

**IN THE UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

JUSTIN H. PHILLIPS,

Plaintiff,

v.

UNITED STATES BUREAU OF THE CENSUS,

Defendant.

COMPLAINT

No. _____

Dr. Justin H. Phillips—through his undersigned attorneys, the Election Law Clinic at Harvard Law School and Selendy Gay Elsberg PLLC—brings this action against the United States Bureau of the Census:

1. The decennial census is the statistical backbone of American society. This constitutionally mandated count is foundational to American democracy and government—not to mention the countless businesses and researchers who depend on the detailed, timely data it provides.

2. The United States Bureau of the Census (“Census Bureau” or “Bureau”) is required to keep all individual responses to census surveys strictly confidential. Federal law prohibits the Bureau from releasing any information that reveals the contents of any individual’s census response. 13 U.S.C. § 9(a)(2).

3. To manage privacy for the 2020 Census, the Bureau introduced a new two-phase algorithm designed to prevent aggregated data from being traced back to individual respondents. First, the privacy processing phase of the algorithm injected the data with random error, also referred to as “noise,” to protect individuals’ privacy. After the privacy-protecting noise was

added, the second, “post-processing” phase modified the data further in order to make the resulting data tables appear logical to laypeople and to make them compatible with conventional software.

4. Unfortunately, the Bureau’s post-processing method may have unintended and harmful consequences. Professional researchers—including Plaintiff Dr. Phillips, Professor of Political Science at Columbia University—are concerned that their research has been harmed because the post-processing phase may have systematically inflated the census-reported populations of sparsely populated and homogeneous areas while shrinking those with greater population density and diversity. This distortion in the 2020 Census would make the data less fit for use in Dr. Phillip’s research while also resulting in an inequitable distribution of political power and resources, likely harming racial minority groups.

5. The public does not know how much distortion the post-processing phase added to the decennial census data because the Bureau has failed to release the intermediate data set that includes the privacy-protecting noise but not the additional changes made during post-processing. Instead, the Bureau has only released the final data that includes distortions introduced by the post-processing phase.

6. Dr. Phillips asked the Bureau for the privacy protected intermediate data under the Freedom of Information Act (“FOIA”), 5 U.S.C. § 552, on July 7, 2022. Despite granting his request for expedited processing, and the passage of more than 116 days, the Bureau still has not given Dr. Phillips the records he requested, nor has it even told him whether it intends to comply, as required by FOIA.

7. Dr. Phillips brings this action to remedy the Bureau’s violation of FOIA.

PARTIES

8. Plaintiff Justin H. Phillips, Ph.D., is a Professor of Political Science and Chair of the Department at Columbia University. Dr. Phillips is a scholar of public opinion research and

related methodologies. In his academic research, he relies on sub-state census data to make accurate inferences about the population. He resides in New York County, New York.

9. Defendant the United States Bureau of the Census is a federal agency within the meaning of FOIA, 5 U.S.C. § 552(f)(1). It is a component agency of the United States Department of Commerce and is headquartered at 4600 Silver Hill Road, Suitland, MD 20746.

JURISDICTION AND VENUE

10. This court has subject matter and personal jurisdiction in this case under 5 U.S.C. § 552(a)(4)(B) and 28 U.S.C. § 1331.

11. Venue is premised on Plaintiff's residence and is proper under 5 U.S.C. § 552(a)(4)(B) and 28 U.S.C. § 1391(c)(1).

FACTS

Evolving Privacy Protections in the Decennial Census

12. Every ten years, the Census Bureau conducts the "actual Enumeration" of each state's population that is required by the Constitution. U.S. Const. art. I, § 2, cl. 3.

13. The publicly released results of this decennial census dictate everything from congressional apportionment to the legality of state and local redistricting plans to the distribution of hundreds of billions of dollars in federal funding.

14. The decennial census also provides the underpinning for the Bureau's population estimates and American Community Survey results for the remainder of the decade.

15. Local governments, academic researchers (such as Dr. Phillips), and businesses rely on census data to help them answer crucial questions in areas ranging from education to law enforcement to public opinion.

16. For many of the census’s key uses—including redistricting, government funding, and research—it is essential that the data not only report the accurate number of residents of each state, but also provide a realistic picture of how population is distributed *within* states.

17. The Census Bureau must balance its mission to provide useful and accurate statistics against the need to protect the privacy of the millions of census respondents. The Bureau is prohibited from releasing “any publication whereby the data furnished by any particular establishment or individual ... can be identified.” 13 U.S.C. § 9(a)(2). Bureau employees take a lifelong oath to protect response information, and the Bureau only publishes individual response data 72 years after its collection.

18. Even aggregated statistics based on individual responses can reveal information that can link back to specific people, especially in places with few people or where respondents have relatively unique attributes.

19. As a result, since at least 1930, the Bureau has taken steps to prevent individual responses from being “reidentified” from aggregated data. At first, the Bureau simply declined to publish certain data tables for geographic areas with very few residents if doing so might reveal individual responses. In recent decades, the Bureau started using more sophisticated techniques to introduce “noise” into the data by swapping household records between geographic areas.

20. Introducing some random intentional inaccuracy into the actual enumeration allows the Bureau to release data that is still useful but protects individual responses.

21. By definition, though, adding noise to census data makes it less accurate.

The 2020 Census’s New Disclosure Avoidance System

22. The Bureau refers to its efforts to protect the confidentiality of individual responses as its “disclosure avoidance system” or “DAS.” With the 2020 Census, the Census Bureau rolled out an all-new DAS that uses different techniques than previous iterations.

23. According to public statements by senior Census Bureau personnel, the Bureau needed to adopt a new approach because advances in computing power were making previous techniques, like swapping, increasingly vulnerable to reverse engineering and therefore reidentification of large swathes of the datasets.

24. At a high level, the 2020 DAS involved three iterations of data. To start, confidential base data was collected and tabulated. Next, a two-phase algorithm called the “Top-Down Algorithm” or “TDA” transformed the data in two steps. In the first phase, the privacy processing phase, the TDA generated privacy-protected “noisy” data by adding random noise so that individual responses cannot be reconstructed. In the second phase, the “post-processing” phase, the TDA further modified the privacy-protected noisy to ensure that all of the values are coherent and sum correctly. So far, only the final data produced by the post-processing phase has been released.

Confidential Base Data

25. The 2020 DAS began with raw data from the confidential individual responses that the Bureau never shares publicly. These individual responses, which the Bureau calls “microdata,” were then tabulated into data tables where each entry indicates a unique combination of traits (*e.g.*, census block, race, sex, age, and marital status) and the number of respondents who share all of those traits. This forms the confidential base data.

Privacy-Protected Noisy Data

26. Next, the first phase of the TDA—the privacy processing phase—injected a predetermined amount of random noise into nearly every statistic that was produced from the confidential base data.

27. There are three statistics, known as “invariants,” to which the Bureau did not add noise: (1) the total population of each state, the District of Columbia, and Puerto Rico; (2) the count of housing units in each census block; and (3) the count and type of group quarters, such as prisons and college dormitories, in each block. All other 2020 Census statistics received noise injections.

28. The privacy phase of the TDA was designed to protect privacy in a mathematically demonstrable manner. This technique, known as “differential privacy” or “DP,” allows the Bureau to precisely quantify the tradeoffs it makes between protecting confidentially and releasing accurate data, by setting what it calls a “privacy loss budget.”

29. This privacy loss budget quantifies the amount of reidentification risk the Census Bureau is willing to tolerate, and is represented by “ ϵ ” (the lower-case Greek letter “epsilon”).

30. After public demonstrations and feedback, the Bureau announced on June 9, 2021, that the privacy loss budget for the 2020 Census would be set at $\epsilon=19.61$ (including $\epsilon=17.14$ for persons data and $\epsilon=2.47$ for housing unit data). According to an official statement, the Bureau reached this decision after giving “significant consideration to the feedback [it] received from ... data users” who analyzed a set of privacy-protected demonstration data released in April 2021.

31. The privacy phase of the TDA used the privacy loss budget to determine exactly how much noise to add to the base data. Once the noise was added, the resulting data protected individual privacy to exactly the degree intended by the value for ϵ that was selected.

32. The data immediately resulting from the privacy phase of the TDA is known as “noisy data.”

33. As a way of representing facts about a population, noisy data may not always make logical sense. For example, the noisy data might indicate that a negative or fractional number of

people live in a certain area, where clearly only a non-negative, whole number is possible. The added noise might also mean that numbers that “should” add up do not. For example, the noisy populations of all counties in a state might not total to the actual population of the state. That is, the noisy counts would often give different populations for the same geographic area, depending on the summary level used.

Final Data

34. The second, “post-processing” phase of the TDA yielded the final data that the Bureau published.

35. The post-processing phase of the TDA began from the “top” at the national level and sequentially worked its way down to smaller and smaller geographic levels. At each level, the algorithm “cleaned up” the noisy data, making additional changes beyond the noise needed to protect privacy in order to ensure that the statistics at each level add up correctly and do not include negative or fractional values.

36. Finally, the results of the post-processing phase were used to generate “synthetic” individual responses (or “microdata”), from which the final data is tabulated and then published.

37. The post-processing phase of the TDA may have introduced systemic distortion. For example, the need to avoid negative numbers means that this phase of the algorithm frequently *adds* people to low-population areas, and ensuring that the state totals sum correctly requires making corresponding *reductions* to large areas. The cumulative effects of the post-processing phase are likely to produce estimates that are biased against racial minority groups.

Publication of the Final Data

38. The Bureau began releasing the cleaned up final data on August 12, 2021. States use this data to draw congressional and state legislative district boundaries.

39. However, the Bureau did not release the privacy protected noisy data that was used to generate the 2020 Census data, even though both the privacy protected noisy data and the final data are designed to sufficiently protect individual privacy.

40. Without the privacy protected noisy data files to compare against the final data, researchers like Dr. Phillips and members of the public have no way to know how much distortion the post-processing phase of the TDA introduced, or whether that distortion had systemic effects, including to the detriment of racial minority groups.

Dr. Phillips's FOIA Request

41. On July 7, 2022, Dr. Phillips requested the privacy protected noisy data files for both the 2020 Census and the demonstration data pulled from the 2010 Census. A true and correct copy of Dr. Phillips's FOIA request is attached to this Complaint as Exhibit A.

42. Dr. Phillips requested that the Bureau expedite the processing of his request and waive his fees as required by FOIA.

43. As Dr. Phillips explained in his request, obtaining the privacy protected noisy data is time-sensitive. The final 2020 Census data is already being used for a wide array of critical government functions and decisions, including redistricting. Scholars including Dr. Phillips are also using the census data for important and timely research, which would be undermined to the extent that the census data includes bias. Without a comparison between the final census data and the intermediate noisy data, the public cannot fully understand the final data's fitness for use because it is unclear how much distortion the post-processing phase introduced.

44. Furthermore, planning is already underway for the upcoming 2030 Census. The Bureau plans to address questions of data privacy during the 2023 fiscal year, so it is crucial that Dr. Phillips and other researchers are given not only the data they need in order to fully evaluate the Bureau's 2020 approach to privacy, but also sufficient time to make findings in order to inform

changes to the upcoming census if they uncover any problematic, systemic distortions introduced by the current algorithms.

45. On July 7, 2022—the same day the Dr. Phillips submitted his request—the Bureau acknowledged receipt of the request and assigned it the tracking number DOC-CEN-2022-002073.

46. On August 26, 2022, the Bureau granted Dr. Phillips’s requests for both expedited processing and a fee waiver. Dr. Phillips was notified of these grants via e-mails sent through the online FOIA system to his counsel. These notifications provided no updates on the status of the request and did not specify when (or whether) Dr. Phillips could expect either a decision or the responsive records themselves.

47. On September 28, 2022, Dr. Phillips’s counsel called the Census Bureau FOIA office to ask about the status of his request. Chauvez Bennett, a FOIA analyst at the Bureau, answered the call and explained that the request had been “sent to the program areas,” but he could give no estimate for when they would provide information to the FOIA office, nor when the FOIA office would provide information to Dr. Phillips. Mr. Bennett noted that Sarabeth Rodriguez was the officer assigned to DOC-CEN-2022-002073, and took a message asking Ms. Rodriguez to call back Dr. Phillips’s attorneys. As of the filing of this Complaint, Ms. Rodriguez has not returned the call.

48. It has now been 116 days (and 80 business days) since Dr. Phillips submitted his FOIA request and the Bureau confirmed receipt, and 66 days (and 44 business days) since the Bureau granted Dr. Phillips’s request for expedited processing. To date, however, the Bureau has not released any of the requested records. It has not informed Dr. Phillips whether it will comply with his request or even when it expects to reach that decision.

COUNT I

Violation of the Freedom of Information Act, 5 U.S.C. § 552: Failure to Provide Timely Decision on Request

49. FOIA requires federal agencies to promptly decide on requests for records. An agency must “determine within 20 days (excepting Saturdays, Sundays, and legal public holidays) after the receipt of any such request whether to comply” and notify the requestor of its decision “immediately.” 5 U.S.C. § 552(a)(6)(A)(i). The 20-day period begins when the “appropriate component of the agency” first receives the request or “ten days after the request is first received by any component of the agency that is designated [by regulation] to receive requests,” whichever comes first. *Id.* § 552(a)(6)(A)(ii).

50. If there are “unusual circumstances,” an agency may delay its response by up to “ten working days” but must provide written notice to the requestor and an estimate of when a response can be expected. *Id.* § 552(a)(6)(B)(i). Accordingly, even if it takes an agency ten or more days to route a request to the correct component, and even if there are other “unusual circumstances” that prevent it from responding quickly, FOIA gives the agency at most 40 business days to indicate whether it will comply with the request. The Census Bureau has failed to comply with this statutory provision.

51. Under FOIA, if an agency “fails to comply with the applicable time limit provisions,” the requestor “shall be deemed to have exhausted his administrative remedies with respect to such request.” *Id.* § 552(a)(6)(C)(i).

52. Dr. Phillips properly requested records within the Census Bureau’s custody on July 7, 2022, and the Bureau acknowledged receipt the same day.

53. The Bureau was obligated under FOIA to decide whether to comply with Dr. Phillips’s request within 20 business days, or by August 5, 2022, and to inform him of its decision

immediately. Even if the Bureau had invoked all extensions contemplated by FOIA, it would have been obligated to make a determination within 40 business days, or by September 1, 2022.

54. Because the Bureau has failed to comply with applicable time limit provisions under FOIA, Dr. Phillips has exhausted his administrative remedies for this request.

55. By failing to decide whether to comply with Dr. Phillips's request and inform him of its decision within the required time, and by failing to release responsive records, the Bureau is violating FOIA.

COUNT II

Violation of the Freedom of Information Act, 5 U.S.C. § 552: Failure to Produce Records

56. On July 7, 2022, Dr. Phillips properly requested records within the Census Bureau's custody and the Bureau acknowledged receipt.

57. The Bureau is obligated under FOIA to "promptly" release requested records. 5 U.S.C. § 552(a)(3)(A).

58. FOIA allows requesters to seek expedited processing if they can demonstrate a "compelling need," 5 U.S.C. § 552(a)(6)(E)(i)(I), including "urgency to inform the public concerning actual or alleged Federal Government activity," *id.* § 552(a)(6)(E)(v)(II). An agency must determine whether to grant a request for expedited processing within ten days. *Id.* § 552(a)(6)(E)(ii)(I). If an agency grants a request for expedited processing, it must then process the request "as soon as practicable." *Id.* § 552(a)(6)(E)(iii).

59. On August 26, 2022, the Bureau granted Dr. Phillips's request for expedited processing.

60. The Bureau has not released any of the records Dr. Phillips requested, nor has it notified Dr. Phillips of any adverse decisions or intent to claim statutory exemptions.

61. Because the Bureau has failed to comply with applicable time limit provisions under FOIA, Dr. Phillips has exhausted his administrative remedies for this request.

62. By failing to produce any of the records Dr. Phillips requested, or to provide an explanation and final decision regarding any records not produced, the Bureau is violating FOIA.

REQUEST FOR RELIEF

WHEREFORE, Dr. Phillips respectfully requests that this court:

- A. Declare that the Census Bureau is violating FOIA by failing to determine whether it will comply with Dr. Phillips' request within the statutory timeframe.
- B. Order the Census Bureau to respond to Dr. Phillips' request expeditiously and within a specified number of days (not to exceed ten business days).
- C. Order the Census Bureau to produce all records it determines it is required to produce as soon as practicable and within a specified number of days (not to exceed sixty calendar days).
- D. Award Dr. Phillips any attorneys' fees and costs he reasonably incurs in this action pursuant to FOIA, 5 U.S.C. § 552(a)(4)(E).
- E. Grant such other relief as the Court deems just and proper.

Dated: October 31, 2022

/s/ Jordan A. Goldstein

Jordan A. Goldstein
Jeffrey Zalesin
SELENDY GAY ELSBERG PLLC
1290 Avenue of the Americas
New York, NY 10104
(212) 390-9000
jgoldstein@selendygay.com
jzalesin@selendygay.com

* Motion for admission pro hac vice forthcoming

‡ Motion for admission as a student attorney forthcoming

/s/ Theresa J. Lee

Theresa J. Lee
Ruth Greenwood*
Jhanelle Bisasor †
Nikol Tang ‡
ELECTION LAW CLINIC
HARVARD LAW SCHOOL
6 Everett Street, Ste. 4105
Cambridge, MA 02138
(617) 496-0370
thlee@law.harvard.edu
rgreenwood@law.harvard.edu

EXHIBIT A



July 7, 2022

Vernon E. Curry, PMP, CIPP/G
FOIA Officer
U.S. Census Bureau, Room 3J235
4600 Silver Hill Road
Washington, DC 20233-3700

RE: Freedom of Information Act Request

Dear Mr. Curry:

The Election Law Clinic at Harvard Law School (“ELC”) submits this request on behalf of Justin H. Phillips, Professor in the Department of Political Science at Columbia University (“Professor Phillips”) to the Bureau of the Census (“Census Bureau” or “Bureau”) pursuant to the Freedom of Information Act (“FOIA”), 5 U.S.C. § 552 *et seq.*

I. Requested Records

Professor Phillips requests the following records:

1. All machine-readable files that describe the 2020 Census data for all fifty states after noise has been injected but before post-processing has occurred. Referred to as the “noisy” measurements stage, these may take the form of “histograms” as described in Abowd et al 2019. These file(s) are also described by Abowd as those produced by line 2 of algorithm 1 and referred to as “A,” as the “noisy answer to a query,” or as a “measurement.”¹
2. All machine-readable files that describe the 2010 Census production settings demonstration data² for all fifty states after noise has been injected but before post-processing has occurred. Referred to as the “noisy” measurements stage, these may take the form of “histograms” as described in Abowd et al 2019. These file(s) are also described by Abowd as those produced by line 2 of algorithm 1 and referred to as “A,” as the “noisy answer to a query,” or as a “measurement.”³

Please note that this request encompasses both digital and physical records.

If you determine that some portions of the requested records are exempt from disclosure, please (1) disclose any reasonably segregable portions of the requested records after deletion of materials exempt pursuant to 5 U.S.C. § 552(b); (2) provide an index of the withheld materials as required

¹ JOHN ABOWD ET AL., CENSUS TOPDOWN: DIFFERENTIALLY PRIVATE DATA, INCREMENTAL SCHEMAS, AND CONSISTENCY WITH PUBLIC KNOWLEDGE 4–6 (2019), <https://systems.cs.columbia.edu/private-systems-class/papers/Abowd2019Census.pdf>.

² The Census Bureau produced these demonstration data by applying the 2020 Census disclosure avoidance system to confidential data from the 2010 Census. *See* U.S. CENSUS BUREAU, DEVELOPING THE DAS: DEMONSTRATION DATA AND PROGRESS METRICS (2022), <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-development.html>.

³ Abowd et al., *supra* note 1, at 4–6.



under *Vaughn v. Rosen*, 484 F.2d 820 (D.C. Cir. 1973), *cert. denied*, 415 U.S. 977 (1974); and (3) provide a detailed justification of the claimed exemptions as required under *Vaughn v. Rosen*. *Id.*

II. Background

The decennial census is the backbone of our nation’s statistical infrastructure. The constitutionally mandated “actual enumeration” of each state’s population, U.S. CONST. art. I, § 2 cl. 3, does far more than dictate the apportionment of congressional seats. Among its many uses, the census determines the legality of state and local redistricting plans, the distribution of public funding at all levels of government, and the results of important research projects conducted by academic researchers and government agencies. It is therefore crucial that the census not only report the total population of each state as accurately as possible, but also present a realistic picture of how population is distributed at lower levels of geography within the states.

The Census Bureau must balance this interest in accuracy against a competing interest in privacy. Under 13 U.S.C. § 9(a)(2), the Census Bureau is required to protect the privacy of individual responses to the decennial census. To maintain this privacy, the Bureau develops a “disclosure avoidance system,” or “DAS,” to mask identifying information within the data before releasing information to the public.⁴ For the 2000 and 2010 censuses, the Bureau relied primarily on “data swapping” for disclosure avoidance. This method identified households that had a high risk of disclosure based on their uniqueness and “swapped” different records between tracts without altering the total populations of census blocks.⁵ The 2010 DAS also incorporated partially synthetic data regeneration, top-coding, bottom-coding, recoding, and noise infusion for large households.⁶

In response to concerns regarding the privacy of census data, the Bureau adopted a new DAS for the 2020 census.⁷ The goal of the new DAS was to reach differential privacy, a mathematical definition of privacy that refers to a viewer’s inability to tell whether particular data was included in the original dataset when viewing an algorithm’s output.⁸ To accomplish this, the Bureau added a pre-determined amount of random noise at particular points in the processing of the dataset.

The 2020 Census made use of the TopDown Algorithm (“TDA”) for disclosure avoidance. The algorithm first infuses differentially private noise into the Census Edited File, then post-processes

⁴ JADE FORD & MOLLY DANAHY, UNDERSTANDING DIFFERENTIAL PRIVACY AND ITS IMPACT ON 2020 CENSUS DATA 1 (2021), <https://campaignlegal.org/sites/default/files/2021-05/Issue%20Brief%20Differential%20Privacy%202021.05.04%20VF.pdf>; *see also* LAURA MCKENNA, DISCLOSURE AVOIDANCE TECHNIQUES USED FOR THE 1960 THROUGH 2010 DECENNIAL CENSUSES OF POPULATION AND HOUSING PUBLIC USE MICRODATA SAMPLES, U.S. CENSUS BUREAU (April 2019), <https://www2.census.gov/adrm/CED/Papers/CY19/2019-04-McKenna-Six%20Decennial%20Censuses.pdf> (describing the history of disclosure avoidance systems).

⁵ *See* McKenna, *supra* note 4, at 5.

⁶ *Id.* at 9.

⁷ Ron Jarmin, *Census Bureau Adopts Cutting Edge Privacy Protections for 2020 Census*, CENSUS BLOG (Feb. 15, 2019), https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html.

⁸ Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy*, 9 FOUND. AND TRENDS IN COMP. SCI. 5–6 (2014), <https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf>.



this data to meet invariants set in advance. Michael Hawes, a senior advisor at the Census Bureau, has described the functioning of the TDA as follows:

At a high level the TDA has five steps. [1] It inputs the Census Edited File microdata and the geographic reference file geocoding. [2] Then it converts those microdata to a functionally equivalent histogram of counts. You can think of this as a fully saturated contingency table of every variable crossed with every other variable with the value of those cells being the number of people at that level of geography who have those specific characteristics, along with all structural zeros which will be impossible values according to the edit rules for the CEF having been removed.

[3] Then the algorithm asks a number of queries and injects noise into those results. We call this the noisy measurement stage. . . .

[4] Armed with these noisy measurements the system must then perform a set of optimization problems. These are designed to ensure consistency across tables and geographies and to ensure that the final histogram is populated with non-negative integer counts.

[5] Finally, the algorithm transforms the resulting histogram back into privacy protected microdata that can be output into the Decennial Tabulation Systems.⁹

While the 2020 DAS might provide improved privacy protections when compared with the data swapping methods used in 2000 and 2010, some experts have voiced concerns that this DAS will amplify existing disparities within the census,¹⁰ which persistently undercounts specific groups including the Black, Latino, and American Indian and Alaska Native populations.¹¹ For example, research analyzing data published by the Census Bureau to demonstrate the application of the 2020 Census DAS found a bias that over-counted the population of smaller counties while undercounting the populations of larger counties.¹² Other research suggests that the 2020 Census DAS may be biased in the direction of reducing the populations of racially and politically diverse neighborhoods and increasing the counts of more homogeneous neighborhoods.¹³ These biases

⁹ MICHAEL HAWES, WEBINAR: DIFFERENTIAL PRIVACY 201 AND THE TOPDOWN ALGORITHM (May 13, 2021), transcript available at <https://www2.census.gov/about/training-workshops/2021/2021-05-13-das-transcript.pdf>.

¹⁰ See, e.g., STEVEN A. OCHOA & TERRY AO MINNIS, PRELIMINARY REPORT: IMPACT OF DIFFERENTIAL PRIVACY & THE 2020 CENSUS ON LATINOS, ASIAN AMERICANS AND REDISTRICTING 1-2, 5-6 (April 2021), <https://www.maldef.org/wp-content/uploads/2021/04/FINAL-MALDEF-AAJC-Differential-Privacy-Preliminary-Report-4.5.2021-1.pdf>.

¹¹ See *Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census*, U.S. CENSUS BUREAU (Mar. 10, 2022), <https://www.census.gov/newsroom/press-releases/2022/2020-census-estimates-of-undercount-and-overcount.html>.

¹² See Ochoa & Minnis, *supra* note 10, at 1–2, 5–6.

¹³ See Christopher T. Kenny et al., *The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 U.S. Census*, 7 SCI. ADV. 1, 1 (2021), <https://imai.fas.harvard.edu/research/files/DAS.pdf>. **Error! Hyperlink reference not valid.**



can harm communities' ability to obtain their fair share of government funding¹⁴ and to enforce their civil rights, including the right to equal political opportunity under the Voting Rights Act.¹⁵ Bias in census data also interferes with academic research in health, public opinion, and many other fields.

Analysts outside the Census Bureau have limited information about the sources of any bias introduced by the 2020 Census DAS. However, research has raised questions about whether post-processing plays a systematic role in shrinking the published counts some types of communities while inflating others.¹⁶ The Census Bureau's own top scientists have acknowledged evidence that post-processing "introduced an upward bias for small populations and a corresponding downward bias for large ones."¹⁷

To analyze the potential impact of the 2020 DAS on accurate population counts, including the effect of post-processing, Professor Phillips requests access to the noisy measurements files underlying the published 2020 Census data and the 2010 Census demonstration data product, as detailed above.

III. Application for Fee Waiver or Limitation of Fees

Professor Phillips asks that all fees for this request be waived or, in the alternative, limited. This request qualifies for such a waiver because the requested disclosure is in the public interest and is non-commercial in nature. In the alternative, this request qualifies for an exemption from search and review costs because Professor Phillips is affiliated with Columbia University, an educational institution, and this request is being made for a scholarly purpose.

In the event that the application for a complete fee waiver is denied and you estimate that more than \$100 in fees will be charged, please contact us before proceeding with any search, review, or duplication.

A. This request is non-commercial and sufficiently benefits the public interest to justify a fee waiver.

Under FOIA, an agency is required to provide records without charge or at a reduced charge if "disclosure of the information is in the public interest because it is likely to contribute significantly to public understanding of the operations or activities of the government and is not primarily in the commercial interest of the requester."¹⁸ The Census Bureau uses the following criteria when deciding whether a request merits a fee waiver:

¹⁴ Andrew Reamer, *Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds*, GEO. WASH. INST. PUB. POL'Y (Apr. 29, 2020), <https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds>.

¹⁵ See Ochoa & Minnis, *supra* note 10, at 1–2, 5–6.

¹⁶ See Kenny et al., *supra* note 13; Aloni Cohen et al., *Census TopDown: Investigating the Impacts of Differential Privacy on Redistricting*, 2ND SYMPOSIUM ON FOUNDATIONS OF RESPONSIBLE COMPUTING 5:22 (2021), <https://drops.dagstuhl.de/opus/volltexte/2021/13873/pdf/LIPIcs-FORC-2021-5.pdf>.

¹⁷ John M. Abowd & Michael B. Hawes, Confidentiality Protection in the 2020 US Census of Population and Housing 24 (Jun. 7, 2022) (pre-print), <https://arxiv.org/pdf/2206.03524.pdf>.

¹⁸ 5 U.S.C. § 552(a)(4)(A)(iii).



- (1) The subject of the requested records must concern identifiable operations or activities of the Federal Government;
- (2) the disclosable portions of the requested records must be meaningfully informative about Government operations or activities in order to be “likely to contribute” to an increased public understanding of those operations or activities;
- (3) the disclosure of the requested information must contribute to the understanding of a reasonably broad audience of persons interested in the subject, as opposed to the individual understanding of the requester;
- (4) the disclosure of the requested information is likely to contribute “significantly” to the public’s understanding of Government operations or activities;
- (5) whether the requester has a commercial interest that would be furthered by the requested disclosure; and
- (6) whether any identified commercial interest of the requester is sufficiently great, in comparison with the public interest in disclosure, such that the disclosure is primarily in the commercial interest of the requester.¹⁹

All six factors favor Professor Phillips’s application for a fee waiver in connection with this request.

First, this request clearly concerns “identifiable operations or activities of the Federal Government,” since it involves data created by the Census Bureau in preparing data products for publication.

Second, the request contributes meaningfully to the public understanding of the Bureau’s activities because it will provide new information about the impact of the 2020 DAS on the accuracy and fitness for use of the released 2020 Census data. As two senior Census Bureau officials—Associate Director for Research and Methodology John M. Abowd and Senior Advisor for Data Access and Privacy Michael B. Hawes—have noted, “[t]ransparency in the decision-making process regarding disclosure limitation is incredibly valuable, as it can help data users better understand any relevant limitations of the data they are using.”²⁰

Stakeholders from across and outside of the political spectrum are actively debating the Census Bureau’s application of differential privacy. For example, the nonpartisan National Conference of

¹⁹ 15 CFR § 4.11(l); *see also* U.S. CENSUS BUREAU, UNITED STATES CENSUS BUREAU FREEDOM OF INFORMATION ACT (FOIA) REFERENCE GUIDE (July 2015), https://www2.census.gov/foia/resources/foia_reference_guide.pdf.

²⁰ Abowd & Hawes, *supra* note 17, at 24.



State Legislatures wrote letters to the U.S. House,²¹ Senate,²² and Census Bureau²³ voicing concerns about the impact of the 2020 DAS on compliance with the “one-person, one-vote” principle and the Voting Rights Act. Similarly, the State of Alabama filed suit in early 2021, challenging the implementation of differential privacy on the grounds that differential privacy will result in data that conflicts with the State’s right to receive accurate “tabulations of population” under 13 U.S.C. § 141. *Alabama v. United States Dep’t of Com.*, 546 F. Supp. 3d 1057 (M.D. Ala. 2021). While the case was ultimately dismissed by a three-judge district court panel, the amicus filings reflect the broad spectrum of legal arguments for and against the use of differential privacy. In support of plaintiffs were amici from sixteen states,²⁴ the Senate of Pennsylvania Republican Caucus,²⁵ the State Government Leadership Foundation,²⁶ a professor of law,²⁷ and a historian.²⁸ The Electronic Privacy Information Center filed an amicus brief in support of Defendants.²⁹ This variety of parties suggests a profound public interest in differential privacy and the balance between privacy and accuracy.

Scholars in statistics, computer science, and research methodology are also debating the effects of differential privacy on the census and redistricting processes. Some experts have highlighted the benefits of differential privacy, including its mathematical rigor and its potential to allow analysts to incorporate confidence intervals when they use census data.³⁰ However, there are also scholars who suggest that the 2020 DAS will make it more challenging both to draw districts that meet

²¹ Letter from Tim Storey, Exec. Dir., National Conference of State Legislatures., to The Honorable Carolyn Maloney, Chair, House Committee on Oversight and Reform, and The Honorable Jim Jordan, Ranking Member, House Committee on Oversight and Reform (May 14, 2020), <https://www.ncsl.org/documents/statefed/House-Census-letter-5-20.pdf>.

²² Letter from Tim Storey, Exec. Dir., National Conference of State Legislatures, to The Honorable Ron Johnson, Chair, Senate Committee on Homeland Security and Governmental Affairs, and the Honorable Gary Peters, Ranking Member, Senate Committee on Homeland Security and Governmental Affairs (June 1, 2020), <https://www.ncsl.org/documents/statefed/Senate-Census-Letter-FINAL0620.pdf>.

²³ Letter from Tim Storey, Exec. Dir., National Conference of State Legislatures, to The Honorable Steven Dillingham, Dir., U.S. Census Bureau (May 26, 2020), <https://www.ncsl.org/documents/statefed/Census-Bureau-letter-May26-FINAL.pdf>.

²⁴ Brief for State of Utah and 15 Other States as Amici Curae Supporting Plaintiffs, *Alabama v. Dept. of Commerce*, 2021 WL 2668810 (M.D. Ala. 2021) (3:21-cv-00211-RAH-ECM-KCN).

²⁵ Brief for (1) Senate of Pennsylvania Republican Caucus, (2) Pennsylvania Senate President Pro Tempore Jake Corman, and (3) Pennsylvania Senate Majority Leader Kim Ward as Amici Curae Supporting Plaintiffs, *Alabama v. Dept. of Commerce*, 2021 WL 2668810 (M.D. Ala. 2021) (3:21-cv-00211-RAH-ECM-KCN).

²⁶ Memorandum in Support of the State Government Leadership Foundation et al., *Alabama v. Dept. of Commerce*, 2021 WL 2668810 (M.D. Ala. 2021) (3:21-cv-00211-RAH-ECM-KCN).

²⁷ Brief for Professor Jane Bambauer as Amicus Curae Supporting Plaintiff’s Complaint for Declaratory and Injunctive Relief, *Alabama v. Dept. of Commerce*, 2021 WL 2668810 (M.D. Ala. 2021) (3:21-cv-00211-RAH-ECM-KCN).

²⁸ Margo Anderson’s Motion for Leave to File Amicus Curiae Brief in Support of Plaintiffs, *Alabama v. Dept. of Commerce*, 2021 WL 2668810 (M.D. Ala. 2021) (3:21-cv-00211-RAH-ECM-KCN).

²⁹ Brief for Electronic Privacy Center as Amicus Curae Supporting Defendants’ Response in Opposition to Plaintiffs’ Motion for Preliminary Injunction and Petition for Writ of Mandamus, *Alabama v. Dept. of Commerce*, 2021 WL 2668810 (M.D. Ala. 2021) (3:21-cv-00211-RAH-ECM-KCN).

³⁰ See JADE FORD AND MOLLY DANAHY, UNDERSTANDING DIFFERENTIAL PRIVACY AND ITS IMPACT ON 2020 CENSUS DATA 11 (2021); Moon Duchin, *Privacy in Census Data: Math Meets Policy* at 41:40, TUFTS SEMINAR (July 7, 2020), <https://www.youtube.com/watch?v=oIRL5vBWq84>.



Voting Rights Act requirements and to identify potentially malapportioned districts. For example, a study by Christopher T. Kenny et al. finds that the 2020 DAS “systematically undercounts the population in mixed-race and mixed-partisan precincts, yielding unpredictable racial and partisan biases.”³¹ This article suggests that the added noise makes it “impossible for states to follow the principle of One Person, One Vote, as it is currently interpreted by courts and policy-makers.”³²

In short, there is deep disagreement over whether the 2020 DAS provided a dataset that is fit for its intended purposes, including redistricting and enforcement of voting and civil rights.

Unfortunately, this debate is taking place against a backdrop of incomplete information. Abowd and Hawes recently acknowledged that “the Census Bureau’s external stakeholders were justifiably disappointed and concerned about the timing, pace, and mechanisms of the agency’s public engagement regarding the development of the 2020 DAS.”³³ The Bureau still has not fully addressed the shortcomings in its public engagement, including its failure to publish noisy data for researchers to compare with post-processed data.

Thus, the public can make only educated guesses about whether any biases in the 2020 DAS are caused by the injection of noise, post-processing, or both. This question has profound policy implications. Without understanding how (if at all) the 2020 DAS introduces bias, stakeholders cannot identify the aspects of the DAS (if any) that need to be reformed or rethought as the Bureau prepares for the 2030 Census. Having access to the noisy measurements file will empower the public to engage more intelligently in debates over the future of privacy protection in census data. Noisy data will also provide insight into the quality of the official 2020 Census data by revealing whether post-processing caused systematic distortions.

Third, this disclosure will be useful and interesting to an expansive pool of stakeholders including state and local governments, academic scholars, think tanks, civil rights groups, and other non-profit organizations. Census data generate some 17,000 publications per year and are “widely used in analyses of the economy, population change, and public health,” as well as being “indispensable tools for federal, state, and local planning.”³⁴ Everyone who works directly with census data or relies on the results of such work would benefit from transparency regarding the 2020 DAS and any biases it may introduce. Releasing this information would help data users distinguish between use cases for which census data is fit and those that may be compromised by the application of differential privacy.

Fourth, the contribution that this request will make to public understanding will be significant.

As noted above, there is significant dispute over the effects of the 2020 DAS. Currently, the public has minimal insight into how, if at all, the 2020 DAS may have made the data less fit for use.

³¹ Christopher T. Kenny et al., *The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 U.S. Census*, 7(41) SCI. ADV. 1, 1 (2021), <https://imai.fas.harvard.edu/research/files/DAS.pdf>

³² *Id.* at 12.

³³ JOHN M. ABOARD & MICHAEL B. HAWES, CONFIDENTIALITY PROTECTION IN THE 2020 US CENSUS OF POPULATION AND HOUSING 24 (Jun. 7, 2022), <https://arxiv.org/pdf/2206.03524.pdf>.

³⁴ Steven Ruggles et al., *Differential Privacy and Census Data: Implications for Social and Economic Research*, 109 AM. ECON. ASS’N PAPERS & PROC. 403, 403 (2019).



Professor Phillips seeks to obtain information that will help measure the potential biases that may have been incorporated when the Bureau conducted post-processing. By revealing this information, this request may settle some of the ongoing debates over the impact of the 2020 DAS on the usefulness of 2020 Census data for redistricting, scholarship, and other use cases.

Fifth, Professor Phillips and ELC have no commercial interest in this FOIA request. Professor Phillips is an academic researcher who is solely interested in studying the impact of the 2020 DAS and communicating those findings to the public. ELC is a non-profit organization that provides legal services to a variety of clients. ELC is submitting this FOIA request on behalf of Professor Phillips solely to further their missions to educate and serve the public.

Sixth, since Professor Phillips and ELC have no commercial interest in documents related to this request, there is no need to balance that interest with the public interest: this request is not primarily in the commercial interest of the requester.

For the reasons listed above, Professor Phillips and ELC request that any fees associated with this request be waived pursuant to 5 U.S.C. § 552(a)(4)(A)(iii) and in conformity with 15 CFR § 4.11(l).

B. Professor Phillips is exempt from search and review fees as a member of an educational institution.

In the event that the Bureau determines Professor Phillips is not eligible for a complete fee waiver, he requests a waiver of search and review fees based on his status as a member of an educational institution.

Pursuant to 5 U.S.C. § 552(a)(4)(A)(ii)(II), fees shall be limited to “reasonable standard charges for document duplication when records are not sought for commercial use and the request is made by an educational or noncommercial scientific institution, whose purpose is scholarly or scientific research.” Professor Phillips is a tenured professor at Columbia University who intends to use the results of this request for scholarly research.

IV. Application for Expedited Processing

Professor Phillips requests that the processing of this request be expedited pursuant to 15 C.F.R. § 4.6(f)(1). This request qualifies for expedited processing for two independent reasons.

First, the request involves “[a] matter of widespread and exceptional media interest involving questions about the Government’s integrity which affect public confidence.”³⁵

The 2020 Census has attracted significant interest from the media and the public both because of the consequences the census has and because of the context in which the census was conducted. Census results guide some of the most significant political developments of each decade, including the redrawing of voting districts and the distribution of trillions of federal dollars through hundreds

³⁵ 15 C.F.R. § 4.6(f)(1)(iii)



of federal spending programs.³⁶ The 2020 Census results were published in the wake of the COVID-19 pandemic, which put the Bureau under immense stress and increased public scrutiny of the results.³⁷ Much of the public debate about the 2020 DAS centers on whether it amplifies inequities in the census that already exist due to unequal undercounts. Given the intense scrutiny of the census generally, it is unsurprising that the implementation of a new DAS would generate “widespread and exceptional” interest and affect public confidence in government functions that rely on census data.

Second, the request involves “[a]n urgency to inform the public about an actual or alleged government activity.”³⁸

While the collection of household responses for the next decennial census will not begin until 2030, the planning phase of the 2030 Census program is already underway. The Government Accountability Office has indicated that the Census Bureau is currently undertaking its “Design selection phase” where it selects “[m]ajor design innovations for Census 2030.”³⁹ This portion of the cycle is scheduled through fiscal year 2024.⁴⁰ The Census Bureau’s budget submission for fiscal year 2023 also highlights the ongoing design selection phase, where the architecture to carry out the 2030 Census is being planned.⁴¹ Further, the Census Bureau indicates that during fiscal year 2023, it will begin work on data privacy, “including a focus on maintaining or improving data accuracy while using differential privacy to ensure disclosure avoidance for respondent data.”⁴² Given the time constraints on conducting this research in time to provide advice and analysis for the 2030 Census, this request involves clear urgency to inform the public.

In order to expedite delivery of these requested documents and in order to reduce possible fees incurred, I am requesting that these documents be delivered either digitally via email or on a data disk via the U.S. Postal Service.

³⁶ See Andrew Reamer, *Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds*, GEO. WASH. INST. PUB. POL’Y (Apr. 29, 2020), <https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds>; Ron Jarmin, *Redistricting Data: What to Expect and When*, CENSUS BLOG (July 28, 2021), <https://www.census.gov/newsroom/blogs/director/2021/07/redistricting-data.html>.

³⁷ See Jarmin, *supra* note 36; Yuriy Rudensky, *How Changes to the 2020 Census Timeline will Impact Redistricting*, BRENNAN CENTER (May 4, 2020), <https://www.brennancenter.org/our-work/research-reports/how-changes-2020-census-timeline-will-impact-redistricting>; Zach Montellaro, *Delayed census data kicks off flood of redistricting lawsuits*, POLITICO (May 1, 2021), <https://www.politico.com/news/2021/05/01/redistricting-lawsuits-485161>.

³⁸ 15 C.F.R. § 4.6(f)(1)(iv).

³⁹ U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-22-104357, 2020 CENSUS: LESSONS LEARNED FROM PLANNING AND IMPLEMENTING THE 2020 CENSUS OFFER INSIGHTS TO SUPPORT 2030 PREPARATIONS 3 (2022), <https://www.gao.gov/assets/gao-22-104357.pdf>.

⁴⁰ *Id.*

⁴¹ U.S. CENSUS BUREAU, U.S. CENSUS BUREAU’S BUDGET FISCAL YEAR 2023, CEN-158 (2022), <https://www2.census.gov/about/budget/congressional-budget-justification-fy-2023.pdf>.

⁴² *Id.* at CEN-163.



Please email copies of responsive documents to:

rgreenwood@law.harvard.edu

Or please mail copies of responsive documents to:

Ruth Greenwood
Election Law Clinic
Harvard Law School
6 Everett Street, Suite 4105
Cambridge, MA 02138

Should you elect, for any reason, to withhold, redact, or deny the release of any record responsive to this request, I ask that you provide me with an explanation for each withholding or redaction, along with pertinent legal citations.

Please confirm receipt of this request and provide me with an estimate of processing time.

Thank you for your time and attention.

Sincerely,

A handwritten signature in black ink that reads "Theresa J. Lee". The signature is written in a cursive, flowing style.

Theresa J. Lee
Election Law Clinic
Harvard Law School
6 Everett Street, Suite 4105
Cambridge, MA 02138
thlee@law.harvard.edu



STATEMENT BY REQUESTER JUSTIN H. PHILLIPS

By signing below, I authorize the Election Law Clinic at Harvard Law School to obtain the requested records on my behalf.

By:  _____

Name: Justin H. Phillips

Date: 7/7/2022