Harvard Data Science Review • Special Issue 2: Differential Privacy for the 2020 U.S. Census

Harnessing the Known Unknowns: Differential Privacy and the 2020 Census

Ruobin Gong[1] Erica L. Groshen[2,3] Salil Vadhan[4]

[1]Department of Statistics, Rutgers University, New Brunswick, New Jersey, United States of America,
[2]New York State School of Industrial and Labor Relations, Cornell University, Ithaca, New York, United States of America,
[3]William Erastus Upjohn Institute for Employment Research, William Erastus Upjohn Unemployment Trustee Corporation, Kalamazoo, Michigan, United States of America,
[4]Department of Computer Science, Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

Published on: Jun 24, 2022

DOI: https://doi.org/10.1162/99608f92.cb06b469

Introduction

This special issue, Differential Privacy for the 2020 U.S. Census: Can We Make Data Both Private and Useful?, provides an entry point to help data scientists across many disciplines adjust to a big change in a key component of our national data infrastructure. The United States Census Bureau is adopting formal differential privacy protections for public products from the 2020 U.S. Decennial Census. This is the first time that a country has released most of its subpopulation counts with formal privacy protections, although certainly not the first time that other official counts have been perturbed for the purpose of disclosure avoidance.

Population censuses are important. Indeed, they may be the oldest statistical products of communal societies. They are mentioned in the Bible (the book of Numbers) and required by Article I, Section 2 of the U.S. Constitution for allocating seats in Congress. After all, as Lord Kelvin noted in 1883:

> [W]hen you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind. (William Thomson, Lord Kelvin, Electrical Units of Measurement [1883])
>
> [Lord Kelvin's observation is often paraphrased to more zippy aphorisms such as, 'You cannot manage what you cannot measure.']

These days, each U.S. decennial census plays a role far beyond simply determining how many seats each state holds in Congress. Statistical frames based on Census Bureau counts underlie nearly all the demographic descriptions and many decisions made by government, business, or other organizations in the United States. Massive federal expenditures are distributed according to population estimates based on census data. Furthermore, a number of active and influential research communities depend upon decennial census data products.

Privacy protection for respondents is also important and getting more difficult to achieve. Such protection has long been required by law, in order to prevent harm and to encourage full and honest responses. Recently, though, growing uses of the decennial census, availability of other data sources, and increased computational firepower make protecting the privacy of census respondents more difficult.

Fortunately, newly developed formal privacy protection systems can both measure the degree of privacy protection and allow adequate transparency to inform statistical inference on protected data. Previous statistical methods used to protect privacy (such as suppression and swapping observations) lack both of these desirable properties.

Nevertheless, adopting a new form of privacy protection for such important data is far from easy. Some of the key challenges include implementation issues confronted by the Census Bureau, understanding analytical implications for data scientists, and managing communication so that all stakeholders can engage effectively with each other and inform the public about the implications of the change.

History of This *HDSR* Special Issue

After extensive research on the success of reconstruction attacks on the 2010 Decennial Census data and privacy protection systems and deliberation by its Data Stewardship Executive Policy Committee, the Census Bureau announced its intention to adopt formal differential privacy (DP) protections for the 2020 Decennial Census in 2018 (Abowd, 2018; U.S. Census Bureau, 2018). In 2019, the bureau began releasing files to help data users adjust and provide input to the changes. Two releases are of particular interest for this volume: (1) DP-protected data from the1940 Decennial Census to accompany unprotected microdata files that were already in the public domain, and (2) DP-protected demonstration files using 2010 Decennial Census data.

Recognizing the importance of this change for many branches of data science, John Eltinge and Erica L. Groshen (both *HDSR* co-editors at the time) and *HDSR* Editor-in-Chief Xiao-Li Meng decided to dedicate a session in the *HDSR* inaugural symposium to this topic. The goals were to sponsor research and produce an *HDSR* special issue that would identify implications of the new protections for research and provide a basis for constructive interdisciplinary dialogue among data users and with the Census Bureau. Three research teams agreed to independently analyze the protected and unprotected 1940 Census data in ways similar to what they expected to do with 2020 Census data. A diverse set of discussants also agreed to provide their views on the Census Bureau's approach and the results. And, the Census Bureau provided their experts to summarize their efforts to date and going forward. This very lively Differential Privacy for the 2020 Census session took place on October 25, 2019.

Since then, the Census Bureau's plans have evolved. In particular, the Bureau made numerous changes to optimize the performance of their DP algorithms and decided to inject less noise than done in the 1940 and 2010 demonstration data products. In the meantime, plans for this special issue also evolved. In particular, the issue has been enriched substantially by contributions after the conference. Two of the 1940 Census analytics teams found they were doing similar but complementary analyses, so they combined their efforts into a single paper that became Asquith et al. (2022). Some conference discussants chose to contribute papers rather than short discussions. *HDSR* published closely related papers that we felt merited republication in the print version of this special issue. The Census Bureau contributed three substantive background papers that will serve as key references on this topic going forward. We also solicited a couple of additional articles that add valuable analysis and insights to the issue.

Goals of This *HDSR* Special Issue

We aim for this special issue to help document, contextualize, and assess the Census Bureau's adoption of differential privacy and the debates around this decision. In doing so, the special issue will serve to inform stakeholders as they try to make use of the 2020 Census data products and understand the impact of differential privacy on such uses of the data products uses. This issue will also provide a reference point for future applications of differential privacy, both within and outside official statistics, recording challenges encountered and solutions found so that they do not need to be rediscovered anew. We also hope that the diverse readership and contributors to our special issue (and *HDSR* in general) will help to spread (both within and across disciplines) important insights and information about this change.

Concretely, the articles in this special issue address three central questions:

Why and how did the Census Bureau adopt differential privacy? The adoption of differential privacy needs to be understood within the history of statistical disclosure control for census data and the bureau's competing obligations to provide useful official statistics and ensure confidentiality for respondents. Moreover, differential privacy is not a single method to simply be deployed. Rather, it is a framework for measuring the privacy protections of a disclosure avoidance system (DAS), and there are many choices for how to implement a DAS while limiting the privacy loss under differential privacy. Thus, the Census Bureau engaged in a multiyear iterative process of releasing demonstration data products (such as the 1940 data that was the subject of the October 2019 *HDSR* session that was the seed for this special issue), soliciting feedback from stakeholders, and modifying the differentially private algorithms before settling on the final DAS used to release the first of the 2020 Census data products (namely, the P.L. 94-171 Redistricting Data Summary File).

Several articles in this special issue trace the Census Bureau's adoption of differential privacy from distinct perspectives. Three of these articles are written by experts at the Census Bureau, reflecting different aspects of the government's work. Abowd et al. (2022) present both the rationale for and the design of the differentially private TopDown Algorithm (TDA) used to produce the 2020 Census Redistricting Data (U.S. Census Bureau, 2021). Hawes (2020) discusses the challenges and lessons that the Bureau has come to learn through its process of implementing differential privacy. Eltinge (2022) deliberates the theoretical considerations in balancing privacy constraints and data quality from the perspective of official statistics agencies. Additionally, Sullivan (2020) underscores the pressing need to protect privacy as a defense to public trust and to the quality of the census.

Will the released differentially private data be fit for use? All of the methods for ensuring differential privacy involve introducing 'random noise' into the calculation of statistics so as to hide the contribution of any individual respondent. Differential privacy provides an accounting framework to measure and control the cumulative privacy loss incurred over all of the statistics calculated and published based on the census. Thus, differential privacy exposes two inherent tensions. One is the privacy-accuracy trade-off; providing a greater level of privacy requires introducing more noise, which leads to less accuracy. The other is the choice of which statistics (and, thus, which uses) to prioritize for accuracy; making some statistics more accurate requires making others less accurate in order to maintain the same level of privacy for respondents. Previous disclosure avoidance methods, such as data swapping, also had privacy-accuracy trade-offs and prioritized some queries over others, but those methods had far less transparency and opportunity for public input. Hence, it was and remains crucial to evaluate the fitness for use of data produced by the differentially private algorithms employed, both to inform the development of the DAS itself and to inform data users about data quality.

To this end, as invited speakers at the 2019 *HDSR* session, Asquith et al. (2022) and Brummet et al. (2022) present assessments of the quality of the DAS demonstration data in various use cases including survey sampling, federal funding allocation, and measurements of residential segregation. Cohen et al. (2022) investigate the impact of the TopDown Algorithm on the ability to accurately perform the analyses needed for elections and redistricting. One discussant at the conference, Heffetz (2022), highlights the multifaceted nature of the privacy-accuracy trade-off and calls for a public discourse of its social and ethical interpretations. A second discussant, Gong (2022), underscores the importance of transparency of privacy mechanisms for drawing reliable statistical inferences.

What was the debate about and how do we move forward? The Bureau's decision to adopt differential privacy was met by resistance, often heated, from several data-user communities. There are a range of factors that likely contributed to the tension. Some of the debate was the result of trade-offs that were not due to adopting differential privacy, but rather were exposed by it. As mentioned above, differential privacy makes explicit the trade-offs between privacy and accuracy, and between the accuracy for different statistics and use cases. Stakeholders who were previously accustomed to treating census data as if it were ground truth, with the disclosure avoidance and other sources of error hidden from view, are now forced to negotiate with the Bureau and each other for accuracy. Given the vast range of uses of census data (including redrawing voting districts for legislative elections, resource allocation, social science research, and policy and private sector analysis) and different priorities placed on these uses and on privacy by different stakeholders, satisfying everyone would be impossible. There is not a unique 'optimal' solution, but rather a range of possibilities that require a policy decision to select among. Thus, it is natural for academic research papers to reach different conclusions depending on what is being measured or evaluated; at the same time, we need to be cautious about political actors exploiting the legitimate discourse for their own gain.

In their capacity as co-chairs to the National Academies of Sciences, Engineering, and Medicine (NASEM) Committee of National Statistics (CNSTAT) Workshop and Expert Meetings on Census Data Quality, Hotz and Salvo (2022) chronicle the development of the Bureau's disclosure avoidance technology over the years, and the various challenges pertaining to the differential privacy revolution in 2020. Drawing from ethnographic fieldwork and theories from science and technology studies, boyd and Sarathy (2022) present a parallel story from the perspective of a variety of stakeholders. Oberski and Kreuter (2020) discuss the scholarly interactions between differential privacy technologies and social scientific insights. Groshen and Goroff (2022) provide essential information and recommendations to social scientists as they approach the analysis of privacy-protected 2020 Decennial Census data.

How the Articles Fit Together

Each of the articles published in this special issue comes from a particular standpoint and addresses at least one of the above questions, and often more than one. We note also that each article reflects a specific moment in time, as both the public discourse and Census DAS algorithm have evolved between the October 2019 *HDSR* session and the present, and some of the articles were completed long before the publication of this special issue.

Census: Importance, History, and Technical Change

This special issue opens with Teresa Sullivan's "Coming to Our Census: How Social Statistics Underpin Our Democracy (and Republic)" (2020). Sullivan is President Emerita and University Professor of Sociology of the University of Virginia, and her article is based on her American Statistical Association President's Invited Address delivered at the 2019 Joint Statistical Meeting. Sullivan regards the decennial census as the cornerstone to the social statistics infrastructure of the United States and underscores its crucial democratic functions to support the representation of populations and the allocation of resources. The article warns of the threat to the census data quality due to mistrust of the government, and highlights the importance of ensuring the confidentiality and security of the census data. Sullivan argues that the professional integrity of the statisticians is the best defense of the census. First published in *HDSR* Issue 2.1 (Winter 2020), Sullivan's article is accompanied by a set of nine invited discussions (Anderson, 2020; Belin, 2020; Chambers, 2020; Citro, 2020; Farley, 2020; Hogan, 2020; Kafadar, 2020; Poston, 2020; Trewin, 2020), including several international perspectives, and a rejoinder from the author [LINK to discussion/rejoinder online].

John Eltinge is Assistant Director for Research and Methodology at the United States Census Bureau, and co-editor of *HDSR*. Eltinge's article, "Disclosure Protection

in the Context of Statistical Agency Operations: Data Quality and Related Constraints" (2022), is an in-depth exploration of the competing considerations from the perspective of statistical agencies and their stakeholders in implementing disclosure limitations while delivering high-quality data products and services. His article asks questions on how disclosure protection changes the way we grapple with the multiple dimensions of data quality and fitness for use, in light of the legal, societal, and operational constraints. These dimensions of data quality include not only *accuracy*, but also *accessibility*, *relevance*, *granularity* (cross-sectional and temporal), *punctuality*, *coherence*, and *interpretability*.

Michael Hawes is Senior Advisor for Data Access and Privacy at the U.S. Census Bureau. First published in Issue 2.2 (Spring 2020), Hawes's article, "Implementing Differential Privacy: Seven Lessons From the 2020

United States Census" (2020), chronicles the challenges that emerge as a result of the Census Bureau's decision to modernize statistical disclosure control. Following the public policy debate around the privacy-accuracy trade-off, Hawes draws attention to the Bureau's efforts in balancing this trade-off while prioritizing a diverse range of use cases. The article highlights the need and the continued efforts by the Bureau to design and iterate the private algorithms specifically adapted to the census data products, while taking into account the various invariants and consistency requirements they must obey.

"The 2020 Census Disclosure Avoidance System TopDown Algorithm" (2022) is authored by the team who worked on the Bureau's implementation of differential privacy (John M. Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, Brett Moran, William Sexton, Matthew Spence, and Pavel Zhuravlev), known as the TopDown Algorithm. The article presents the mathematics, design, and testing of the algorithm, discusses the motivation for adopting differential privacy as its privacy-loss accounting framework, and summarizes the policy considerations that motivated specific aspects of the design of TDA. It focuses on the version of the algorithm that the Census Bureau used to release the 2020 Census Redistricting Data (P.L. 94-171) Summary File in August 2021. The article describes the tuning and testing of the TDA over time in order to optimize utility based on stakeholder feedback, but leaves for future work a discussion of how users can best analyze the released data.

Empirical Evaluations

The following articles focus mostly on fitness for use, but in so doing, they inform questions about how differential privacy was implemented and the debate over its impact. They analyze how the additional noise added by privacy protection might affect standard uses of 2020 Decennial Census data. The first two articles (Asquith et al. [2022] and Brummet et al. [2022]) are based on the 1940 Census data created by the team at IPUMS (Integrated Public Use Microdata Series) using the Census Bureau's software and the demonstration products released by the Bureau in 2019. (Importantly, readers should note that for both the 1940 and 2010 demonstration data products, the Census Bureau injected notably more noise than they subsequently injected into 2020 Decennial Census data products. In addition, these evaluations only assess the *apparent* fitness for use of privatized data, because the analyses were conducted on the 1940 demonstration data *as is*, without statistically accounting for the privacy protections that had been applied). The third article (Cohen et al., 2022) is based on the 2018 version of the TDA evaluated on reconstructed 2010 Census microdata. Based at universities (Tufts University, Boston University, and the Universities of Michigan, Chicago, and Minnesota) and research institutions (the Upjohn Institute for Employment Research and the National Opinion Research Center), the teams include quantitative methodologists such as statisticians, computer scientists, and mathematicians, and social scientists such as demographers, economists, sociologists, and geographers.

In "Assessing the Impact of Differential Privacy on Measures of Population and Racial Residential Segregation," (2022), the authors (Brian Asquith, Brad Hershbein, Tracy Kugler, Shane Reed, Steven Ruggles, Jonathan Schroeder, Steve Yesiltepe, and David Van Riper) examine the absolute and relative accuracy of population counts in total and by race for multiple geographic levels and compare commonly used measures of residential segregation. They show how the accuracy varies by the global privacy loss budget and by the allocation of the privacy loss budget to geographic levels and queries. The authors also demonstrate that the differentially private data can indicate either notably more or notably less segregation in particular areas than the untreated data.

In “The Effect of Differentially Private Noise Injection on Sampling Efficiency and Funding Allocations: Evidence From the 1940 Census” (2022), the authors (Quentin Brummet, Edward Mulrow, and Kirk Wolter) consider three separate uses of the decennial census data: (1) oversampling populations in surveys, (2) screening operations for surveys of rare populations, and (3) allocating federal funds to specific areas. They find that for use cases that involve large populations, the effects of noise injection are relatively modest, but for rare populations and small areas, sampling-frame coverage special issues and misallocations of funds can be severe.

In “Private Numbers in Public Policy: Census, Differential Privacy, and Redistricting” (2022), the authors (Aloni Cohen, Moon Duchin, JN Matthews, and Bhushan Suwal) consider applications of the decennial census data in redistricting, where the data is used to balance districts, to describe the demographic composition of districts, and to detect signals of racial polarization. Based on a close look at nine localities in Texas and Arizona, they find reassuring evidence that TopDown did not threaten these redistricting functions, relative to legal and practical standards already in play. They also compare the discrepancies introduced by TopDown to previously documented sources of error in census data.

Commentary and Critique

In his reflections, “What Will It Take to Get to Acceptable Privacy-Accuracy Combinations?” (2022) on the Asquith et al. and Brummet et al. studies (both in this special issue), Ori Heffetz, a professor at Hebrew University of Jerusalem and Hebrew University of Jerusalem and the Samuel Curtis Johnson Graduate School of Management at Cornell University, takes us on a deep dive into the meaning of the privacy-accuracy trade-off. Heffetz emphasizes the application-specific nature of accuracy standards, highlighting an entanglement between privacy error and other types of error, both within the privatized data set and in relation to other sources of data. On the other hand, Heffetz argues that the privacy standards as currently employed by the Census DAS are insufficient from the technical point of view of differential privacy, particularly as we anticipate further spending of the privacy loss budget on future data releases. Heffetz concludes by calling for a *societal infrastructure* that supports the normative discussion about an acceptable trade-off, viewing the trade-off through the lens of a cost-benefit analysis that takes into account the variety of social, economic, and ethical perspectives.

In "Transparent Privacy Is Principled Privacy" (2022), Ruobin Gong, a statistician at Rutgers University, discusses an important advantage brought forth by differential privacy: the transparency of the privacy mechanism. The probabilistic mechanism with which data is privatized can be made public without harming the privacy guarantee, and can be leveraged using statistical methodologies to obtain trustworthy inference from privatized data. Gong further argues that mandated invariants imposed on the privatized data may diminish transparency and result in limited statistical usability, and calls for the release of the pre-'postprocessed' census noisy measurements to support social and scientific research.

Broader Perspectives

A technology scholar at Microsoft Research and a doctoral student of computer science at Harvard University, danah boyd and Jayshree Sarathy contribute an article titled "Differential Perspectives: Epistemic Disconnects Surrounding the U.S. Census Bureau's Use of Differential Privacy" (2022). Drawing from the theories of science and technology studies (STS) and their ethnographic fieldwork, boyd and Sarathy present an account of the Census Bureau's decision to adopt differential privacy from the perspective of a variety of stakeholders. They provide context for the development, including the key events and decisions, of the 2020 Census Disclosure Avoidance System. Analyzing the controversy that arose, the authors tap deep into the epistemic, ideological, and political perspectives driving the communities with vested interests in the integrity and the quality of census data products.

The opinion piece "Differential Privacy and Social Science: An Urgent Puzzle" by Daniel L. Oberski and Frauke Kreuter (2020) was first published in Issue 2.1 (Winter 2020). Daniel Oberski is an Associate Professor in Data Science, Methodology, and Statistics at the University of Utrecht. Kreuter is Co-Director of the Social Data Science Center (SoDa) and faculty member in the Joint Program in Survey Methodology (JPSM) at the University of Maryland and Professor of Statistics and Data Science at the Ludwig Maximilian University of Munich, Germany. In their article, Oberski and Kreuter argue that "the discussion on implementing differential privacy has been clouded by incompatible subjective beliefs about risk, each perspective having merit for different data types." Second, they study both challenges and positive consequences for social science research if differential privacy is widely implemented. They conclude with a call for interdisciplinary collaboration to solve the urgent puzzle that differential privacy raises.

V. Joseph Hotz, Professor of Economics at Duke University, and Joseph J. Salvo, a demographer at the University of Virginia and the National Conference on Citizenship, present "A Chronicle of the Application of Differential Privacy to the 2020 Census" (2022). Hotz and Salvo trace the development of the Census Bureau's DAS from the early 20th century to the current census, and discuss the technical, legal, and pragmatic challenges along this path of evolution. In their roles as co-chairs to the NASEM CNSTAT 2020 Census Data Products Workshop and the subsequent Expert Group Meetings, Hotz and Salvo describe how the workshop's key findings from the assessment of the Bureau's initial 2010 Demonstration Data Files for a variety of use cases sparked rounds of iterations between the Bureau and the data-user communities, leading to accuracy improvements in the public data release. At the same time, the authors raise a number of issues that have become clearer since the CNSTAT Workshop. These topics concern the differences between the theoretical underpinning of the Bureau's DAS and the consequences of its application in practice. In light of the latter findings, Hotz and Salvo highlight the need for the Census Bureau to conduct a new round of simulated reconstruction and reidentification attacks on the 2020 Census products that it plans to release. They also call for a reexamination of the Census Bureau's legal obligation to protect the confidentiality of its respondents in light of the modern data technology landscape.

In "Disclosure Avoidance and the 2020 Census: What Do Researchers Need to Know?" (2022), Erica L. Groshen and Daniel L. Goroff address all three questions summarized in this foreword, as they provide essential information to social science researchers who will analyze the privacy-protected 2020 Decennial Census data. Groshen is a labor economist now at the Cornell University ILR School and formerly Commissioner of the United States Bureau of Labor Statistics. Goroff is an applied mathematician who is a program director for the Alfred P. Sloan Foundation. The authors highlight what is new about the 2020 Census's approach to disclosure avoidance, as well as what seems new but is actually little changed from recent censuses. They also examine strategies, trade-offs, and rationales associated with processing and releasing the decennial results. Finally, they offer specific conclusions to the Census Bureau and researchers to help promote appropriate and well-informed analysis of 2020 Census data.

Bottom Line

Of necessity, this special issue is an interim report. From the beginning, given the complexity of the transformation and the duration of the implementation, we did not expect to provide the final word on the use of differential privacy in the 2020 Census. Rather, we seek to help assemble a foundation that future research and policymaking can build on, both within and beyond the context of disclosure avoidance for census data. In that spirit, we offer some high-level lessons learned thus far.

1. Lest anyone think otherwise, modernizing disclosure avoidance protections for 2020 Census products is necessary and has far-reaching implications. This change will affect many stakeholders (including the public, the Census Bureau, researchers, and policymakers) in many ways. The change is not that disclosure avoidance distortions and other types of error are new for decennial census products. The change is that stakeholders are directly confronted by the accuracy-privacy trade-offs (rather than these being managed by the bureau behind the scenes) and related norms and practices for data analysis and communication, where many questions remain open. This has engendered far broader discussion of disclosure methods than ever before. While the Bureau has practiced disclosure limitation for decades, and disclosure experts have been doing research on them for decades, this is the first time that all who care about decennial census data are talking about what disclosure is, what it means, and what its implications are.
2. The process is far from over for the Census Bureau and data users. As of June 2022, when this foreword is being written, many planned data sets have not been finalized or released and use of the released data has

just begun. Using the new data appropriately will require analysts to rethink methods and sources used to analyze decennial censuses. Their experiences and requests will provide valuable feedback to the Census Bureau. The Census Bureau's plans for products and access will necessarily continue to adjust over the next few years, if not longer. Furthermore, even once 2020 products and processes are settled, new ones will need to be devised as part of advance planning for the 2030 Decennial Census, which is already underway.

3. The complexities of implementation, not to mention delays due to the COVID-19 pandemic, have required the Census Bureau to make many midstream adjustments. While privacy protections and noise injection are not new, this is the first instance of a statistical agency applying formal privacy protections to a population census. The Census Bureau has had to negotiate a variety of unexpected technical, legal, organizational, personnel, funding, and political challenges throughout the implementation process. As a consequence, preparation and keeping abreast of plans, progress, and implications have been difficult for observers.
4. Many stakeholders have a lot of work ahead. For example, analysts who rely directly and indirectly on decennial census products to make statistical inferences may need to reexamine their choices of data sources and methodologies and adjust them appropriately. In another realm, policymakers likely should reconsider the appropriateness of existing triggers, such as knife-edge program qualification criteria (for example, Housing and Urban Development Community Block Grants, Rural Business Development Grants, and the Rural Microentrepreneur Assistance Program), that will now rely on intentionally 'fuzzed' data. Furthermore, technical data users need to build more holistic frameworks for measuring various forms of uncertainty. Noise injected through DAS intersects other known and unknown sources of error in the data. Little is understood about their interactions, and much more research and evaluation are needed.
5. Future privacy protection efforts for official data will reflect lessons learned from this effort. This experience will help the Census Bureau and the other statistical agencies as they decide if and how to extend formal privacy protections to other data series in the years to come. Yet, observers should avoid leaping too quickly to conclusions about likely impact on any particular data set, because the exact implementation of differential privacy protections must be specific to the nature of the data involved. Furthermore, litigation or legislation on some of the topics raised here may be inevitable and could also affect future implementation.
6. Communication across perspectives and disciplines is essential, challenging, and needs to start early. For some stakeholders, this change is long overdue, even as others feel blindsided and question the science and the urgency. More communication among the various communities could help convert these gaping cross-disciplinary disconnects into opportunities for fruitful collaboration. Indeed, these exchanges should be underway well before the changes impact people's work.

We hope that readers will find this special issue useful as they prepare for changes ahead and for orienting themselves in the constellation of stakeholders involved in the implementation of formal privacy for the 2020 Decennial Census. Given the political implications of the data, many public exchanges about the 2020 Decennial Census data have been highly contentious. By contrast, we are heartened by the constructive discourse represented in this special issue, where the authors are driven by a shared goal of finding the best balance between privacy and data usability, even as they may reach different conclusions. Thus, we hope that this special issue will serve as a model and starting point for bridging communication gaps and encouraging future collaboration.

Acknowledgments

The special issue editors thank the many authors and anonymous reviewers of the contributions in this special issue for their hard work to advance understanding of this challenging subject matter. Many people at the Census Bureau, including John Abowd, John Eltinge, Michael Hawes, and Simson Garfinkel were generous with fact-checking and insights. Founding *HDSR* Editor-in-Chief Xiao-Li Meng expertly helped guide the effort throughout and Rebecca McLeod led the production process masterfully. All views expressed herein as well as errors and deficiencies are our own.

Disclosure Statement

Harvard Data Science Review and special issue editors gratefully acknowledge support from the Alfred P. Sloan Foundation for production of the hardcopy version of this special issue. Ruobin Gong and Erica L. Groshen have no other financial or nonfinancial disclosures to share for this editorial. Salil Vadhan's work on this special issue was supported by Cooperative Agreements CB16ADR0160001 and CB20ADR0160001 with the Census Bureau and a Simons Investigator Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.

References

Abowd, J. M. (2018). The U.S. Census Bureau adopts differential privacy. In Y. Guo & F. Farooq (Eds.), *KDD '18 Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (p. 2867). ACM. https://doi.org/10.1145/3219819.3226070

Abowd, J., Ashmead, R., Cumings-Menon, R., Garfinkel, S., Heineck, M., Heiss, C., Johns, R., Kifer, D., Leclerc, P., Machanavajjhala, A., Moran, B., Sexton, W., Spence, M., & Zhuravlev, P. (2022). The 2020 Census disclosure avoidance system TopDown Algorithm. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.529e3cb9

Anderson, M. J. (2020). Historical lessons: How statistics underpin our democracy. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.490ac65e

Asquith, B., Hershbein, B., Kugler, T., Reed, S., Ruggles, S., Schroeder, J., Yesiltepe, S., & Van Riper, D. (2022). Assessing the impact of differential privacy on measures of population and racial residential segregation. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.5cd8024e

Belin, T. R. (2020). The U.S. Census as a crucible for trust in government. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.bac26d97

boyd, d., & Sarathy, J. (2022). Differential perspectives: Epistemic disconnects surrounding the U.S. Census Bureau's use of differential privacy. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.66882f0e

Brummet, Q., Mulrow, E., & Wolter, K. (2022). The effect of differentially private noise injection on sampling efficiency and funding allocations: Evidence from the 1940 Census. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.a93d96fa

Chambers, R. (2020). Should the Census have more spine? *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.b0966dd6

Citro, C. F. (2020). Are we up to the challenges of protecting federal statistics? *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.1a3cd97f

Cohen, A., Duchin, M., Matthews, J., & Suwal, B. (2022). Private numbers in public policy: Census, differential privacy, and redistricting. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.22fd8a0e

Cohen, A., Duchin, M., Matthews, J., & Suwal, B. (2022). Private numbers in public policy: Census, differential privacy, and redistricting. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.22fd8a0e

Eltinge, J. L. (2022). Disclosure protection in the context of statistical agency operations: Data quality and related constraints. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.1cfad278

Farley, R. (2020). The importance of Census 2020 and the challenges of getting a complete count. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.8a0cc85c

Gong, R. (2022). Transparent privacy is principled privacy. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.b5d3faaa

Groshen, E. L., & Goroff, D. (2022). Disclosure avoidance and the 2020 Census: What do researchers need to know? *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.aed7f34f

Hawes, M. B. (2020). Implementing differential privacy: Seven lessons from the 2020 United States Census. *Harvard Data Science Review*, 2(2). https://doi.org/10.1162/99608f92.353c6f99

Heffetz, O. (2022). What will it take to get to acceptable privacy-accuracy combinations? *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.5d9b1a8d

Hogan, H. (2020). Distrust in the governments brings risk to the census. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.11f3e977

Hotz, V. J., & Salvo, J. (2022). A chronicle of the application of differential privacy to the 2020 Census. *Harvard Data Science Review*, (Special Issue 3). https://doi.org/10.1162/99608f92.ff891fe5

Kafadar, K. (2020). statisticians' role in ensuring accuracy and integrity of federal data. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.aeb6e98b

Oberski, D. L., & Kreuter, F. (2020). Differential privacy and social science: An urgent puzzle. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.63a22079

Poston, Jr., D. L. (2020). The Decennial Census and congressional apportionment. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.2b99e39a

Sullivan, T. A. (2020). "Coming to our census: How social statistics underpin our democracy (and republic)": Author's Response. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.addb8baf

Sullivan, T. A. (2020). Coming to our census: How social statistics underpin our democracy (and republic). *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.c871f9e0

Thomson, W. (1883). Electrical units of measurement. *Popular lectures and addresses*, *1*(73).

Trewin, D. (2020). An Australian perspective on Teresa Sullivan's "Coming to our census." *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.53fef7af

U.S. Census Bureau. (2018). *2020 Census program management review*. https://www2.census.gov/programs-surveys/decennial/2020/program-management/pmr-materials/2018-08-03/pmr-all-materials-2018-08-03.pdf

U.S. Census Bureau. (2021). *2020 Census Redistricting Data (Public Law 94-171) Summary File United States*. machine-readable data files/prepared by the U.S. Census Bureau. https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html

U.S. Const. art. I, §2.
