

**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 “ $\epsilon$ ” (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  $\epsilon$  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