MIT Case Studies in Social and Ethical Responsibilities of Computing

Differential Privacy and the 2020 US Census

Simson Garfinkel¹

¹Data Science Program, George Washington University

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

ABSTRACT

In 2017 the US Census Bureau announced that it would update the statistical methods it used to meet its statutory obligation to protect the confidentiality of information furnished to the agency. The new system would be based on an approach called “differential privacy” that was invented in 2006 in the context of cryptography. Unlike the Census Bureau’s previous system—a technique based on swapping responses from easy-to-identify residential units—the confidentiality protection afforded by differential privacy involves adding carefully structured random values (called “noise”) to every intermediate computation, and then implementing “postprocessing” algorithms to make the noise-infused data resemble the legacy data produced by the previous Census Bureau methods.¹ Although for decades the Census Bureau intentionally added error to the decennial census tables and publicly released microdata to protect the privacy of respondents, it had never revealed how much error was added nor the exact mechanism. For 2020, the Census Bureau decided to engage stakeholders directly and tell data users how much error would be introduced. As the Census Bureau raced against the clock to implement the new system, thousands of stakeholders became alarmed that differential privacy might introduce so much randomness into the published data that they would be unusable for many purposes, including drawing new districts for the US House of Representatives as well as academic research in demography and economics. In the interest of transparency, the Census Bureau produced multiple Demonstration Data Products (DDP) using the 2010 Census data. Still, the bureau’s Disclosure Avoidance System (DAS) team made significant changes to the system between the release of the final DDP and the official redistricting statistics for the 2020 Census. The Census Bureau never released information about how it had decided on the “privacy loss budget” for the 2020 Census Redistricting File.

Keywords: differential privacy, disclosure avoidance, statistical disclosure limitation, US Census Bureau

Simson Garfinkel

Data Science Program, George Washington University

Author Disclosure(s): Between 2017 and 2021, the author was the Senior Computer Scientist for Confidentiality and Data Access at the US Census

Bureau, where he was the lead computer scientist working on the 2020 Disclosure Avoidance System. Between 2017 and 2019 the author was also the Chair of the Census Bureau's Disclosure Review Board. The author was previously Chief of the Census Bureau's Center for Disclosure Avoidance Research. The author left the Census Bureau in April 2021 and was not personally involved in the setting of the privacy loss budget for the 2020 Census. The author is also a part-time faculty member at George Washington University and a member of the Association for Computing Machinery's US Technology Policy Council, where he co-chairs the USTPC's Digital Governance Committee. This article is written in the author's private capacity from public sources and does not represent the views or policy of the US Government, the Department of Commerce, the US Census Bureau, or the US Department of Homeland Security.

Learning Objectives

- Describe the concept of differential privacy.
- Discuss the tension between using data for the public good and protecting the confidentiality of those who provided the data.
- Name potential social and technical costs of increased transparency on government statistical agencies.
- Discuss challenges facing engineers trying to deploy systems for high-stakes projects on tight deadlines.
- Distinguish how the constraints facing a government agency's use of technology are different from a private company making similar decisions.
- Identify challenges of public engagement on technical issues in which different communities have interests that are inherently in opposition and different ways of measuring success.

Introduction: The US Census

Seats in the US House of Representatives are allocated based on the population of each state, so the Founders wrote into the US Constitution a requirement that the newly created nation conduct an "actual Enumeration," or census, of each state's population within three years of the first meeting of the US Congress, and every ten years following.

First overseen by Thomas Jefferson in 1790, the US census is the world's longest-running statistical program. Although the Census Bureau is designed to be a scientific organization that is largely protected from politics, the profound impact that the decennial count has on raw political power all but guarantees that many of its

technical decisions will be end up in court. This case study explores the controversial decision by the Census Bureau to intentionally add error to the results of the 24th Census, better known as the 2020 Census, to protect the privacy of respondents using a mathematical technique called “differential privacy” (DP). The Census Bureau did not have a working implementation of its DP system when it decided to use the method. It had no backup plan for the possibility that its scientists might not create a functional approach in time for the legally mandated statistical publication.

In addition to apportioning the House of Representatives, data created as part of the Decennial Census of Population and Housing establishes baseline demographic statistics that calibrate the American Community Survey and many of the nation’s other statistical programs.² The US Department of Justice also uses census data for enforcing Section 2 of the Voting Rights Act³ and ultimately to control the distribution of an estimated \$1.5 trillion in federal spending.⁴ There is thus considerable interest in having the released statistics be as accurate as possible.

Complaints about the census go back to the first. “One thing is certain: our real numbers will exceed, greatly, the official returns of them,” quipped George Washington when the 1790 Census turned up just 3.9 million people, including 700,000 slaves.⁵

As evidenced by the 1790 count, the census also reflects some of the nation’s darkest history. Until the ratification of the 14th Amendment, the Constitution stated that an enslaved person would count only “three fifths” toward each state’s enumeration—a mathematical formula designed to give the slave-holding southern states political leverage over the North. Likewise, the Constitution explicitly excluded “Indians not taxed” from the count, and thus Native Americans were not systematically counted until the end of the 19th Century.⁶ Arguments over who gets counted and how they are counted continue to be part of the decennial exercise. The adoption of differential privacy added a scientific controversy to the mix.

Interactive Gallery

Hover over the boxes to visualize historical documents and additional information about race data.

Visit the web version of this article to view interactive content.

Courtesy of US Census Bureau

The first law governing the decennial census required that the agents conducting the census post the enumeration returns in “two of the most important places” in each

enumeration district so that they could be publicly inspected and corrected before the results were sent to Congress.⁷ This practice was stopped in 1850 over concerns that the lack of confidentiality was lowering the response rate. In 1870, census superintendent Francis Amasa Walker increased confidentiality protections, and the Census Act of 1879,⁸ which governed the 1880 Census, required that enumerators take an oath to avoid improperly disclosing respondent data.⁹ A version of that oath remains in effect today for all Census Bureau employees.

The Census Bureau is thus charged with collecting and publishing statistics that are accurate, yet simultaneously protecting the confidentiality of information that is provided by respondents. This charge is codified today in Title 13, Section 9 of the Census Act ([Figure 1](#)). (Here the term “confidentiality” is used for data that the Census Bureau collects and holds in confidence; the term “privacy” is typically reserved for data that respondents keep to themselves and do not share with the government. This case study also uses the term “privacy” to describe information protected with differential privacy.)

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may, except as provided in section 8 or 16 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998 or section 2(f) of the Census of Agriculture Act of 1997—

(1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or

(2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or

(3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports.

No department, bureau, agency, officer, or employee of the Government, except the Secretary in carrying out the purposes of this title, shall require, for any reason, copies of census reports which have been retained by any such establishment or individual. Copies of census reports which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.

(b) The provisions of subsection (a) of this section relating to the confidential treatment of data for particular individuals and establishments, shall not apply to the censuses of governments provided for by subchapter III of chapter 5 of this title, nor to interim current data provided for by subchapter IV of chapter 5 of this title as to the subjects covered by censuses of governments, with respect to any information obtained therefor that is compiled from, or customarily provided in, public records.

Figure 1

Title 13, Section 9 of the US Census Act.

In addition to downloading data from the Census Bureau’s website (or acquiring the data from one of the many organizations that download, enhance, and repackage the data), several hundred researchers in the United States have the ability to directly access the Census Bureau’s raw, confidential data. To do so, each researcher must have an approved research project and undergo a background investigation. Even then, the data can only be accessed from one of the government’s Federal Statistical Research Data Centers (FSRDCs). Before a researcher can remove data from the FSRDC for publication, or even for mere discussion with uncleared colleagues, the data must be reviewed and approved for release by the Census Bureau’s Disclosure Review Board to ensure that the release will not violate Title 13 Section 9.

US law also recognizes a public interest in *eventually* making the census data public. For example, historians and family researchers use the data to help understand the past. Currently, that “eventually” is 72 years, meaning that the raw data collected in 2020 will be made publicly available in 2092.^{[10](#)}

Disclosure Avoidance and the 2020 Census

Field operations for the 2020 Census started in August 2019 with address canvassing, in which the Census Bureau created a database of every place in the United States where people lived. This database, called the “frame,” included houses, apartments, “group quarters” like college dorms and nursing homes, and designated places for allocating people experiencing homelessness.^{[11](#)} As part of this operation, the Census Bureau obtained addresses from other federal agencies and Native American tribes, and used satellite imagery to validate 65% of all residences in the United States. Roughly 40,000 people were then hired to validate the remaining 35% of the addresses in person.^{[12](#)}

The 2020 Census originally planned to collect six pieces of information for each resident: their name, age, sex, race,^{[13](#)} ethnicity,^{[14](#)} and in the case of those living in residences, their relationship to the “head” of the household. The most important computation that the Census Bureau performs with these data is the tally of how many people are living in each state, which determines the number of seats that each state will have in the House of Representatives for the rest of the decade.^{[15](#)} This information was supposed to be delivered to the president by December 31, 2021, followed by the 2020 Census Redistricting Data Files on March 31, a file that would contain block-by-block tabulations of the number of people by race and voting age for each “census block.”^{[16](#)} It would be followed by a demographic profile that would have detailed age

pyramids by sex and demographic information about families, such as the number of households headed by same-sex couples.

Some kinds of demographic information can be very sensitive for some populations. The United States is a hugely varied place, and much of that variation is captured in demographic data. The collected data identifies elderly women living alone and teenage children living with a single parent. It identifies same-sex couples with children, families that are mixed-race, and even adults in possibly polygamous relationships. The raw data also make it easy for a housing authority to find units that have more occupants than allowed—as might be the case if someone in subsidized housing is renting out a spare room. In 2020, the Donald Trump administration further attempted to collect a seventh piece of information about each respondent: whether or not the person was a citizen of the United States. Many civil rights advocates argued that even if the citizenship information was not used for immigration enforcement—which would violate Title 13—the mere attempt to collect such data would decrease the response rate within immigrant communities.

Protecting the confidentiality of these data from improper disclosure requires more than simply stripping the names from the individual records: it requires a variety of statistical techniques so that the Census Bureau data publications cannot be reverse-engineered to reveal the underlying confidential data. The Census Bureau calls such techniques “disclosure avoidance,” because they are designed to prevent an improper disclosure of confidential information. (See [Table 1](#))

Table 1			
<i>Legacy Disclosure Avoidance Methods</i> ¹⁷			
Method	Explanation	Example	Disadvantage
Suppression	Data are suppressed in a table.	Values in a table from less than 10 houses replaced with a “D.”	Removes information that is useful to data users. Ironically, suppression reveals information that is useful when reconstructing the suppressed information using other data.
Controlled Rounding or “coarsening”	Values are rounded so that specific values cannot be discerned.	The number of delinquent children by county in a table might be rounded to the nearest five children.	It may be possible to use other data to recover or sharpen the coarsened information.

Limiting or “coarsening” geographic detail	Locations are reported using less specific geographical area.	Street, city, and county in a microdata record are replaced with the record’s state.	It may be possible to use other data to recover or sharpen the coarsened information.
Top-coding and bottom-coding	Values above or below specific thresholds are reported at the threshold.	Incomes over \$90,000 are reported as “\$90,000 or more.”	Outliers are hidden from the analysis, which may have public policy implications.
Data swapping and rank swapping	For selected records, some attributes are swapped or exchanged to hide the identity of the data subjects.	First name and education level for a classroom are swapped with those of a classroom in another county.	Artifacts are introduced into the data, decreasing analytical utility. If the artifacts can be detected by data users, it may be possible to undo the swapping mechanism.
Blank and impute	Attributes for some records are removed and replaced with values generated by a statistical model.	Instead of reporting that a person’s income is \$54,320, the person’s income is reported as \$57,230 based on a model.	Artifacts are introduced into the data, decreasing analytical utility. If the imputed values can be detected by data users, it may be possible to undo the swapping mechanism.
Noise injection or “infusion”	Random noise is added to values.	Noise drawn from a distribution is added to every income in the microdata (additive noise), or incomes are multiplied by random values between 0.95 and 1.05 (multiplicative noise).	Bias can be introduced with the noise if negative values are rounded up to zero. It is difficult to know how much noise is the right amount to add.

Note: All of these methods have the advantage that they are relatively straightforward to explain and implement. They all have the disadvantage that they do not “compose,” which means that they cannot be combined in more-or-less arbitrary ways with predictable and incremental results. For example, applying suppression to a data set that has already been top-coded may not yield additional protection. Indeed, confidential information can be revealed if a single data set serves as the basis for two separate data products, if the two data products are protected using different methods.

The Census Bureau started applying disclosure avoidance techniques to the decennial census in 1960 because of advances in computing and has increased the amount of protection each decade.^{[18](#)}

In 1960, the increased availability of computing resources over the previous decade made it useful for the Census Bureau to distribute a so-called “public use microdata

sample” (PUMS) of the 1960 Census to researchers. The PUMS consisted of 180,000 individual records, chosen at random, and distributed on punch cards and magnetic tape, allowing researchers to perform their own analyses of the data.^{[19](#)}

To protect the confidentiality of the respondents in the PUMS, the “direct identifiers” (names) were removed from the records, and the addresses were replaced with a code that denoted a Public Use Microdata Area (PUMA) that did not cross state lines and had a population threshold of 250,000. In the 1970 Census, the 250,000 threshold was also applied to neighborhood characteristics. In the 1980 Census, household income was rounded to \$10 and top-coded at \$75,000; age was top-coded at 90. Starting in 1990, these and other “confidentiality edits” were applied to the confidential microdata, rather than just to the PUMS, and the edited data were used to produce both the PUMS *and* the official tables. In 2000 the Census Bureau started the practice of “data swapping,” in which some of the attributes of highly identifiable households were swapped with other households (but always within the same state); it also added noise infusion to protect some characteristics in large households. The 2010 Census also used “partially synthetic data” to protect the characteristics of some individuals living in group quarters. But despite the fact that disclosure avoidance has been used for five decades, the Census Bureau did not publicize the details of the confidentiality protecting techniques, and many data users were not even aware of them.

In addition to the confidentiality edits, the Census Bureau made many other changes to the data between what was collected and what was used to produce the data publications. One of the most controversial techniques, called “hot-deck imputation,” was used to synthesize the characteristics of occupied housing units that had not filled out the survey, had not been home when visited by an enumerator, and for which a “proxy enumeration” was not collected from a neighbor. In these cases, Census Bureau employees considered similar households and then created synthetic data. Hot-deck imputation was so controversial that it was the subject of the US Supreme Court case *Utah v. Evans* (2002).^{[20](#)} In that case, the technique increased North Carolina’s population by 0.4% but only increased Utah’s population by 0.2%. As a result of the imputation, Utah lost a seat in the House of Representatives while North Carolina gained a seat. Ultimately, fewer than 100 imputed counts were responsible for the switch. The US Supreme Court sided with Donald Evans, the Secretary of Commerce, stating that the use of imputation did not violate the US Constitution or the Census Act, and North Carolina got to keep the additional seat.

Differential Privacy

Statisticians have long understood that multiple data products made from the same confidential data can inadvertently reveal protected facts. Mathematically, it is possible to combine the multiple statistical releases into a single system of equations and then solve the equations, revealing the underlying confidential data. Consider a simple example: if three people live on a block and the statistical agency reports that their median age is 30 and their mean age is 44, then there are only 30 possible combinations of integer ages that the three people can have. If another data publication reveals that the oldest person on the block is 80, then the three people's ages must be 22, 30, and 80.^{[21](#)}

Given these sorts of vulnerabilities, in 2018 the Census Bureau announced that the 2020 Census required a radical new approach to protect respondent confidentiality. In fact, it said, the protections that had been used for the 2010 Census could now be undone using a combination of fast computers, modern algorithms, and commercially available data.

The root of the problem was that the Census Bureau had published dozens of different statistical tables based on the 2010 confidential data. One reported the precise number of individuals of each race on each census block. Another reported the precise number of people for each year of age within each census tract. Others reported still more precise statistics. Using powerful computers that were now available, it was possible to construct a set of simultaneous equations encompassing all of the published tables and solve for the underlying confidential data. This kind of attack is called a “database reconstruction.”^{[22](#)} Once the underlying data are reconstructed, each person's age, sex, and census block can be matched against a commercial database, revealing that person's name. But the reconstructed data *also* includes each person's self-reported race and ethnicity—information that is generally not commercially available, and especially not generally available for children.

Database reconstruction isn't just a hypothetical risk: the Census Bureau conducted a database reconstruction attack on the published 2010 data and discovered that it could accurately reconstruct the confidential microdata for 144 million people, or 46% of the US population. For another 76 million, all five variables were correct except for age, which was off by a single year. Finally, using commercially available data from 2010, the Census Bureau discovered that it could reidentify data from 52 million people, or 17% of the US population.^{[23](#)}

How could the Census Bureau’s 2010 disclosure avoidance protection have been so terribly lacking? The techniques were designed to protect outliers—people whose age or race made them readily identifiable. The protection system was never designed to protect against a database reconstruction attack that used *all* of the published tables, because such an attack wasn’t considered feasible in 2010. But the emergence of cloud computing combined with advances in commercially available “solvers” made the reconstruction attack not just feasible, but straightforward.

For its new protection system, the Census Bureau turned to a technique called “differential privacy” (DP). Developed in 2006 by Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, DP is not based on assumptions about the amount of computing power or external data available to the attacker: it is based on adding intentional error to published statistics to create uncertainty about the confidential data on which the published statistics are based.²⁴ Unlike other protection frameworks, DP doesn’t create safe, protected data from which the underlying confidential values cannot be reverse-engineered. Instead, DP formalizes a concept of a “privacy loss” that results from the use of confidential data to create published statistics, and provides an accounting framework for tracking the maximum amount of privacy loss that occurs if the same confidential data are used in multiple publications.

Visit the web version of this article to view interactive content.

Protecting Privacy with MATH (Collab with the Census)

DP is a “noise infusion” technique, and its use would have similarities with previous disclosure avoidance techniques that the Census Bureau had employed:

- The noise would be added to the confidential data to produce a new data set that would be used to create the official tabulations. Thus, a successful reconstruction attack would reconstruct the noisy data, rather than the original confidential data. (The Census Bureau’s reconstruction attack had similarly reconstructed the swapped 2010 data, implying that the individuals successfully reconstructed and reidentified had not been swapped in the 2010 data.)
- The amount of noise added is tunable: the more noise, the less privacy loss. In the case of “pure” DP, the noise is measured on a scale labeled by ϵ (epsilon) that varies from 0 to ∞ . When $\epsilon = 0$, an infinite amount of noise is added, and the published data have no relationship to the underlying confidential information. When $\epsilon = \infty$, no

noise is added, and the published data have no privacy protection (or “infinite” privacy loss). (See [Figure 2](#))

- The privacy mechanism doesn’t specify how much noise should be used. For the 2010 Census, the details of the privacy protection mechanism were seen as technical issues and resolved by statisticians working within the Census Bureau’s Center for Disclosure Avoidance Research (CDAR). For the 2020 Census, the decision of how much noise to add was seen as a public policy issue and decided by the Census Bureau’s Data Stewardship Executive Policy Committee, chaired by the Census Bureau’s Deputy Director.

Figure 2

Figure 2. Differential Privacy algorithms incorporate a tunable parameter ϵ , which controls the balance between data set accuracy and privacy loss: $\epsilon = 0$ corresponds to the addition of infinite noise to the raw data, which results in a publication that causes no privacy loss but also contains no useful information about the confidential data, whereas $\epsilon = \infty$ corresponds to the addition of no noise, which yields no loss of data accuracy but incorporates no privacy protection within the resulting data set. (Source: US Census Bureau.)

Nevertheless, differential privacy had several characteristics that made it and its application fundamentally different from what the Census Bureau had done in the past:

- DP is based on a mathematical definition of “privacy loss.” DP provides an *upper bound* on the maximum *increase* in privacy loss that can result from the data release.

The previous techniques were not based on mathematically provable, rigorous definitions of privacy or privacy loss.

- The definition of privacy loss is based on mathematically distinguishing between 1) what a data intruder can learn using a data release that includes an arbitrary person's data, and 2) what a data intruder can learn using a data set that does not include that same arbitrary person's data. This is where the name "differential privacy" comes from: it is the difference between the two data sets.
- Every person in the census would have their confidential data protected by differential privacy. In contrast, the previous protection mechanism for the 2010 census was only applied to individuals that Census Bureau statisticians thought were at risk of re-identification.
- In the past, the Census Bureau had not released the specifics of the disclosure avoidance mechanism. With DP, the Census Bureau would release every detail regarding the disclosure avoidance process, including the source code of the software and the values of all privacy protection parameters.²⁵ The only information that would not be released would be the actual noise values, produced by billions of random numbers, that would then be added to the confidential data.²⁶

Protecting the 2020 Census

Dr. John Abowd became the Census Bureau's Associate Director for Research and Methodology and Chief Scientist on June 1, 2016. A distinguished professor of economics at Cornell University, Abowd had come to the Census Bureau to modernize its privacy protection systems.²⁷

The decennial census may be the Census Bureau's reason for existing, but it is just one of more than a hundred data products that the agency produces, and Abowd wanted to modernize them all. In addition to the decennial census, he initiated projects to develop "formal privacy" protection approaches for the 5-year Economic Census, the annual American Community Survey, and the use of confidential data by academic researchers in the FSRDCs.²⁸

Abowd's goals were extraordinarily ambitious. Planning for the 2020 Census had been underway since December 11, 2011. When Abowd started work in 2016, the critical systems for the 2020 Census were supposed to be already designed and nearing completion in preparation for the bureau's 2018 "End-to-End" test.²⁹ The confidentiality protection system would have to be written from scratch, and it was one of the most critical systems, a metaphorical bottleneck through which all of the 2020 Census data would flow.

Abowd was confident that a new system could be created, tested, and deployed in time, in part because he had led the team in 2008 that had deployed differential privacy to the Census Bureau's "OnTheMap" data product.³⁰ But OnTheMap was designed from the beginning to use noise infusion for confidentiality protection: retrofitting it to use differential privacy had not been very difficult. In contrast, the 2020 Census was steeped in decades of practice, and it thus had many procedures that would require updating to make it formally private.

The first challenge was the lack of differential privacy experts within the Census Bureau itself. The Census Bureau addressed this challenge by signing Cooperative Agreements with universities that had differential privacy experts, issuing contracts to small businesses that had expertise in the area, engaging a prestigious thinktank to review its program, and initiating a training program for its existing staff.³¹ Within a year, three of differential privacy's four inventors (Dwork, Nissim, and Smith) were connected in some way with the project.

Somewhat to the surprise of those involved, the team was able to implement a sophisticated DP algorithm in time for the 2018 End-to-End test. Called "TopDown," the algorithm computed statistics, applied noise, and then recomputed statistics at each geographic level of interest, from the entire nation, to each state, each county, each census tract, and ultimately each block ([Figure 3](#)). The algorithm, combined with the software to initialize the Amazon cloud-computing cluster and perform the necessary data transformations, is called the Disclosure Avoidance System (DAS).

Data Protection Process

Figure 3

Creating differentially private data for the 2020 Census redistricting files. For the 2020 US Census, the Census Bureau incorporated differential privacy at each geographic level of interest, from individual census blocks up to the national population as a whole. (Source: US Census Bureau.)

The Census Bureau performed these computations on clusters at Amazon Web Services, running on clusters typically with between ten and twenty-four computers,

each with 96 cores and 768GiB of RAM. The cost for each run was typically between \$100 and \$10,000, depending on which part of the algorithm was being tested and how much test data was being used. Over the following years, several thousand runs were performed, with a total cost of computer time running into millions of dollars. Nearly all of these runs were performed using the confidential data from the 2010 Census.

Given the growing interest among data users regarding the adoption of differential privacy, the Census Bureau committed to making multiple public releases of both protected data and code so that data users could see for themselves the impact of the new technology. These releases, based on reprocessing confidential data from the 2010 Census, were called the 2010 Demonstration Data Products. They were designed to let data users explore how the DAS was improving as the Census Bureau's development team raced to the publication deadline. The Census Bureau hoped that data users would download each release and provide constructive feedback as to where the algorithm needed to be improved. In total, six data releases were made between October 2019 and June 2021 (see the appendix).^{[32](#)}

Reaction of Data Users

"How Will Statistical Agencies Operate When All Data Are Private?" is the rhetorical question that Abowd asked in 2017 when he accepted a prestigious award. "Data collected by statistical agencies must be published in some form... Publications are the public good that justifies the expense of taxpayer revenue on the agency. And the quality of those publications, measured in terms of their usefulness to the society that financed them, is the social benefit from that public good."^{[33](#)}

At the same time, he noted, for decades that public good had been balanced with the need to protect privacy. Although the Census Bureau had previously told data users that some error had been added for confidentiality protection, it had never quantified the amount. Without knowing how much error had been added, many data users acted as if the data were perfectly accurate. Now for the first time, data users would be told how much error was being intentionally added in the interest of protecting confidentiality! What's more, a lot more error would be added to the 2020 data than had been added in the past, a result of what had been learned from the reconstruction and reidentification experiment with the 2010 Census data.

There is a strict trade-off between data quality and confidentiality protection: increasing one decreases the other. With the Census Bureau finally providing

transparency regarding the trade-off, officials expected that there would be a desire to increase accuracy at the expense of confidentiality.

In fact, many data users did not seem interested in balancing accuracy with privacy protection at all, and the fact that DP allowed such a balance to be made was a problem, not a feature. Steve Ruggles, director of the Minnesota Population Center (MPC), became a primary critic. Ruggles's objections to the use of differential privacy started before the early demonstration data sets from the DAS were publicly released, and grew during the summer of 2018 when the Census Bureau released data from the 1940 Census that had been processed with a "beta" version of the DAS that had not been optimized for statistical accuracy.

Two of MPC's researchers, David Van Riper and Tracy Kugler, had compared a data set of 1940 data processed using the TopDown algorithm with the original 1940 data and discovered that DP added more percentage error to small populations than large ones. This was in fact the expected result, as DP uses additive noise. So as expected, the state's small Hispanic population was far more impacted by DP than its significantly larger non-Hispanic population. This would present a problem for researchers trying to study small, marginalized populations. Of course, DP was also protecting those same populations.

"It is not clear how this new disclosure avoidance system will be implemented, but there is significant danger that it will reduce or even eliminate the usability of public use data for many common research and policy applications," Ruggles wrote in an email that he sent to thousands of data users in October 2018. Before the end of the year, more than four thousand researchers signed a petition expressing concern regarding the use of DP in the 2020 Census and the intended use in the American Community Survey. "Adoption of differential privacy will have far-reaching consequences for research. It is possible—even likely—that scientists, planners, and the public will lose the free access we have enjoyed for six decades to reliable public Census Bureau data describing American social and economic change," Ruggles and 13 other academics wrote in a report they published later that year. "We believe that the differential confidentiality approach is inconsistent with the statutory obligations, history, and core mission of the Census Bureau."³⁴

"The numbers might be off by five, 10, 20 people, and if you're dealing with exact percentages... [t]hat could mean a lot," Jeffrey M. Wice, a national redistricting attorney, told the Associated Press. Census data does not need to be exact for most purposes, "as long as we know it's really pretty close," said Justin Levitt, an election

law professor at Loyola Law School in Los Angeles. But “there's certainly a point where blurry becomes too blurry.”³⁵

Following the first Demonstration Data Products release, the Committee on National Statistics (CNSTAT) of the National Academies of Sciences, Engineering and Medicine held a two-day workshop at which analyses of the data were discussed. Census Bureau presentations emphasized that much of the error came not from the DP but from the “postprocessing” that turned all of the noise-infused counts into nonnegative, integer values.³⁶ As a result, Census Bureau presenters noted that it should be possible to improve the accuracy of the DAS without compromising privacy. And, of course, the final decision on the “privacy loss budget”—the tradeoff between privacy and accuracy—had yet to be made.

Privacy certainty took a backseat to accuracy at the CNSTAT workshop. While the workshop’s organizers devoted fifteen hours to discuss the impact of DP on data accuracy, just ninety minutes were scheduled for discussion of privacy issues. And whereas the data sessions featured quantitative analyses with graphs and charts and measures of error, the privacy discussion lacked any quantification. This was ironic, given that the existential justification for differential privacy was that its formal approach allowed the Census Bureau to quantify privacy loss. Where was the quantification? “The fundamental tension between privacy and accuracy runs through all the workshop sessions, but the workshop by design emphasized data users and their needs, which tend overwhelmingly to focus on the data accuracy side of the proposition,” the workshop’s organizers acknowledged.³⁷

Almost entirely absent from the CNSTAT meeting was the typical collection of civil society organizations and academic researchers who actively research and write about the intersection of technology, privacy, and public policy. Those organizations tend to be staffed by lawyers, not data scientists, and they often lack in-house experts who can download gigabyte-sized data sets and perform privacy analyses. These organizations typically respond to notices in the *Federal Register* by regulatory agencies, or file suit in federal court. But whereas the Census Bureau had filed a Federal Register Notice asking data users for their accuracy requirements, it never published details on its plans to implement DP or how it would set the privacy loss budget.³⁸

Many civil liberties organizations concerned about privacy and technology seemed generally pleased that the Census Bureau was using DP, the so-called “gold standard” of privacy protection, and they did not focus on the specific question of the privacy loss budget or the trade-off between accuracy and privacy. These organizations were also

focused on the efforts of the Trump administration to ask respondents if they were citizens of the United States. When that effort failed, the administration directed the Census Bureau to infer which respondents were noncitizens using administrative records such as driver's license data, and then to produce a set of tables that had data from noncitizens removed. These efforts consumed considerable attention from privacy and civil liberties advocates, who chose to fight the administration's efforts in court, rather than focusing on the ways that DP might be used to protect such information once it was collected.^{[39](#)}

Unexpected Complications

Less than a month after the CNSTAT workshop, reports started circulating of a mysterious pneumonia-like disease in the Chinese city of Wuhan. On March 11, 2020, the World Health Organization declared the COVID-19 pandemic.

On schedule, the Census Bureau sent out "invitations" in early March, directing residents of the United States to fill out the online census collection instrument. But the pandemic impacted virtually every aspect of human life in the United States, and the 2020 Census was no different. The Census Bureau moved to 100 percent telework in March and "paused" field operations, as the pandemic-related lockdowns would likely interfere with door-to-door enumeration efforts. The 2020 Internet Collection Instrument was left running until October 15, 2020 (it was originally scheduled to be decommissioned on July 31, 2020).^{[40](#)} This meant that the final apportionment counts, which were originally scheduled to be delivered to the president by December 31, 2020, would not be delivered until April 26, 2021, and the release of the redistricting data was pushed from March 31, 2021 to August 12, 2021. The extra time gave the Census Bureau's statisticians more opportunity to improve the DAS.

The design of the TopDown algorithm had been guided by the idea that statistics were computed using a so-called "Geographic Spine" that specified each census block as being contained within a specific census tract, county, and state, and as a result, statistics based on these geographies were significantly more accurate than computations based on a collection of blocks or tracts that were not cleanly contained within a specific county. However, it turns out that many communities in the United States are defined by towns, other civil divisions, and tribal lands—rather than counties. The DAS team developed an improved geographic approach and in June 2020 the chief scientist and the associate director for demographic programs jointly announced that the DAS team had developed a solution that it was now working to deploy.^{[41](#)}

In March 2021, the state of Alabama filed suit against the Census Bureau, stating that the redistricting data should be released at the end of March as originally planned, rather than in August—and that the Census Bureau should be prohibited from using differential privacy.⁴² Sixteen other states joined the lawsuit.⁴³ The panel of three federal judges hearing the case denied both requests, and the lawsuit was dropped on September 9, 2021.⁴⁴

In April 2021, the Census Bureau made available its *fifth* demonstration data set. Unlike previous data sets, this one focused solely on data that would be produced by the 2020 Census for redistricting. The redistricting data is block-level data that includes sixty-three different race categories and two ethnicities, but only provides counts of individuals who are eighteen and over or under age eighteen. It also includes a separate housing file indicating the number of occupied and unoccupied units on each block. The Census Bureau announced that the DP algorithm would create inconsistencies between these two files, such as blocks containing people but having no occupied housing units, or having occupied housing units but zero household population. Nevertheless, the joint statement issued by the Census Bureau’s chief scientist and its associate director for demographic programs stated that “the data meet the accuracy criteria we developed after extensive discussions with the redistricting community and the Civil Rights Division at the U.S. Department of Justice.”⁴⁵

The Redistricting File Decision and Release

In June 2021, the Census Bureau announced that it had set the final parameters for the 2020 Redistricting File, and that it had further modified the DAS to make it more accurate than it had previously been. The privacy loss budget was set to $\epsilon=17.14$ for statistics based on people (the “persons file”) and $\epsilon = 2.47$ for the statistics based on housing units (the “housing unit file”).⁴⁶ These values were far higher than the values that had been used in any of the demonstration data products, surprising many census watchers.

The DAS geographic processing was changed to incorporate an “optimized spine” that was designed to improve the accuracy for cities, boroughs, and towns that did not neatly match up with state counties, and that separately computed statistics for tribal areas.⁴⁷ The postprocessing system also implemented a “multipass” postprocessing system, in which it computed the total population of every geographical area before computing the racial mix. Finally, the underlying differential privacy mechanism was changed from “pure DP,” which used the Laplace or Geometric distribution, to so-

called “Zero-Concentrated Differential Privacy (zCDP)” which uses a Gaussian distribution. The Gaussian distribution is broader than the Laplace distribution, but its “tails” are much smaller, so there would be fewer outliers created by DP itself ([Figure 4](#)).

Zero-Concentrated Differential Privacy (zCDP)

Figure 4

For the 2020 US Census, the Census Bureau adopted “zero-concentrated differential privacy” (zCDP), in which the random noise that is inserted to protect privacy of the raw data is drawn from a Gaussian distribution rather than a Laplace (or Geometric) distribution. Drawing random noise from a Gaussian distribution introduces fewer statistical outliers. (Source: US Census Bureau.)

The 2020 redistricting file was uploaded to the Census Bureau’s webserver on August 12, 2021.^{[48](#)} Two months later, the Census Bureau published a 37-page guide explaining in detail and all in one place for the first time the privacy risks caused by the decennial census, the details of the Census Bureau’s database reconstruction experiment, the reasons why the Census Bureau had adopted differential privacy, the concept of the accuracy/privacy loss tradeoff, what it means to add noise to a hypothetical census block, and how the TopDown algorithm works.^{[49](#)} The guide stated that whereas between 1 and 5 percent of blocks had some kind of inconsistent or implausible result as a result of DP, less than 0.1 percent of block groups were affected, and no counties were impacted. The guide provided many details of the

allocation of the privacy loss budget as well, but it did not provide any details regarding how the setting for the privacy loss budget had been decided.

Most of the press coverage of the first major release of data for the 2020 Census made little mention of DP—probably a good outcome for the Census Bureau, which instead celebrated the release of the redistricting numbers and did not dwell on the controversies involved in their production. The DAS team presumably continued working: it still had to release the demographic data—the tables with information about sex and age—as well as the tables with detailed race and tribal affiliations, which would be produced with an entirely different DP engine. For each of those releases, an additional privacy loss budget would need to be established and allocated.

Differential privacy was developed to protect a national census, and the 2020 Census was the first attempt anywhere to use the technology on such a scale. Although nearly all of the deep technical problems that the DAS team encountered were in the postprocessing part of the DAS, the postprocessing was only necessary because of the decision to use DP in the first place.

The decision to use DP for such a high-stakes project, and without a backup plan, almost certainly brought funding, talent, and attention to the project that it otherwise could never have obtained. The decision also helped the Census Bureau reestablish its reputation as a leader in data science, privacy protection, and advanced computing. But data users would probably have been more supportive of the project if the Census Bureau had established clear metrics for statistical accuracy that it was committed to meeting. Earlier engagement with users also might have led to an earlier identification and resolution of the scientific issues relating to geographic processing and the need for total count accuracy.

Discussion Questions and Activities

1. What are the threats that justify some protection for privacy or confidentiality for enumerated persons?
2. The Census Bureau's publications do not contain names, so assigning names to data that the Census Bureau releases requires consulting additional data sources. The Census Bureau clearly thinks that it has an ethical obligation to protect against such attacks, even though much of this data is already publicly available. Do you agree?
3. The history of census-taking also informs opinions about data confidentiality. In Germany and in other nations, census activities have triggered widespread protests because of the fear that the data could be used to purge immigrants or others. The

Nazi regime used the results of the 1939 German census and other Western European censuses to assist in the internment and murder of millions of Jews. Likewise, data from the US 1940 census were used to assist in the internment of approximately 120,000 Japanese Americans in the western United States between 1942 and 1945. Are such concerns relevant today?

4. None of the abuses of census data described above would be prevented by differential privacy, since they all involve abuses by the census-taking authority, rather than by the people or agencies that made use of the census publications. Likewise, many of the civil society groups that were concerned about the politicization of the 2020 Census by the Trump administration were concerned about the proposed inclusion of a citizenship question, and then by proposed use of administrative data to identify “unauthorized immigrants.” Given these issues, do you think that the attention focused on differential privacy was warranted?
5. Many of the loudest opponents of the DP approach were academic researchers. What responsibilities do researchers have toward the people whose data they analyze, and how are these responsibilities enforced? Could the risk to US persons’ confidentiality be mitigated by imposing security and other obligations on data users, such as a promise to not reidentify data? What about legal prohibitions against reidentification?[50](#)
6. Many people critical of the Census Bureau’s use of differential privacy believed that it was unnecessary because the information being protected—a person’s address, age, sex, race, ethnicity and relationship status—is widely available from other sources. Others have argued that this information is a result of government compulsion, is often expensive in other contexts, thus not available for everybody, and that it is especially not widely available for children, unauthorized immigrants, and various hard-to-count populations. How should the Census Bureau decide the appropriate measures and mechanisms that it should follow to protect respondent confidentiality?
7. Differential privacy represents for many people a new way of thinking about privacy issues. Traditionally, privacy issues have been viewed through the framework of Fair Information Practice Principles (FIPS) developed between the 1960s and the 1990s. These are a series of mostly procedural protections for data systems. They include: consideration of the purpose for which data are collected, the uses that will be applied to the data, whether the data are held securely, what user choice is available in the system, and whether the data are accurate and up to date. Another approach is Helen Nissenbaum’s “contextual integrity,” which considers confidentiality violations that occur when information norms are not respected.[51](#) Compare and

contrast the confidentiality goals of each of these mechanisms. What are the advantages and disadvantages of each? Would the use of the FIPS or contextual integrity help resolve the conflict between confidentiality and data uses? How would the Census Bureau enforce such an approach?

8. Differential privacy was invented in 2006, but the Census Bureau didn't start factoring the technique into its confidentiality protection system for the 2020 Census until 2016. As a result, the algorithm and implementation software were still under development months after the 2020 census data had been collected. Given the compressed time tables and the substantial pushback from academic users and states, should the Census Bureau have spent an extra decade developing and socializing its differential privacy technique, and not deployed differential privacy until the 2030 census?
9. The confidentiality provisions of the Census Act (Title 13, Section 9) have not been updated in decades. How would you update them to take into account advanced mathematical techniques like database reconstruction and the existence of vast amounts of personal information online?
10. What else could the Census Bureau have done to ease the transition to differential privacy?
11. What would you need to know in order to decide if $\epsilon=17.14$ is an appropriate value for the release of the redistricting data?
12. If a court had ordered the Census Bureau not to use differential privacy to protect the 2020 Census, what other options were realistically available?
13. How would you change the law to recognize the tension between the need to produce accurate statistics and the need to protect privacy?

Acknowledgments

The author is grateful to Claire Bowen, danah boyd, Daniel Goroff, Chris Hoofnagle, Eddie Hunsinger, Philip Leclerc, David Van Riper, and the MIT reviewers for their comments on previous versions of this case study.

Bibliography

Abowd, John M. "The U.S. Census Bureau Tries to Be a Good Data Steward in the 21st Century, 9th Annual FDIC Consumer Research Symposium." Distinguished Guest Lecture, Arlington, VA, October 18, 2019.

<https://www.fdic.gov/news/events/consumersymposium/2019/documents/presentations/abowd-presentation.pdf>.

Abowd, John. "Balancing Privacy and Accuracy: New Opportunity for Disclosure Avoidance Analysis." *Research Matters* (blog). US Census Bureau, October 29, 2019. https://www.census.gov/newsroom/blogs/research-matters/2019/10/balancing_privacyan.html.

Abowd, John. "Census Bureau Awards Cooperative Agreements to Georgetown University and Purdue University." US Census Bureau, October 18, 2016. https://www.census.gov/newsroom/blogs/research-matters/2016/10/census_bureau_awards.html.

Abowd, John. Declaration of John M. Abowd, State of Alabama, et al., v. United States Department of Commerce, et. al, Case No. 3:21-CV-211-RAH-ECM-KCN. <https://censusproject.files.wordpress.com/2021/04/2021.04.13-abowd-declaration-alabama-v-commerce-ii-final-signed.pdf>.

Abowd, John M. "How Will Statistical Agencies Operate When All Data Are Private?" *Journal of Privacy and Confidentiality* 7, no. 3 (2016). <https://doi.org/10.29012/jpc.v7i3.404>.

Abowd, John. "Modernizing Disclosure Avoidance; What We've Learned, Where We Are Now." *Research Matters* (blog). US Census Bureau, March 13, 2020. https://www.census.gov/newsroom/blogs/research-matters/2020/03/modernizing_disclosu.html.

Abowd, John. "Staring Down the Database Reconstruction Theorem, American Association for the Advancement of Science Annual Meeting." Cornell University (blog), February 16, 2019. <https://blogs.cornell.edu/abowd/special-materials/>.

Abowd, John M. "The U.S. Census Bureau Adopts Differential Privacy." In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). Video, 34 min. Association for Computing Machinery, New York, NY. <https://doi.org/10.1145/3219819.3226070>.

Abowd, John, and Victoria A. Velkoff. "Modernizing Disclosure Avoidance: A Multipass Solution to Post-processing Error." *Research Matters* (blog). US Census Bureau, June 18, 2020. https://www.census.gov/newsroom/blogs/research-matters/2020/06/modernizing_disclosu.html.

Abowd, John, and Victoria A. Velkoff. "Modernizing Privacy Protections for the 2020 Census: Next Steps." *Research Matters* (blog). US Census Bureau, April 28, 2021.

https://www.census.gov/newsroom/blogs/random-samplings/2021/04/modernizing_privacy.html.

Abowd, John M., Gary L. Benedetto, Simson L. Garfinkel et al., “The Modernization of Statistical Disclosure Limitation at the US Census Bureau.” Census Working Paper, August 2020, Washington, DC. <https://www.census.gov/library/working-papers/2020/adrm/modernization-statistical-disclosure-limitation.html>.

Catt, Mary. “ILR’s John Abowd to Lead US Census Bureau Research.” *Cornell Chronicle*, December 10, 2015. <https://news.cornell.edu/stories/2015/12/ilrs-john-abowd-lead-us-census-bureau-research>.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. “Calibrating Noise to Sensitivity in Private Data Analysis.” In *Proceedings of the Third Conference on Theory of Cryptography*, ed. S. Halevi and T. Rabin, 265–284. Berlin, Heidelberg: Springer-Verlag, 2006. http://dx.doi.org/10.1007/11681878_14.

Federal Committee on Statistical Methodology. “Statistical Policy Working Paper 22 (Second version),” 2005. Report on Statistical Disclosure Limitation Methodology. <https://www.hhs.gov/sites/default/files/spwp22.pdf>.

Harvard Data Science Review. “Inaugural Symposium, Differential Privacy for 2020 US Census, Session 2,” 2019. Slides, 138. <https://assets.pubpub.org/j2yr11kl/11587735061843.pdf>.

Ho, Dale. “The Trump Administration’s Census Cover Up.” American Civil Liberties Union, July 27, 2018 (updated March 21, 2019). <https://www.aclu.org/blog/voting-rights/trump-administrations-census-cover>.

Garfinkel, Simson. “Walker and the ‘Indian Question.’” *Technology Review*, August 24, 2021. <https://www.technologyreview.com/2021/08/24/1030428/walker-and-the-indian-question/>.

Garfinkel, Simson L. “Deploying Differential Privacy for the 2020 Census of Population and Housing.” Keynote presentation, Privacy Enhancing Technologies Symposium Stockholm, Sweden, July 16, 2019. <https://simson.net/ref/2019/2019-07-16%20Deploying%20Differential%20Privacy%20for%20the%202020%20Census.pdf>.

Garfinkel, Simson L., and Philip Leclerc. Randomness Concerns When Deploying Differential Privacy. In *Proceedings of the 19th Workshop on Privacy in the Electronic*

Society (WPES'20), 73–86. New York: ACM, 2020.

<https://doi.org/10.1145/3411497.3420211>.

Garfinkel, Simson, John M. Abowd, and Christian Martindale. “Understanding Database Reconstruction Attacks on Public Data.” *Communications of the ACM* 62, no. 3 (March 2019): 46–53. <https://doi.org/10.1145/3287287>.

Garfinkel, Simson, John Abowd, and Sarah Powazek. “Issues Encountered Deploying Differential Privacy.” In *Proceedings of the Workshop on Privacy in the Electronic Society*, 133–137. New York: ACM, 2018.

<https://dl.acm.org/doi/10.1145/3267323.3268949>.

Gatewood, George. *A Monograph on Confidentiality and Privacy in the U.S. Census*. Washington, DC: US Census Bureau, July 2001.

<https://www.census.gov/history/pdf/ConfidentialityMonograph.pdf>.

Hawes, Michael, and Michael Ratcliffe. “Understanding the 2020 Census Disclosure Avoidance System: Differential Privacy 201 and the TopDown Algorithm.” Workshop Presentation, US Census Bureau, May 13, 2021.

<https://www2.census.gov/about/training-workshops/2021/2021-05-13-das-presentation.pdf>.

JASON. *Formal Privacy Methods for the 2020 Census*. The MITRE Corporation, JSR-19-2F, April 2020. <https://irp.fas.org/agency/dod/jason/census-privacy.pdf>.

Machanavajjhala, Ashwin, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. “Privacy: Theory Meets Practice on the Map.” In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE '08)*, 277–286. New York: IEEE Computer Society, 2008. <https://doi.org/10.1109/ICDE.2008.4497436>.

Marks, Rachel, and Nicholas Jones. “Collecting and Tabulating Ethnicity and Race Respondents in the 2020 Census.” Workshop Presentation, US Census Bureau, February 2020. <https://www2.census.gov/about/training-workshops/2020/2020-02-19-pop-presentation.pdf>.

McDermott, Jennifer, and Mike Schneider. “Researchers Question Census Bureau’s New Approach to Privacy.” Associated Press, September 28, 2019. <https://apnews.com/article/ri-state-wire-census-2020-us-news-ap-top-news-redistricting-88307905204e4d5aaad10db486514ccb>.

McKenna, Laura. "Disclosure Avoidance Techniques Used for the 1960 Through 2010 Census." Working Paper. Census Technical Report ADRM-2020-007, US Census Bureau, Washington, DC, 2019. <https://www.census.gov/library/working-papers/2019/adrm/six-decennial-censuses-da.html>.

McKenna, Laura. "U.S. Census Bureau Reidentification Studies." Research and Methodology Directorate, US Census Bureau, Washington, DC, April 2019. <https://www2.census.gov/adrm/CED/Papers/CY19/2019-04-Reidentification%20studies-20210331FinRed.pdf>.

National Academies of Sciences, Engineering, and Medicine. *2020 Census Data Products: Data Needs and Privacy Considerations: Proceedings of a Workshop*. Washington, DC: The National Academies Press, 2020. <https://doi.org/10.17226/25978>. Video: <https://vimeo.com/showcase/6661772>.

Nissenbaum, Helen. "Privacy as Contextual Integrity." *Washington Law Review* 79, no. 1 (2004): 119–157. <https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/>.

Phillips, Mark, Edward S. Dove, and Bartha M. Knoppers. "Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?" *Bioethical Inquiry* 14 (2017): 527–539. <https://doi.org/10.1007/s11673-017-9806-9>.

Reamer, Andrew. "Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds." GW Institute of Public Policy, Washington, DC, April 29, 2020. <https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds>.

Reich, Henry. "Protecting Privacy with MATH (Collab with the Census)." Minute Physics, Video, September 12, 2019. <https://youtu.be/pT19VwBAqKA>.

Riccardi, Nicholas, and Mike Schneider. "Alabama Sues to Stop Redistricting Delay, Privacy Initiative." AP News, March 10, 2021. <https://apnews.com/article/alabama-redistricting-lawsuits-census-2020-courts-9627c66d67629942d2f961bf781f4309>. For the lawsuit, see *State of Alabama v. US Department of Commerce*, Civil Action 2:18-cv-00772-RDP, US District Court for the Northern District of Alabama, May 21, 2021.

Ruggles, Steven, "Comparability of the Public Use Files of the U.S. Census of Population, 1880-1980." *Social Science History* 15, no. 1 (Spring 1991). <https://users.pop.umn.edu/~ruggles/Articles/SSH1991.pdf>.

Ruggles, Steven, Catherine Fitch, Diana Magnuson, and Jonathan Schroeder. "Differential Privacy and Census Data: Implications for Social and Economic Research." *AEA Papers and Proceedings* 109 (2019): 403–408.
<https://doi.org/10.1257/pandp.20191107>.

Ruggles, Steven, et al. "Implications of Differential Privacy for Census Bureau Data and Scientific Research." Minnesota Population Center, University of Minnesota, Working Paper No. 2018-6, Minneapolis, MN, December 2018, Version 5.1.
<https://doi.org/10.18128/MPC2018-6>.

Schneider, Mike. "16 States Back Alabama's Challenge to Census Privacy Tool." AP News, April 13, 2021. <https://apnews.com/article/race-and-ethnicity-alabama-florida-census-2020-us-news-3407a2d05be23972741294bc05838801>.

Schneider, Mike. "Census on Track for August Data Release After Court Ruling." AP News, June 30, 2021. <https://apnews.com/article/census-2020-courts-government-and-politics-bc46bd098e0d2234cd456d697f14a658>.

Scott, Tony. "The People's Code." CIO Council (blog entry), August 11, 2016.
<https://www.cio.gov/2016/08/11/peoples-code.html>.

Scott, Tony. "Federal Source Code Policy: Achieving Efficiency, Transparency, and Innovation through Reusable and Open Source Software." Executive Office of the President, Office of Management and Budget, Memorandum No. M-16-21, Washington, DC, August 8, 2016.
https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2016/m_16_21.pdf.

Seltzer, William. "Excluding Indians Not Taxed: Federal Census and Native-Americans in the 19th Century." Paper prepared for presentation at the 1999 Join Statistical Meetings, Baltimore, MD, August 8–12, 1999.
<https://margoanderson.org/govstat/ind.pdf>.

Tennant, Brad. "'Excluding Indians Not Taxed:' *Dred Scott*, *Standing Bear*, *Elk* and the Legal Status of the Native Americans in the Latter Half of the Nineteenth Century." *International Social Science Review* 86, Nos. 1/2 (2011): 24–43.
<https://www.jstor.org/stable/41887472>.

US Census Bureau. *2020 Census Operational Plan: A New Design for the 21st Century*. Version 1.1, November 2015. <https://www2.census.gov/programs->

[surveys/decennial/2020/program-management/planning-docs/2020-oper-plan.pdf](https://www.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan.pdf).

US Census Bureau. "Soliciting Feedback from Users on 2020 Census Data Products," Notice, 83 Fed. Reg. 34111, July 19, 2018.

<https://www.federalregister.gov/documents/2018/07/19/2018-15458/soliciting-feedback-from-users-on-2020-census-data-products>.

US Census Bureau. "New Briefing to Announce First Major Field Operations for the 2020 Census." Archived Video: August 12, 2019.

<https://www.census.gov/newsroom/press-kits/2018/2020-adcan.html>.

US Census Bureau Fact Sheet. "How the 2020 Census Counts People Experiencing Homelessness." US Census Bureau Fact Sheet, December 2020.

<https://www.census.gov/content/dam/Census/library/factsheets/2020/dec/census-counts-homeless.pdf>.

US Census Bureau. "2020 Census Apportionment Results." April 26, 2021.

<https://www.census.gov/data/tables/2020/dec/2020-apportionment-data.html>.

US Census Bureau. "Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results." June 9, 2021. <https://www.census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html>.

US Census Bureau. "2020 Census Redistricting Data Files Press Kit." August 12, 2021.

<https://www.census.gov/newsroom/press-kits/2021/2020-census-redistricting.html>.

US Census Bureau. "2020 Census Operational Adjustments Due to COVID-19." Last updated October 8, 2021. <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/operational-adjustments.html>.

US Census Bureau. "Disclosure Avoidance for the 2020 Census: An Introduction."

November 2, 2021. <https://www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html>.

US Census Bureau. "About the Topic of Race." N.d.

<https://www.census.gov/topics/population/race/about.html>.

US Census Bureau. "1960 (Population)." N.d.

https://www.census.gov/history/www/through_the_decades/index_of_questions/1960_population.html.

US Census Bureau. “Disclosure Avoidance Blogs.” N.d.

<https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/blogs.html>.

Zayatz, Laura, Jason Lucero, Paul Massell, and Asoka Ramanayake. “Disclosure Avoidance for Census 2010 and American Community Survey Five-Year Tabular Data Products.” Working Paper, Statistical Research Division, U.S. Census Bureau, Washington, DC, 2009. <https://www.census.gov/library/working-papers/2009/adrm/rrs2009-10.html>.

Appendix:

The 2010 Data Demonstration Materials

Currently, the 2010 Data Demonstration Products (DDP) can be downloaded from the Census Bureau’s website at: <https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/>.

The initial Data Demonstration Product included just tables, but the following releases also included the Privacy Protected Microdata File (PPMF) from which the tables were created.

The data releases consisted of:^{[52](#)}

- (1) 2010 Original Demonstration Data Product Version 2019-10-29
- (2) 2010 Summary Metrics and Demonstration PPMF Version 2020-05-27
- (3) 2010 Summary Metrics and Demonstration PPMF Version 2020-09-17
- (4) 2010 Summary Metrics and Demonstration PPMF Version 2020-11-16
- (5) 2010 Summary Metrics and Demonstration PPMF Version 2021-04-28
- (5) 2010 Summary Metrics, Demonstration PPMF Version 2021-06-08, and production settings for the 2020 redistricting file.

2019-10-29

The 2010 Census Edited File processed by the DAS. The person-level tables received a privacy-loss budget of $\epsilon = 4$, and the housing unit-level tables received a privacy-loss budget of $\epsilon = 0.5$

- [01-Redistricting_File--PL_94-171/](#) - The privacy-protected microdata file processed by the DAS and tabulated using the same schema as the Public Law 94 Redistricting File release of the 2010 Census.
- [02-Demographic_and_Housing_Characteristics/](#) - The privacy-protected microdata file processed by the DAS and tabulated using a schema similar to the Summary File 1 release of the 2010 Census.

2020-09-17

The DAS run that generated the PPMFs 2020-09-17 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 $\epsilon = 0.5$.

2020-11-16

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 $\epsilon = 0.5$.

2021-04-28

This release included two sets of tables with two privacy loss budgets.

Set #1: Global PLB of $\epsilon = 12.2$ (tuned to accuracy targets)

- Detailed Summary Metrics
- Person-level data ($\epsilon = 10.3$)
- Unit-level data ($\epsilon = 1.9$)

Set #2: Global PLB of $\epsilon = 4.5$ (for algorithmic comparison to prior demonstration data releases)

- Detailed Summary Metrics
- Person-level data ($\epsilon = 4.0$)
- Unit-level data ($\epsilon = 0.5$)

2021-06-08 and 2021-08-12

Included in this release are Detailed Summary Metrics (DSM) and Privacy-Protected Microdata Files (PPMFs) at the chosen Privacy-loss Budget of 17.14 for person files

and 2.47 for housing units.

Press Play ► Simson Garfinkel joined the Data Skeptic podcast to discuss his work modernizing the Census Bureau disclosure avoidance system from private to public disclosure avoidance techniques using differential privacy.

Visit the web version of this article to view interactive content.

Footnotes

1. Laura McKenna, “Disclosure Avoidance Techniques Used for the 1960 Through 2010 Census,” Working Paper, Census Technical Report ADRM-2020-007, US Census Bureau, <https://www.census.gov/library/working-papers/2019/adrm/six-decennial-censuses-da.html>. ↵
2. <https://www.census.gov/programs-surveys/decennial-census.html>. ↵
3. US Department of Justice, “Section 2 of the Voting Rights Act,” available at <https://www.justice.gov/crt/section-2-voting-rights-act> ↵
4. Andrew Reamer, “Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds,” GW Institute of Public Policy, Washington, DC, April 29, 2020, <https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds>. ↵
5. George Washington, letter to Gouverneur Morris, 28 July 1791, available at <https://teachingamericanhistory.org/document/letter-to-gouverneur-morris/>. ↵
6. For a discussion of the phrase “excluding Indians not taxed,” see Brad Tennant, “‘Excluding Indians Not Taxed’: *Dred Scott*, *Standing Bear*, *Elk* and the Legal Status of the Native Americans in the Latter Half of the Nineteenth Century,” *International Social Science Review* 86 (2011): 24-43, <https://www.jstor.org/stable/41887472>; and William Seltzer, “Excluding Indians Not Taxed: Federal Census and Native-Americans in the 19th Century,” Paper presented at the 1999 Joint Statistical Meetings, Baltimore, MD, August 8-12, <https://margoanderson.org/govstat/ind.pdf>. ↵
7. George Gatewood, “A Monograph on Confidentiality and Privacy in the U.S. Census” (Washington, DC: U.S. Census Bureau, July 2001), <https://www.census.gov/history/pdf/ConfidentialityMonograph.pdf>. ↵

8. U.S. Census Bureau, “1880 Overview,” available at https://www.census.gov/history/www/through_the_decades/overview/1880.html. See also “1880 Instructions,” at https://www.census.gov/history/www/through_the_decades/census_instructions/1880instructions.html, and Forty-Fifth Congress , Session III, Chapter 195, 1879. p. 473. Archived at https://www.census.gov/history/pdf/1879_Census_Act.pdf ↵
9. Walker, who also served as Commissioner of Indian Affairs, left public service in 1881 to become MIT’s third president. See Simson Garfinkel, “Walker and the ‘Indian Question,’” *Technology Review*, August 24, 2021, <https://www.technologyreview.com/2021/08/24/1030428/walker-and-the-indian-question/>. ↵
10. Somewhat surprisingly, it is not the Census Bureau that publishes the “census and survey records of the Bureau of the Census containing identifying individuals enumerated in population censuses,” but the archivist, as detailed in Title 44 section 2108 of the US Code. <https://www.govinfo.gov/content/pkg/USCODE-2008-title44/html/USCODE-2008-title44.htm>. For the Census Bureau’s explanation of the “72-Year Rule,” see https://www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html. ↵
11. US Census Bureau Fact Sheet, “How the 2020 Census Counts People Experiencing Homelessness,” December 2020, <https://www.census.gov/content/dam/Census/library/factsheets/2020/dec/census-counts-homeless.pdf>. ↵
12. US Census Bureau, “Archived Video: August 12, 2019 – New Briefing to Announce First Major Field Operations for the 2020 Census,” <https://www.census.gov/newsroom/press-kits/2018/2020-adcan.html>. ↵
13. “The U.S. Census Bureau must adhere to the 1997 Office of Management and Budget (OMB) standards on race and ethnicity,” states the U.S. Census Bureau’s website. The official race categories for the 2020 Census were White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, and Some Other Race. See <https://www.census.gov/topics/population/race/about.html>. ↵

14. Ethnicity is defined by OMB as either “Hispanic or Latino” or “Not Hispanic or Latino.” According to a presentation for the 2020 Census, “OMB defines “Hispanic or Latino” as a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race. People who identify as Hispanic, Latino, or Spanish may be any race.” See Rachel Marks and Nicholas Jones, “Collecting and Tabulating Ethnicity and Race Respondents in the 2020 Census,” Workshop Presentation, February 2020, at <https://www2.census.gov/about/training-workshops/2020/2020-02-19-pop-presentation.pdf>. [↵](#)
15. US Census Bureau, “2020 Census Apportionment Results,” April 26, 2021, <https://www.census.gov/data/tables/2020/dec/2020-apportionment-data.html>. [↵](#)
16. US Census Bureau, “2020 Census Redistricting Data Files Press Kit,” August 12, 2021, <https://www.census.gov/newsroom/press-kits/2021/2020-census-redistricting.html>. Here the word “block” is a bit confusing, because it refers to a census block, from which all other Census Bureau geographies are built, and not to city blocks. In urban places such as Manhattan, a census block may in fact be a city block. In rural places, a single census block may be a mile or longer on a side. In the 2010 Census there were 11,151,873 blocks, of which 4,893,926 had zero population, and 1,832,637 had between 1 and 10 people. Just 31 blocks had more than 5,000 people, and 9,888 blocks had more than 1,000. [↵](#)
17. These and other legacy disclosure avoidance techniques are described in Federal Committee on Statistical Methodology, “Statistical Policy Working Paper 22 (Second version, 2005),” Report on Statistical Disclosure Limitation Methodology, <https://www.hhs.gov/sites/default/files/spwp22.pdf>. [↵](#)
18. For more information on the 1960 survey, see US Census Bureau, “1960 (Population),” https://www.census.gov/history/www/through_the_decades/index_of_questions/1960_population.html. [↵](#)
19. Laura McKenna, “Disclosure Avoidance Techniques”; Steven Ruggles, “Comparability of the Public Use Files of the U.S. Census of Population, 1880-1980,” *Social Science History* 15, no. 1 (Spring 1991), <https://users.pop.umn.edu/~ruggles/Articles/SSH1991.pdf>. [↵](#)
20. *Utah et al. v. Evans*, Secretary of Commerce *et al.*, 536 U.S. 452 (2002), available at <https://supreme.justia.com/cases/federal/us/536/452/>. [↵](#)

21. Simson Garfinkel, John M. Abowd, and Christian Martindale, “Understanding Database Reconstruction Attacks on Public Data,” *Communications of the ACM* 62, no. 3 (March 2019): 46-53, <http://dx.doi.org/10.1145/3287287>. Note that the examples in this article are for illustrative purposes only; in practice, official statistics agencies tend to publish median ages, and not means, and they typically do not publish mean ages for such small areas. [↵](#)
22. Garfinkel, Abowd, and Martindale, “Understanding Database Reconstruction Attacks.” [↵](#)
23. John Abowd, “Declaration of John Abowd, U.S. Census Bureau, State of Alabama v. U.S. Department of Commerce,” Appendix B, United States District Court for the Middle District of Alabama Eastern Division, filed April 13, 2021, available at <https://censusproject.files.wordpress.com/2021/04/2021.04.13-abowd-declaration-alabama-v-commerce-ii-final-signed.pdf>. [↵](#)
24. Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, “Calibrating Noise to Sensitivity in Private Data Analysis,” in *Proceedings of the Third Conference on Theory of Cryptography (TCC’06)*, ed. S. Halevi and T. Rabin (Berlin: Springer-Verlag, 2006), 265–84, https://doi.org/10.1007/11681878_14. [↵](#)
25. The Census Bureau had traditionally not released any of the software that it used for creating the decennial census, the American Community Survey, the Economic Census, or any of its other premiere data products. However, over the previous two decades the US government had increasingly grown accustomed to the idea of using open-source software and even contributing software back to the community. In August 2016, Tony Scott, the chief information officer of the US government, released the Federal Source Code Policy, which stated that it was the policy of the US government to release its internally developed software when there was no compelling reason not to do so. See Tony Scott, “The People’s Code,” CIO Council (blog entry), August 11, 2016, <https://www.cio.gov/2016/08/11/peoples-code.html>. The source code policy itself is Tony Scott, “Federal Source Code Policy: Achieving Efficiency, Transparency, and Innovation through Reusable and Open Source Software,” Executive Office of the President, Office of Management and Budget, Memorandum No. M-16-21, Washington, DC August 8, 2016, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2016/m_16_21.pdf. [↵](#)

26. Simson L. Garfinkel and Philip Leclerc, "[Randomness Concerns when Deploying Differential Privacy](https://doi.org/10.1145/3411497.3420211)," in *Proceedings of the 19th Workshop on Privacy in the Electronic Society (WPES'20)* (New York: ACM, 2020), 73–86, <https://doi.org/10.1145/3411497.3420211>. [↵](#)
27. Mary Catt, "ILR's John Abowd to lead US Census Bureau research," *Cornell Chronicle* (December 10, 2015), <https://news.cornell.edu/stories/2015/12/ilrs-john-abowd-lead-us-census-bureau-research>. [↵](#)
28. John M. Abowd, Gary L. Benedetto, Simson L. Garfinkel et al., "The Modernization of Statistical Disclosure Limitation at the US Census Bureau," Census Working Paper, August 2020, <https://www.census.gov/library/working-papers/2020/adrm/modernization-statistical-disclosure-limitation.html>. [↵](#)
29. US Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century*, Version 1.1 November 2015, <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan.pdf>. [↵](#)
30. OnTheMap is an interactive application on the Census Bureau's website that allows the user to draw an arbitrary geometry and determine a variety of statistics, such as how many commuters live within the boundary, where they work, and their income. The application had been under development when Dwork et al. published their paper on differential privacy in 2006, and the OnTheMap group retrofitted their system to incorporate differential privacy. For a discussion of OnTheMap, see Ashwin Machanavajjhala et al., "Privacy: Theory meets Practice on the Map," in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE '08)* (New York: IEEE Computer Society, 2008), 277–286, <https://doi.org/10.1109/ICDE.2008.4497436>. [↵](#)
31. John Abowd, "Census Bureau Awards Cooperative Agreements to Georgetown University and Purdue University," US Census Bureau, October 18, 2016, https://www.census.gov/newsroom/blogs/research-matters/2016/10/census_bureau_awards.html; JASON, *Formal Privacy Methods for the 2020 Census*, The MITRE Corporation, JSR-19-2F, April 2020, <https://irp.fas.org/agency/dod/jason/census-privacy.pdf>. [↵](#)
32. For a discussion of the 2010 Data Demonstration Products, see <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2010-demonstration-data-products.html>. [↵](#)

33. John M. Abowd, “How Will Statistical Agencies Operate When All Data Are Private?” *Journal of Privacy and Confidentiality* 7, no. 3 (2017), <https://doi.org/10.29012/jpc.v7i3.404>. ↵
34. Steven Ruggles et al., “Implications of Differential Privacy for Census Bureau Data and Scientific Research,” Working Paper No. 2018-6, Minnesota Population Center, Minneapolis, MN, December 2018. https://assets.ipums.org/_files/mpc/MPC-Working-Paper-2018-6.pdf ↵
35. Jennifer McDermott and Mike Schneider, “Researchers Question Census Bureau’s New Approach to Privacy,” Associated Press, September 28, 2019, <https://apnews.com/article/ri-state-wire-census-2020-us-news-ap-top-news-redistricting-88307905204e4d5aaad10db486514ccb>. ↵
36. John Abowd, “Modernizing Disclosure Avoidance: What We’ve Learned, Where We Are Now,” Research Matters, *Research Matters* (blog), US Census Bureau, March 13, 2020, https://www.census.gov/newsroom/blogs/research-matters/2020/03/modernizing_disclosu.html. ↵
37. National Academies of Sciences, Engineering, and Medicine, *Census Data Products: Data Needs and Privacy Considerations: Proceedings of a Workshop* (Washington, DC: The National Academies Press, 2020), <https://doi.org/10.17226/25978>. The video proceedings of the workshop can be found at <https://vimeo.com/showcase/6661772>. ↵
38. US Census Bureau, “Soliciting Feedback from Users on 2020 Census Data Products,” Notice, 83 Fed. Reg. 34111, July 19, 2018, <https://www.federalregister.gov/documents/2018/07/19/2018-15458/soliciting-feedback-from-users-on-2020-census-data-products>. ↵
39. See, for example, Dale Ho, “The Trump Administration’s Census Cover Up,” American Civil Liberties Union, July 27, 2018 (updated March 21, 2019), <https://www.aclu.org/blog/voting-rights/trump-administrations-census-cover>. ↵
40. US Census Bureau, “2020 Census Operational Adjustments Due to COVID-19,” last updated October 8, 2021, <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/operational-adjustments.html>. ↵
41. John Abowd and Victoria A. Velkoff, “Modernizing Disclosure Avoidance: A Multipass Solution to Post-processing Error,” *Research Matters* (blog), US Census

Bureau, June 18, 2020, https://www.census.gov/newsroom/blogs/research-matters/2020/06/modernizing_disclosu.html. ²

42. Nicholas Riccardi and Mike Schneider, “Alabama Sues to Stop Redistricting Delay, Privacy Initiative,” AP News, March 10, 2021,

<https://apnews.com/article/alabama-redistricting-lawsuits-census-2020-courts-9627c66d67629942d2f961bf781f4309>. For the lawsuit, see *State of Alabama v. US Department of Commerce*, Civil Action 2:18-cv-00772-RDP, US District Court for the Northern District of Alabama, May 21, 2021. ³

43. Mike Schneider, “16 States Back Alabama’s Challenge to Census Privacy Tool,” AP News, April 13, 2021, <https://apnews.com/article/race-and-ethnicity-alabama-florida-census-2020-us-news-3407a2d05be23972741294bc05838801>. ⁴

44. Mike Schneider, “Census on Track for August Data Release after Court Ruling,” AP News, June 30, 2021, <https://apnews.com/article/census-2020-courts-government-and-politics-bc46bd098e0d2234cd456d697f14a658>. ⁵

45. John Abowd and Victoria A. Velkoff, “Modernizing Privacy Protections for the 2020 Census: Next Steps,” *Research Matters* (blog), US Census Bureau, April 28, 2021, https://www.census.gov/newsroom/blogs/random-samplings/2021/04/modernizing_privacy.html. ⁶

46. US Census Bureau, “Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results,” June 9, 2021, <https://www.census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html>. ⁷

47. Michael Hawes and Michael Ratcliffe, “Understanding the 2020 Census Disclosure Avoidance System: Differential Privacy 201 and the TopDown Algorithm,” US Census Bureau, May 13, 2021, <https://www2.census.gov/about/training-workshops/2021/2021-05-13-das-presentation.pdf>. ⁸

48. The 2020 Census Redistricting File can be downloaded from https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/01-Redistricting_File--PL_94-171/. ⁹

49. US Census Bureau, “Disclosure Avoidance for the 2020 Census: An Introduction,” November 2, 2021, <https://www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html>. ¹⁰

50. For example, a law in Australia criminalizes re-identification. See Mark Phillips, Edward S. Dove, and Bartha M. Knoppers, “Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?” *Bioethical Inquiry* 14 (2017): 527–539, <https://doi.org/10.1007/s11673-017-9806-9>. ↵

51. Helen Nissenbaum, “Privacy as Contextual Integrity,” *Washington Law Review* 79, no. 1 (2004): 119–157, <https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/>. Approaches based on contextual integrity dictate that search engine companies like Google should not disseminate terms that users enter into search engines because, in context, search queries are like asking a question to a librarian. ↵

52. <https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/ppmf20210428/2021-04-28-das-metrics-overview.pdf>. ↵