5:535–552.

McKenna, L. 2019. "Disclosure Avoidance Techniques Used for the 1960 Through 2010 Census." https://www.census.gov/library/working-papers/2019/adrm/six-decennial-censuses-da.html. Retrieved April 23, 2021.

McKenna, L. 2019b. "U.S. Census Bureau Reidentification Studies," https://www.census.gov/library/working-papers/2019/adrm/2019-04-ReidentificationStudies.html. Retrieved April 23, 2021.

Skinner, C. and N. Shlomo. 2008. "Assessing Identification Risk in Survey Micro-data Using Log-Linear Models. Journal of the American Statistical Association," 103(483): 989-1001. Retrieved April 23, 2021, from http://www.jstor.org/stable/27640138

Steorts, R. C., R. Hall and S. E. Fienberg. 2016. "A Bayesian Approach to Graphical Record Linkage and Deduplication," Journal of the American Statistical Association, 111(516):1660-1672, DOI: 10.1080/01621459.2015.1105807.

Warner, S. L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." Journal of the American Statistical Association, 60(309): 63–69.

Wasserman, L., and S. Zhou. 2010. "A Statistical Framework for Differential Privacy." Journal of the American Statistical Association, 105(489): 375–389.

Appendix

U.S. Department of Justice
Civil Rights Division
Deputy Assistant Attorney General
Washington, D.C 20530

March 15, 2000

Dr. Kenneth Prewitt
Director
Department of Commerce
Bureau of the Census
Washington, D.C. 20233

Dear Dr. Prewitt:

This is in regard to the Census 2000 Redistricting Data and confirms the February 25, 2000 telephone conversation between Marshall Turner of your staff and my Deputy, Anita Hodgkiss.

At a November 29, 1999 meeting here at the Civil Rights Division, you and your staff discussed with Ms. Hodgkiss the possible need to apply certain disclosure avoidance techniques to the detailed race and ethnicity statistics included in the Census 2000 Redistricting Data files in order to meet the confidentiality requirements of Title 13, U.S.C. We have reviewed the information you provided at the November 29 meeting and we do not believe that the application of these disclosure avoidance techniques will impair the use of these data for enforcement of civil rights programs.

We greatly appreciate your kind assistance.

Sincerely,

Bill Lann Lee
Acting Assistant
Attorney General
Civil Rights Division

EXHIBIT 2

2020 Disclosure Avoidance System Updates

The Census Bureau is working closely with our data users as we modernize the privacy protections for the 2020 Census. We are reporting 2020 Disclosure Avoidance System (DAS) developments here, in our blogs [/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html#blogs] , and in our digital newsletter (Subscribe [https://public.govdelivery.com/accounts/USCENSUS/signup/15409] | Archived Issues [/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/newsletters.html]).

We appreciate your engagement and encourage you to email comments and suggestions to 2020DAS@census.gov [mailto:2020DAS@census.gov]

EXPAND ALL | COLLAPSE ALL

4/19/2021: New Demonstration Data Will Feature Higher Privacy-loss Budget, Satisfies Redistricting Accuracy Targets

Based on the results of over 600 experimental data runs to optimize and tune the parameters of the new 2020 Census Disclosure Avoidance System (DAS) algorithm, the Census Bureau's Data Stewardship Executive Policy Committee (DSEP) has chosen the privacy-loss budget (PLB) for the forthcoming set of demonstration data.

The global privacy-loss budget (PLB) for the persons file in the next demonstration data set will be 10.3 and the PLB for the housing units data will be 1.9. As discussed previously [https://lnks.gd/l/eyJhbGciOiJIUzI1NiJ9.eyJidWxsZXRpbl9saW5rX2lkIjoxMDIsInVyaSI6ImJwMjpjbGljayIsImJ1bGxldGluX2lkIjoiMjAyMTA0MTkuMzkwMzAyNDEiLCJ1cmwiOiJodHRwczovL2NvbnRlbnQuZ292ZGVsaXZlcnkuY29tL2FjY291bnRzL1VTQ0VOU1VTL2J1bGxldGlucy8yYmU2MjZmIn0.Kvvn7zjEufv5BK4NDFTZSUt9b-y_xUTwJ9aA1szbUXs/s/600744772/br/102755986114-l] , the four demonstration products released to date used a PLB of 4.0 for persons and 0.5 for housing units—significantly lower than we anticipate using for the final 2020 Census data. Those earlier demonstration data were purposefully "tuned" to privacy and not "tuned" for producing highly accurate redistricting data. We held the PLB roughly the same across those four releases to allow us to compare the effects of incremental algorithmic improvements in the system.

While significantly larger than the PLB used in the previous data, the 10.3 PLB is still allocated in a manner that provides a level of protection for every census record and every published characteristic. For those of you trying to understand the increase in accuracy attributable to a shift from a PLB of 4 to a PLB of 10.3, it's important to understand that the PLB is logarithmic—meaning every additional number in the PLB scale represents an exponential increase in the PLB. The forthcoming demonstration data, released as Privacy-Protected Microdata Files (PPMFs) will help data users see that increase in PLB reflected in the accuracy of population counts and demographic characteristics at various levels of geography.

The Census Bureau announced the new PLB in a declaration [https://lnks.gd/l/eyJhbGciOiJIUzI1NiJ9.eyJidWxsZXRpbl9saW5rX2lkIjoxMDMsInVyaSI6ImJwMjpjbGljayIsImJ1bGxldGluX2lkIjoiMjAyMTA0MTkuMzkwMzAyNDEiLCJ1cmwiOiJodHRwczovL3d3dzIuY2Vuc3VzLmdvdi9hYm91dC9wb2xpY2llcy9mb2lhL3JlY29yZHMvYWxhYmFtYS12cy1kb2MvYWxhYmFtYS1paS00MS1kZWZzLXBpLW9wcG9zaXRpb24tYW5kLWRlY2xhcmF0aW9ucy5wZGY_dXRtX2NhbXBhaWduPTIwMjE

wNDE5bXNkZWNzMWNjZHRhciZ1dG1fbWVkaXVtPWVtYWlsJnV0bV9zb3VyY2U9Z292ZGVsaXZlcnkifQ
.IYoT5T_U5HtPCSb9OaKQCaKGaQYRXJT8WqdLCygxb5U/s/600744772/br/102755986114-l]
submitted in response to litigation on April 13, 2021.

New Demonstration Data Will Satisfy Redistricting Use Case Needs

In the same declaration
[https://lnks.gd/l/eyJhbGciOiJIUzI1NiJ9.eyJidWxsZXRpbl9saW5rX2lkIjoxMDQsInVyaSI6ImJwMjpjbGljayIsImJ1bGxldGluX2lkIjoiMjAyMTA0MTkuMzkwMzAyNDEiLCJ1cmwiOiJodHRwczovL3d3dzIuY2Vuc3VzLmdvdi9hYm91dC9wb2xpY2llcy9mb2lhL3JlY29yZHMvYWxhYmFtYS12cy1kb2MvYWxhYmFtYS1paS00MS1kZWZzLXBpLW9wcG9zaXRpb24tYW5kLWRlY2xhcmF0aW9ucy5wZGY_dXRtX2NhbXBhaWduPTIwMjEwNDE5bXNkZWNzMWNjZHRhciZ1dG1fbWVkaXVtPWVtYWlsJnV0bV9zb3VyY2U9Z292ZGVsaXZlcnkifQ.8K5vlizgZ5pa4CUJ5c-Bi7925EnrAQthSR_6yquNmRA/s/600744772/br/102755986114-l] we previewed some of the high-level results from the upcoming demonstration data release regarding the accuracy criteria established for the P.L. 94-171 redistricting data (see our previous newsletter
[https://lnks.gd/l/eyJhbGciOiJIUzI1NiJ9.eyJidWxsZXRpbl9saW5rX2lkIjoxMDUsInVyaSI6ImJwMjpjbGljayIsImJ1bGxldGluX2lkIjoiMjAyMTA0MTkuMzkwMzAyNDEiLCJ1cmwiOiJodHRwczovL2NvbnRlbnQuZ292ZGVsaXZlcnkuY29tL2FjY291bnRzL1VTQ0VOU1VTL2J1bGxldGlucy8yY2I3NDViIn0.EfM4F8PuQdagdlQInOrl4UunuOWs6krpR9kESSPi_ck/s/600744772/br/102755986114-l] for criteria details).

We report that the new demonstration data will fully satisfy those specialized accuracy criteria. Specifically, populations, voting-age populations, and the proportion of the largest OMB-designated race and ethnicity groups are all reliable for redistricting and Voting Rights Act scrutiny. Because new districts cannot be drawn before the 2020 P.L. 94-171 Redistricting Data Summary File is released, counties, block groups, minor civil divisions, incorporated places, and census designated places were all used as on- and off-spine geographic entities for tuning purposes.

Added Noise from Differential Privacy is Less Than Inherent Census Errors

The declaration
[https://lnks.gd/l/eyJhbGciOiJIUzI1NiJ9.eyJidWxsZXRpbl9saW5rX2lkIjoxMDYsInVyaSI6ImJwMjpjbGljayIsImJ1bGxldGluX2lkIjoiMjAyMTA0MTkuMzkwMzAyNDEiLCJ1cmwiOiJodHRwczovL3d3dzIuY2Vuc3VzLmdvdi9hYm91dC9wb2xpY2llcy9mb2lhL3JlY29yZHMvYWxhYmFtYS12cy1kb2MvYWxhYmFtYS1paS00MS1kZWZzLXBpLW9wcG9zaXRpb24tYW5kLWRlY2xhcmF0aW9ucy5wZGY_dXRtX2NhbXBhaWduPTIwMjEwNDE5bXNkZWNzMWNjZHRhciZ1dG1fbWVkaXVtPWVtYWlsJnV0bV9zb3VyY2U9Z292ZGVsaXZlcnkifQ.L3-mkPTkwDl4ZY44T71zYeLeUZqHAtW6Sao8np54oeY/s/600744772/br/102755986114-l] also revealed high-level results of an analysis comparing the error caused by the new differentially private methods to the other sources of error that are inherent in census data (coverage error, measurement error, etc.) based on post-census analyses.

Our internal analyses have shown that 2010 Census operations resulted in an average county-level estimation of uncertainty in terms of total population of +/- 960 people (averaging 1.6% of the county census counts). The new demonstration data, by comparison, has an average error of only +/- 5 people at the county level (reflecting a mean absolute percent error of 0.04% of the counties' population) as noise from differential privacy.

At the block level the differentially private data have an average population error of +/- 3 people, which includes both housing unit and group quarters populations. Compare that with the simulated error inherent in the census that puts the average uncertainty of block population counts at +/- 6 people.

We'll share more information about these results in our next newsletter.

Next Steps

Per the calendar below, we will release the new demonstration data by April 30. We'll release two versions to aide in your analysis: a version using the new (10.3, 1.9) PLB, and one using the earlier, development-focused PLB (4.0, 0.5).

We'll look forward to your feedback after that release.

2021 Key Dates, Redistricting (P.L. 94-171) Data Product:

By April 30:

- Census Bureau releases new Privacy-Protected Microdata Files (PPMFs) and Detailed Summary Metrics.
- Two versions: Candidate strategy run using new PLB and old PLB.

By late May:

- Data users submit feedback.

Early June:

- The Census Bureau's Data Stewardship Executive Policy (DSEP) Committee makes final determination of PLB, system parameters based on data user feedback for P.L. 94-171.

Late June:

- Final DAS production run and quality control analysis begins for P.L. 94-171 data.

By August 16:

- Release 2020 Census P.L. 94-171 data as Legacy Format Summary File*.

September:

- Census Bureau releases PPMFs and Detailed Summary Metrics from applying the production version of the DAS to the 2010 Census data.
- Census Bureau releases production code base for P.L. 94-171 redistricting summary data file and related technical papers.

By September 30:

- Release 2020 Census P.L. 94-171 data** and Differential Privacy Handbook.

* Released via Census Bureau FTP site.

** Released via data.census.gov.

- 4/7/2021: Meeting Redistricting Data Requirements: Accuracy Targets
- 2/23/2021: The Road Ahead: Upcoming Disclosure Avoidance System Milestones
- 2/3/2021: Fine-Tuning the Disclosure Avoidance System to Ensure Accuracy
- 11/25/2020: Invariants Set for 2020 Census Data Products
- 11/16/20: Now Online: Revised Privacy-Protected Microdata Files

10/23/20: Issue Discovered in the Privacy-Protected Microdata File

9/24/2020: New Metrics and Demonstration Data

9/17/20: New Privacy-Protected Microdata Files (PPMFs)

8/26/20: New Frequently Asked Question re: DAS Development Schedule

7/14/20: New Privacy-Protected Census Demonstration Data

7/1/20: Census Bureau Partners with Committee on National Statistics to Produce New Demonstration Data Files

6/26/20: New Frequently Asked Questions

6/1/20: New Metrics and DAS Updates Presentations from CNSTAT Expert Meeting on Disclosure Avoidance

5/27/20: Release of "2010 Demonstration Metrics 2;" First Set of Post-Baseline Quality Metrics Results

5/18/20: New 2020 Census Data Products Newsletter

4/20/20: New Frequently Asked Questions

4/17/20: Video: DAS Development Status Update

3/27/20: Draft Quality Measures and Benchmark

Related Information

Developing the DAS: Demonstration Data and Progress Metrics [/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-da development.html]

2020 Disclosure Avoidance System: Frequently Asked Questions [/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates/2020-das-faqs.html]

Disclosure Avoidance and the 2020 Census [/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html]

Last Revised: April 20, 2021