Exhibit 1

FAIR LINES AMERICA FOUNDATION, INC., 2308 Mount Vernon Ave. Ste. 716 Alexandria, VA 22301-1328 Plaintiff, v. UNITED STATES DEPARTMENT OF COMMERCE 1401 Constitution Avenue, NW Washington, D.C. 20230; and UNITED STATES BUREAU OF THE CENSUS, 4600 Silver Hill Road Washington, D.C. 20233-3700, Defendants.	)	Civ. A. No. 1:21-cv-01361 (ABJ)

DECLARATION OF JOHN M. ABOWD

I, John M. Abowd, make the following Declaration pursuant to 28 U.S.C. § 1746, and declare that under penalty of perjury the following is true and correct to the best of my knowledge:

In this Declaration I:

- Provide background on how the Census Bureau applies the confidentiality provisions set forth in 13 U.S.C. §§ 8(b) & 9 for the 2020 Census, including the use of differential privacy;
- Explain the importance of maintaining the redactions made by the Disclosure Review Board (DRB) to portions of the 988 pages of records produced in this lawsuit on May 25, 2021, and the additional 23 pages produced on July 6, 2021; and
- Explain why Title 13 confidentiality prohibits the Census Bureau from producing to Plaintiff the state-by-state numbers representing individuals added to the 2020 Census totals through a process known as Group Quarters Count Imputation (GQCI).
- Explain the harm (e.g., delay, impact on DAS, etc.) on the Census Bureau's efforts to protect the confidentiality of the information it collects if disclosure of the information withheld in this litigation is required.

BACKGROUND

1. I am the Chief Scientist and Associate Director for Research and Methodology at the United States Census Bureau. I have served in this capacity since June 2016. My statements in this declaration are based on my personal knowledge or on information supplied to me in the course of my professional responsibilities.
2. I received my Ph.D. in economics from the University of Chicago with specializations in econometrics and labor economics in 1977 (M.A. 1976). My B.A. in economics is from the University of Notre Dame.

3. I have been a university professor since 1976 when I was appointed assistant professor of economics at Princeton University. I was also assistant and associate professor of econometrics and industrial relations at the University of Chicago Graduate School of Business. In 1987, I was appointed associate professor of industrial and labor relations with indefinite tenure at Cornell University where I am currently the Edmund Ezra Day Professor (emeritus). My current position at the Census Bureau is part of the Career Senior Executive Service.
4. I am a member and fellow of the American Association for the Advancement of Science, American Statistical Association, Econometric Society, and Society of Labor Economists (president 2014). I am an elected member of the International Statistical Institute. I am also a member of the American Economic Association, International Association for Official Statistics, National Association for Business Economists, American Association for Public Opinion Research, Association for Computing Machinery, and American Association of Wine Economists. I regularly attend and present papers at the meetings of these organizations.
5. I have served on the American Economic Association Committee on Economic Statistics. I have also served on the National Academy of Sciences Committee on National Statistics, the Conference on Research in Income and Wealth Executive Committee, and the Bureau of Labor Statistics Technical Advisory Board for the National Longitudinal Surveys (chair: 1999-2001).
6. I have worked with the Census Bureau since 1998, when the Census Bureau and Cornell University entered into the first of a sequence of Intergovernmental Personnel Act agreements and other contracts. Under those agreements, I served continuously as Distinguished Senior Research Fellow at the Census Bureau until I assumed my current position as Chief Scientist in 2016, under a new Intergovernmental Personnel Act contract. Since March 29, 2020, I have been in the Associate Director position at the Census Bureau as a Career Senior Executive Service employee.

7. From 2011 until I assumed my position as Chief Scientist at the Census Bureau in 2016, I was the lead Principal Investigator of the Cornell University node of the NSF-Census Research Network, one of eight such nodes that worked collaboratively with the Census Bureau and other federal statistical agencies to identify important theoretical and applied research projects of direct programmatic importance to the agencies. The Cornell node produced the fundamental science explaining the distinct roles of statistical policymakers and computer scientists in the design and implementation of differential privacy systems at statistical agencies.

8. I have published more than 100 scholarly books, monographs, and articles in the disciplines of economics, econometrics, statistics, computer science, and information science. I have been the principal investigator or co-principal investigator on 35 sponsored research projects. I was a founding editor of the Journal of Privacy and Confidentiality—an interdisciplinary journal, and I continue to serve as an editor and on the governance board. My full professional resume is attached to this report as Appendix A.

9. I have worked on and managed Census Bureau projects that were precursors to the Census Bureau's current program to implement differential privacy for the 2020 Census of Population and Housing. I was one of three senior researchers who founded the Longitudinal Employer-Household Dynamics (LEHD) program at the Census Bureau, which is generally acknowledged as the Census Bureau's first 21st Century data product: built to the specifications of local labor market specialists without additional survey burden, and published beginning in 2001 using state-of-the-art confidentiality protection via noise infusion. This program produces detailed public-use statistical data on the characteristics of workers and employers in local labor markets using large-scale linked administrative, census, and survey data from many different sources. In 2008, my work with LEHD led to the first production implementation worldwide of differential privacy as part of a product of the LEHD program called

OnTheMap. The LEHD program also implemented other prototype systems to protect confidential information, including allowing the public to access synthetic micro-data confirmed via direct analysis of the confidential data on validation servers. A differentially private version of this system is under development at the Census Bureau but not for use with the 2020 Census.

IMPORTANCE OF CONFIDENTIALITY

10. Though participation in the census is mandatory under 13 U.S. Code § 221, in practice, the Census Bureau must rely on the voluntary participation of each household in order to conduct a complete enumeration.
11. One of the most significant barriers to conducting a complete and accurate enumeration are individuals' concerns about the confidentiality of census data. The Census Bureau's pre-2020 Census research showed that 28% of respondents were "extremely concerned" or "very concerned" and a further 25% were "somewhat concerned" about the confidentiality of their census responses.[1] These concerns are even more pronounced in minority populations and represent a major operational challenge to enumerating traditionally hard-to-count populations.[2]
12. To secure voluntary participation, Congress first established confidentiality protections for individual census responses in the Census Act of 1879. These confidentiality protections were later expanded and codified in 13 U.S. Code §§ 8(b) & 9, which prohibits the Census Bureau from releasing "any publication whereby the data furnished by any particular establishment or individual under this title can be identified[,]" and

---

[1] U.S. Census Bureau (2019) "2020 Census Barriers, Attitudes, and Motivators Study Survey Report" https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-study-survey.pdf, p. 38-39.

[2] Ibid, p.39-42.

allows the Secretary to provide aggregate statistics so long as those data "do not disclose the information reported by, or on behalf of, any particular respondent[.]" Title III of the Foundations for Evidence Based Policymaking Act of 2018 also requires statistical agencies to "protect the trust of information providers by ensuring the confidentiality and exclusive statistical use of their responses."[3]

13. The broader scientific community generally concurs about the importance of rigorous protection of confidentiality by statistical agencies. For example, the National Academy of Sciences' definitive guidebook for federal statistical agencies states "Because virtually every person, household, business, state or local government, and organization is the subject of some federal statistics, public trust is essential for the continued effectiveness of federal statistical agencies. Individuals and entities providing data directly or indirectly to federal statistical agencies must trust that the agencies will appropriately handle and protect their information."[4] The report also notes that respondents expect statistical agencies not to "release or publish their information in identifiable form."[5] The National Academies also broadly exhort statistical agencies to "continually seek to improve and innovate their processes, methods, and statistical products to better measure an ever-changing world."[6]

14. The Census Bureau enjoys higher self-response rates than private survey companies in large part because the public generally trusts the Census Bureau to keep its data safe. The Census Bureau makes extensive outreach efforts to assure respondents and

---

[3] Title III of the Foundations for Evidence Based Policymaking Act of 2018, § 3563.

[4] National Academies of Sciences, Engineering, and Medicine 2021. Principles and Practices for a Federal Statistical Agency: Seventh Edition. Washington, DC: The National Academies Press. https://doi.org/10.17226/25885, p. 37-38.

[5] Ibid., p. 38.

[6] Ibid., p. 4.

other data providers about the Bureau's commitment to protection of confidential data. The criminal fines and imprisonment penalties that Census Bureau employees would face by unlawfully disclosing respondent information are frequently cited by the Census Bureau in these outreach efforts.[7]

15. This trust in the Census Bureau is particularly important for the decennial census, given the "civic ceremony" aspect of the census, akin to the civic ceremony aspect of elections and voting. The decennial census is an exercise where the nation comes together every ten years, under a strict promise of confidentiality, to provide information to help govern our nation. Were the Census Bureau to expose confidential information, there is no doubt that self-response rates would drop, increasing survey cost across programs by increasing in-person follow up, and decreasing the quality of the census overall.

PRIVACY AND CONFIDENTIALITY PROTECTION AT THE CENSUS BUREAU[8]

16. Protecting the privacy of our respondents and the confidentiality of their data is at the core of the Census Bureau's mission. Our privacy promise to respondents is key to promoting response to our censuses and surveys.

---

[7] https://www.census.gov/content/dam/Census/library/factsheets/2019/comm/2020-confidentiality-factsheet.pdf.

[8] The terms "privacy" and "confidentiality" are related, but technically distinct. Generally speaking, protecting privacy entails adherence to the full suite of Fair Information Practice Principles, and includes elements of collection and use limitation, purpose specification, and openness, among others. Confidentiality protection, more specifically, is a component of protecting privacy, and typically refers to the protection of data against unauthorized disclosure, access, or use. In the statistical and technical communities, however, "privacy protection" often refers specifically to the various statistical disclosure limitation methods used to protect the confidentiality of individuals' data. It is this latter conception of privacy protection, specifically statistical safeguards against disclosure, that I will be using throughout this declaration when using the generic term "privacy." And it is this conception of privacy protection which, for the Census Bureau, includes the

17. Data collected from the decennial census support a wide array of critical government and societal functions at the federal, state, tribal, and local levels. In addition to apportioning seats in the U.S. House of Representatives and supporting the redistricting of those seats, census data also support the allocation of over $675 billion in federal funding each year based on population counts, geography, and demographic characteristics.[9] Census data also support important public and private sector decision-making at the federal, state, tribal, and local levels, and serve as benchmark statistics for other important surveys and data collections throughout the decade.[10]

18. The Census Bureau publishes an enormous number of statistics calculated from its collected data. Following the 2010 Census, for example, the Census Bureau published over 150 billion independent statistics about the characteristics of the 308,745,538 persons in the resident population that were enumerated in the census. To serve their intended governmental and societal uses, the majority of these statistics needed to be published at very fine levels of detail and with geographic precision often down to the individual census tract or block.

19. While it would be quite difficult to ascertain from any single one of those published statistics the identity of any individual census respondent or the contents of that respondent's census response, the volume and detail of information published by the

---

methods the Bureau implements to protect the confidentiality of the census data covered by the confidentiality provisions of 13 U.S.C. §§ 8(b) and 9.

[9] Hotchkiss, M., & Phelan, J. (2017). Uses of Census Bureau data in federal funds distribution: A new design for the 21st century. United States Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/working-papers/Uses-of-Census-Bureau-Data-in-Federal-Funds-Distribution.pdf.

[10] 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.

Census Bureau, taken together, pose a serious challenge for protecting the privacy and confidentiality of census data. Combining information from multiple published statistics or tables can make it easy to pick out those individuals in a particular geographic area whose characteristics differ from those of the rest of their neighbors. These individuals, who have unique combinations of the demographic characteristics reported in statistical summaries, are known as "population uniques" and their records have traditionally been the target of the mechanisms that the Census Bureau uses to protect confidentiality in its data publications.

20. Traditional statistical disclosure limitation methods,[11] like those used in the 2010 Census, cannot defend against modern challenges posed by enormous cloud computing capacity and sophisticated software libraries. That does not mean traditional statistical disclosure limitation methods usually fail—they usually do not fail. But as computer scientists bring their expertise from the field of cryptography to the field of safe data publication, they have exposed significant vulnerabilities in traditional confidentiality protection methods. [12] The Census Bureau's own internal analysis, for example, confirmed that a modern database reconstruction-abetted re-identification attack can reliably match a large number of 2010 Census responses to the names of those respondents—a vulnerability that exposed information of *at least* 52 million Americans

[11] The technical field that addresses confidentiality is known as "statistical disclosure limitation." At the Census Bureau, it is known as "disclosure avoidance." It is also called "statistical disclosure control" by some statisticians and "privacy-preserving data analysis" by some computer scientists.

[12] "Official Statistics at the Crossroads: Data Quality and Access in an Era of Heightened Privacy Risk," *The Survey Statistician*, 2021, Vol. 83, 23-26 (available at Survey_Statistician_2021_January_N83_03.pdf (isi-iass.org)). The paper is based on talks that I gave in 2019 to the Committee on National Statistics and the Joint Statistical Meetings. It summarizes the research in Abowd, J.M. and I. Schmutte "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review,* Vol. 109, No. 1 (January 2019):171-202, DOI:10.1257/aer.20170627.

and potentially up to 179 million Americans.[13] To defend against this known vulnerability, the Census Bureau explored different confidentiality methods that explicitly defend against database reconstruction attacks and concluded that the best tool to protect against this modern attack while also preserving the accuracy and usability of data products comes from the body of scientific work called "differential privacy."

THE HISTORY OF INNOVATION IN THE DECENNIAL CENSUS

21. The decennial census, known officially as the *Decennial Census of Population and Housing,* is the flagship statistical product of the U.S. Census Bureau. Though the Census Bureau conducts hundreds of surveys every year, the once-every-decade enumeration of the population of the United States, mandated by Article I, Section 2 of the U.S. Constitution, is the single largest and most complex data collection regularly conducted by the United States government. Since the very first U.S. census in 1790, the collection, processing, and dissemination of census data have posed unique challenges and have required the Census Bureau to improve its operations every decade.
22. The challenges faced by the Census Bureau have led to remarkable innovations. Herman Hollerith's electric tabulation machine, developed for the 1890 Census, revolutionized the field of data processing and led Hollerith to form the company that eventually became IBM.[14] To conduct the 1950 Census, the Census Bureau commissioned the development of the first successful civilian digital computer, UNIVAC I.[15]

---

[13] See Appendix B for a summary of the Census Bureau's simulated reconstruction and re-identification attacks.

[14] https://www.census.gov/history/www/census_then_now/notable_alumni/herman_hollerith.html.

[15] https://www.census.gov/history/www/innovations/technology/univac_i.html.

With each passing decade, the Census Bureau develops, tests, and deploys innovations to its statistical methods, field data collection methods, and data processing operations.

23. That spirit of innovation includes the Census Bureau's more recent implementation of modern privacy protections. Prior to the 1990 Census, the primary mechanism that the Census Bureau employed to protect the confidentiality of individual census responses was to withhold publication of (or "suppress") any table that did not meet certain household, population, or demographic characteristic thresholds. The 1970 Census, for example, suppressed tables reflecting fewer than five households, and would only publish tables of demographic characteristics cross-tabulated by race if there were at least five individuals in each reported race category.[16] These suppression routines helped to protect privacy by reducing the detail of data published about individuals who were relatively unique within their communities. By the 1990 Census, however, the Census Bureau transitioned away from suppression methodologies for two reasons: first, data users were dissatisfied with missing details caused by suppression and second, the Bureau realized that the suppression routines it had been using were insufficient to fully protect against re-identification.[17]

---

[16] Zeisset, P. (1978), "Suppression vs. Random Rounding: Disclosure Avoidance Alternatives for the 1980 Census," https://www.census.gov/content/dam/Census/library/working-papers/1978/adrm/Suppression%20vs.%20Random%20Rounding%20Disclosure-Avoidance%20Alternatives%20for%20the%201980%20Census.pdf.

[17] McKenna, L. (2018), "Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing," https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf, p. 6.

24. For the 1990 Census, the Bureau began using a technique known as noise infusion to safeguard respondent confidentiality. Noise infusion helps to protect the confidentiality of published data by introducing controlled amounts of error or "noise" into the data. The goal of noise infusion is to preserve the overall statistical validity of the resulting data while introducing enough uncertainty that attackers would not have any reasonable degree of certainty that they had isolated data for any particular respondent. The noise infusion used in 1990 was a very simple form of data swapping between paired households in a geographic area with similar attributes, and for small block groups the Census Bureau replaced the collected characteristics of households with imputed characteristics.[18]

25. For the 2000 and 2010 censuses, the Census Bureau began to infuse noise using a more advanced "data swapping" method. The Census Bureau first identified households most vulnerable to re-identification—especially households on smaller-population blocks whose residents had differing demographic characteristics from the remainder of their block. While every non-imputed[19] household record in the Census Edited File (CEF) had a chance of being selected for data swapping, records for more vulnerable households (typically those on low-population blocks) were selected with greater

[18] Ibid., p. 6-7. An "imputed characteristic" is the prediction of a statistical model used in place of a missing characteristic, when used in standard editing procedures, or in place of a collected characteristic, when used for confidentiality protection.

[19] When a respondent household provides only a count of the number of persons living at that address or when the housing unit population count is itself imputed, the Census Bureau imputes all characteristics: sex, age, race, ethnicity, and relationship to others in the household. Such persons are called "whole-person census imputations" in technical documentation. When a household consists entirely of whole-person census imputation records, it is called an "imputed" household. A "non-imputed" household contains at least one person whose characteristics were collected on the census form for the household.

probability. Then, the records for all members of those selected households were exchanged with the records of households in nearby geographic areas that matched on key characteristics. For the 2000 and 2010 censuses, those key matching characteristics were (1) the whole number of persons in the household, and (2) the whole number of persons aged 18 or older in the household. These swapping criteria resulted in the total population and total voting age population for each block being held "invariant"—that is, while noise was added to all remaining characteristics, no noise was added to the block-level total population or block-level voting age population counts.[20] The selection and application of these particular invariants is not an innate feature of data swapping; invariants are implementation parameters that can be applied to (or removed from) any counted characteristic under any noise infusion methodology.

THE DISCLOSURE AVOIDANCE METHODS USED FOR THE 2010 CENSUS ARE NO LONGER SUFFICIENT

26. While the Census Bureau's disclosure avoidance methodologies for the 2000 and 2010 censuses were considered sufficient at the time, advances in technology in the years since have reduced the confidentiality protection provided by data swapping.
27. Disclosure avoidance has been a recognized branch of statistics since the 1970s, but it has only been since the late 1990s that it has evolved into a distinct scientific field of study in both statistics and computer science. Prof. Latanya Sweeney's 1997 revelation that she had re-identified then Massachusetts Governor William Weld's medical records in a purportedly "deidentified" public database[21] prompted the Census Bureau

[20] Ibid. p. 8-10.

[21] Sweeney, L. (2002). "k-anonymity: a model for protecting privacy." International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5); 557-570, also recounted in Ohm, P. (2009) "Broken promises of privacy: Responding to the surprising failure of anonymization." UCLA l. Rev. 57: 1701.

and many other statistical agencies to re-examine the efficacy of their disclosure avoidance techniques.

28. *Re-identification attacks.* Prior to 2016, disclosure risk assessments usually focused on assessing the vulnerability of microdata releases (data products that contain individual records for all or some of the data subjects deidentified by removing names and addresses), rather than the rules used for aggregated data releases (data compiled and aggregated into tables). Simulated "re-identification attacks" analyze the risk that an external attacker could use individuals' characteristics that are included on a published microdata file (e.g., location, age, and sex) and link those records to a third-party data source (e.g., commercial data or voter registration lists) that contains those characteristics along with the individuals' names and addresses. The resulting rates of "putative" (suspected) and confirmed linkages show the overall degree of vulnerability of the data. If those linkage rates are deemed too large, then additional disclosure avoidance is necessary to mitigate the disclosure risk.

29. The general problem with relying exclusively on re-identification studies to assess disclosure risk is that they can only provide a "best-case" approximation of the underlying disclosure risk of the data. If a real attacker has access to more sophisticated tools (e.g., optimization algorithms or computing power) or to higher quality external data (e.g., with better age and address information) than the tools or data used in the simulated attack, then the real disclosure risk will be substantially higher than what is estimated via the study. This limitation is particularly vexing for statistical agencies that must rely on a "release and forget" approach to data publication, where disclosure avoidance safeguards must be selected without foreknowledge of the better tools and external data that attackers may have at their disposal after the data are published.

30. Re-identification studies also underestimate the risk from releasing aggregated data. The Census Bureau has long relied on re-identification studies to assess the disclosure

risk of its microdata releases, but the majority of Census Bureau data products are aggregated data releases. Over the past decade, aggregated data releases have become increasingly vulnerable to sophisticated "reconstruction attacks" that have emerged as computing power has improved and gotten substantially cheaper.

31. *Reconstruction attacks.* The theory behind a "reconstruction attack" is that the release of *any* statistic calculated from a confidential data source will reveal a potentially trivial, but non-zero, amount of confidential information.[22] As a consequence, if an attacker has access to enough aggregated data with sufficient detail and precision, then the attacker may be able to leverage information from each statistic in the aggregated data to reconstruct the individual-level records that were used to generate the published tables. This process is known as a "reconstruction attack," and it adds a new degree of disclosure vulnerability against which statistical agencies must defend. While the statistical and computer science communities have been aware of this vulnerability since 2003, only over the last few years have computing power and the sophisticated numerical optimization software necessary to perform these types of reconstructions advanced enough to permit reconstruction attacks at any significant scale.

32. The risk of reconstruction and re-identification attacks is real and substantiated. The Census Bureau has been approached by Prof. Sweeney and others who claim that they have identified specific vulnerabilities in our standard disclosure avoidance methodologies.[23] The vulnerabilities in the disclosure avoidance protections for the Census Bureau's Survey of Income and Program Participation (SIPP) identified by Prof.

---

[22] Dinur, I. and Nissim, K. (2003) "Revealing Information while Preserving Privacy" PODS, June 9-12, San Diego, CA. https://doi.org/10.1145/773153.773173.

[23] McKenna, L. (2019b). "U.S. Census Bureau Reidentification Studies," available at https://www.census.gov/library/working-papers/2019/adrm/2019-04-ReidentificationStudies.html.

Sweeney led the Census Bureau to immediately implement permanent changes to the disclosure avoidance rules used for SIPP data, including increased noise infusion and delayed reporting of survey participants' major life events.[24]

33. Statistical releases do not all need to be of the same type, or contain the same data fields, to enable re-identification by reconstruction. For example, a 2015 interagency report published by the National Institute of Standards and Technology (NIST) written by my colleague Simson Garfinkel provided examples of using disparate data sets to reconstruct hidden underlying data.[25] Some of these examples are quoted here:

34. "*The Netflix Prize:* Narayanan and Shmatikov showed in 2008 that in many cases the set of movies that a person had watched could be used as an identifier.[26] Netflix had released a dataset of movies that some of its customers had watched and ranked as part of its "Netflix Prize" competition. Although there was [sic] no direct identifiers in the dataset, the researchers showed that a set of movies watched (especially less popular films, such as cult classics and foreign films) could frequently be used to match a user profile from the Netflix dataset to a single user profile in the Internet Movie Data Base (IMDB), which had not been de-identified and included user names, many of which were real names. The threat scenario is that by rating a few movies on IMDB, a person might inadvertently reveal *all* of the movies that they had watched, since the person's IMDB profile could be linked with the Netflix Prize data."[27] (emphasis in original)

---

[24] McKenna, L. (2019b). p. 2-3.

[25] Garfinkel, S. (2015) "De-Identification of Personal Information," National Institute of Standards and Technology, available at http://dx.doi.org/10.6028/NIST.IR.8053 at 26-27.

[26] Narayanan, A. and Shmatikov V. "Robust De-anonymization of Large Sparse Datasets," *IEEE Symposium on Security and Privacy* (2008): 111-125.

[27] Garfinkel, S. (2015), p. 26-27.

35. "*Credit Card Transactions:* Working with a collection of de-identified credit card transactions from a sample of 1.1 million people from an unnamed country, Montjoye *et al.* showed that four distinct points in space and time were sufficient to specify uniquely 90% of the individuals in their sample.[28] Lowering the geographical resolution and binning transaction values (*e.g.,* reporting a purchase of $14.86 as between $10.00 and $19.99) increased the number of points required."[29]

36. "*Mobility Traces:* Montjoye *et al.* showed that people and vehicles could be identified by their "mobility traces" (a record of locations and times that the person or vehicle visited). In their study, trace data from a sample of 1.5 million individuals was processed, with time values being generalized to the hour and spatial data generalized to the resolution provided by a cell phone system (typically 10-20 city blocks).[30] The researchers found that four randomly chosen observations of an individual putting them at a specific place and time was sufficient to uniquely identify 95% of the data subjects.[31] Space/time points for individuals can be collected from a variety of sources, including purchases with a credit card, a photograph, or Internet usage. A similar study performed by Ma *et al.* found that 30%-50% of individuals could be identified with 10 pieces of side information.[32] The threat scenario is that a person who revealed five place/time pairs (perhaps by sending email from work and home at four

---

[28] Montjoye, Y-A. et al. "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, 30 (January 2015) Vol 347, Issue 6221.

[29] Garfinkel, S. (2015), p. 27.

[30] De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. Scientific reports, 3(1).

[31] Ibid., p. 1-5.

[32] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip and N. S. V. Rao (2013) "Privacy Vulnerability of Published Anonymous Mobility Traces," in IEEE/ACM Transactions on Networking, vol. 21, no. 3, pp. 720-733, June 2013, doi: 10.1109/TNET.2012.2208983.

times over the course of a month) would make it possible for an attacker to identify his or her entire mobility trace in a publicly released dataset. As above, the attacker would need to know that the target was in the data."[33]

37. The same general principles apply to census data. The difference between census data and the examples above is that census data can be combined in vastly more ways with other information because all the tables published from census data share basic standardized identifiers including location, age, sex, race, ethnicity, and marital status. Even if each of these identifiers is not included in every table, their use and combinations across many different tables creates the disclosure risk. The Census Bureau understood this emerging risk even before the 2010 Census. As field collection for the 2010 Census was first beginning, the Census Bureau had already flagged the heightened disclosure risk of releasing detailed block level population data, even with the 2010 Census swapping mechanism in place.

38. For example, during a January 2010 meeting of the Census Bureau's Data Stewardship Executive Policy (DSEP) Committee, the chair of the Disclosure Review Board voiced her concerns about the 2010 Census swapping mechanism's ability to adequately protect future censuses, noting specifically the challenge posed by "continuing to release data at the block level, as block populations continue to decrease (e.g., 40% of blocks in North Dakota have only 1 household in them)."[34] Based on this warning, DSEP decided that "the problem of block population size and disclosure avoidance is real,

---

[33] Garfinkel, S. (2015), p. 27-28.

[34] The Data Stewardship Executive Policy Committee is chaired by the Deputy Director/Chief Operating Officer and composed of career senior executives with expertise in confidentiality practice, the uses of Census Bureau data, and policy. The Committee is the parent organization for the Disclosure Review Board (DRB), which reviews and approves individual data releases to ensure that no confidential data is released.

and that it deserves attention in the context of 2020 planning." DSEP Meeting Record, January 14, 2010. See Appendix C.

39. After tracking this growing risk of reconstruction and re-identification attacks for several years, the Census Bureau decided in 2015 to establish a new team to comprehensively evaluate the Census Bureau's disclosure avoidance methods to determine if they were sufficient to protect against these disclosure risks.[35]

2010 CENSUS SIMULATED RECONSTRUCTION-ABETTED RE-IDENTIFICATION ATTACK

40. The results from the Census Bureau's 2016-2019 research program on simulated reconstruction-abetted re-identification attack were conclusive, indisputable, and alarming. Appendix B, attached to this declaration, provides an overview of that simulation and the results. The bottom line is that our simulated attack showed that a conservative attack scenario using just 6 billion of the over 150 billion statistics released in 2010 would allow an attacker to accurately re-identify *at least* 52 million 2010 Census respondents (17% of the population) and the attacker would have a high degree of confidence in their results with minimal additional verification or field work. In a more pessimistic scenario, an attacker with access to higher quality commercial name and address data than those used in our simulated attack could accurately re-identify around 179 million Americans or around 58% of the population.

41. Emerging attack scenarios and our own internal simulated attacks show that were the Census Bureau to use the disclosure avoidance mechanism implemented for the 2010 Census again for the 2020 Census, the results would be vulnerable to reconstruction and re-identification attacks because of the parameters of the swapping mechanism's 2010 implementation: an overall insufficient level of noise, the invariants preserved without noise, and the geographic and demographic detail of the published summary

[35] DSEP Meeting Record, February 5, 2015. See Appendix D.

data. The Census Bureau can no longer rely on the swapping implementation used in 2010 if it is to meet its obligations to protect respondent confidentiality under 13 U.S. Code §§ 8(b) & 9. Protecting against new technology-enabled re-identification attacks, while maintaining the high quality of the decennial census data products, requires the implementation of a disclosure avoidance mechanism that is better able to protect against these new, sophisticated vectors of attack.

DISCLOSURE AVOIDANCE OPTIONS CONSIDERED FOR THE 2020 CENSUS

42. Faced with compelling evidence of the inherent vulnerability of the 2010 Census swapping mechanism to protect against reconstruction-abetted re-identification attacks, the Census Bureau began exploring the available data protection strategies that it could employ for the 2020 Census. This process was overseen by the Data Stewardship Executive Policy Committee. The three disclosure avoidance methods the Census considered were *Enhanced Data Swapping*, *Suppression*, and *Differential Privacy*.

43. The Census Bureau decided that differential privacy was the best tool after analyzing the various options through the lens of economics. Efficiently protecting privacy can be viewed as an economic problem because it involves the allocation of a scarce resource—confidential information—between two competing uses: public data products and privacy protection. If we produce more accuracy, we will have less privacy, and vice versa. And just like in the classic economic example of the trade-off between producing guns and butter, the tradeoff between privacy and accuracy can be analyzed with a production possibility curve. Our empirical analysis showed that differential privacy offered the most efficient trade-off between privacy and accuracy—our calculations showed that the efficiency of differential privacy dominated traditional

methods.[36] In other words, regardless of the level of desired confidentiality, differential privacy will always produce more accurate data than the alternative traditional methods considered by the Census Bureau.

44. *Differential Privacy.* Differential privacy, first developed in 2006, is a framework for quantifying the precise disclosure risk associated with each incremental release from a confidential data source.[37] In turn, this allows the Census Bureau to quantify the precise amount of statistical noise required to protect privacy. This precision allows the Census Bureau to calibrate and allocate precise amounts of statistical noise in a way that protects privacy while maintaining the overall statistical validity of the data.

45. The Census Bureau first began using differential privacy to protect its statistical data products in 2008, with the launch of its OnTheMap tool for employee commuting statistics and its heavily used extension OnTheMap for Emergency Management. In the years since, the Census Bureau has also successfully used differential privacy in a number of other innovative statistical products, such as the Post-Secondary Employment Outcomes and Veteran Employment Outcomes products. Differential privacy is also being used by many of the major technology firms, including Apple[38], Google,[39]

---

[36] See Abowd, J. M., & Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. American Economic Review, 109(1), 171-202.

[37] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Springer, Berlin, Heidelberg.

[38] Differential Privacy Team. (2017). "Learning with Privacy at Scale." Apple Machine Learning Journal, 1(8).

[39] Erlingsson, U., V. Pihur, and A. Korolova. (2014). "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response." Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14, 1054–1067.

Microsoft,[40] and Uber.[41] Other statistical agencies, such as the Statistics of Income Division of the Internal Revenue Service, have also begun implementing differential privacy.[42] Internationally, the Australian Bureau of Statistics,[43] the Office of National Statistics in the United Kingdom,[44] and Statistics Canada[45] explicitly recognize the threat from combining multiple statistical tabulations to re-identify respondent information and recommend output noise infusion systems, including differential privacy.

46. Given its obligations to produce high quality statistics from the decennial census while also protecting the confidentiality of respondents' census records under 13 U.S. Code §§ 8(b) & 9, the Census Bureau's Data Stewardship Executive Policy Committee

[40] Ding, B., J. Kulkarni, and S. Yekhanin. (2017). "Collecting Telemetry Data Privately." Advances in Neural Information Processing Systems 30.

[41] Near, J. (2018) "Differential Privacy at Scale: Uber and Berkeley Collaboration," Enigma 2018 (January) USENIX Assoc. https://www.usenix.org/node/208168.

[42] Bowen, C. et al. (2020) "A Synthetic Supplemental Public-Use File of Low-Income Information Return Data: Methodology, Utility, and Privacy Implications," (July) Tax Policy Center, The Brookings and Urban Institutes. https://www.urban.org/sites/default/files/publication/102547/a-synthetic-supplemental-public-use-file-of-low-income-information-return-data_2.pdf.

[43] Australian Bureau of Statistics, (2019) "Protecting the Confidentiality of Providers," January 2019, *1504.0 - Methodological News*, https://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/1504.0Main%20Features9999Jan%202019?opendocument&tabname=Summary&prodno=1504.0&issue=Jan%202019&num=&view=, accessed on March 31, 2021.

[44] United Kingdom Office for National Statistics, (2021) "Policy on Protecting Confidentiality in Tables of Birth and Death Statistics," https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/policyonprotectingconfidentialityintablesofbirthanddeathstatistics#annex-a-understanding-the-legal-and-policy-framework, accessed on March 31, 2021.

[45] Statistics Canada, (2021) "A Brief Survey of Privacy Preserving Technologies," March 2021, *Data Science Network for the Federal Public Service*, https://www.statcan.gc.ca/eng/data-science/network/privacy-preserving, accessed on March 31, 2021.

determined that the Census Bureau should proceed with the deployment and testing of differential privacy for use in the 2020 Census.[46]

47. The best disclosure avoidance option that offers a solution capable of addressing the new risks of reconstruction-abetted re-identification attacks, while preserving the fitness-for-use of the resulting data for the important governmental and societal uses of census data, is differential privacy. I have summarized here what I consider to be the most important reasons that the Census Bureau decided to adopt differential privacy as the privacy risk accounting framework for the 2020 Census Disclosure Avoidance System (DAS).

48. **Disclosure avoidance must be proactive.** The fundamental objective of disclosure avoidance protections is to proactively prevent disclosures. Just like corporations are not expected to wait until they have suffered a major data breach before upgrading their IT security systems to protect against known threats, statistical agencies should

[46] On May 10-11, 2017, DSEP decided that "any request for disclosure avoidance of proposed publications for the 2020 Census be routed to the 2020 DAS team before going to the DRB" meaning that all 2020 Census publications would be subject to differential privacy. See Appendices E and F. On February 15, 2018, DSEP suspended publication of "all proposed tables in Summary File 1 and Summary File 2 for the 2020 Census at the block, block-group, tract, and county level except for the PL94-171 tables, as announced in Federal Register Notice 170824806–7806–01…" acknowledging that "…these data in many cases were accurate to a level that was not supported by the actual uses of those data, and such an approach is simply untenable in a formally private system." DSEP further decided that "SF1 and SF2 will be rebuilt based on use cases." See Appendix G. In parallel with these decisions by DSEP, the disclosure risks identified by the preliminary results of the simulated reconstruction attack also led to this issue being added to the Census Bureau's risk management portfolio. On April 17, 2017, the risk of reconstruction attacks was proposed for inclusion in the Research and Methodology Directorate's risk registry. On September 12, 2017, it was escalated and included on the Enterprise-level Risk register. Finally, on January 30, 2018, it was further escalated to the Enterprise-level Issue register, with the development and use of the 2020 Census Disclosure Avoidance System as an identified resolution action to be taken.

not wait until they suffer a confirmed breach before improving their disclosure avoidance protections to account for known threats. The expectation, for both IT security and disclosure avoidance, is to remain vigilant about emerging threats and risks, and to take appropriate action *before* those risks lead to a breach.

49. **The disclosure risk landscape has fundamentally changed since 2010.** Traditional methods of assessing disclosure risk rely on knowing what tools and resources an attacker might leverage to undermine confidentiality protections. These tools, however, are ever evolving. Over the last decade, technological advances have made powerful cloud computing environments, with sophisticated optimization algorithms capable of performing large-scale attacks, cheap and easily available. While these tools were not yet a viable attack model in 2010, they certainly represent a credible threat today.[47]

50. **Internal research has conclusively proven the fundamental vulnerabilities of the 2010 swapping methodology.** The Census Bureau has performed extensive empirical analysis of the disclosure risk inherent to the 2010 Census swapping methodology as detailed in Appendix B. No disclosure avoidance technique can produce usable data with absolutely zero risk of re-identification, but the re-identification rates from our internal experiments on the 2010 Census swapping methodology are orders of magnitude higher than what they were intended to be. The privacy threat landscape has evolved over the last decade and compels the Census Bureau to adapt its protections accordingly.

---

[47] DSEP drew this conclusion from the simulated reconstruction-abetted re-identification attack in Appendix B. The Office of National Statistics reached the same conclusion in its 2018 "Privacy and data confidentiality methods: a Data and Analysis Method Review (DAMR)" at [Privacy and data confidentiality methods: a Data and Analysis Method Review (DAMR) – GSS (civilservice.gov.uk)] (cited on April 10, 2021).

51. **The Census Bureau determined that differential privacy was the only method that could adequately protect the data while preserving the value of census data products.** When our internal research demonstrated the vulnerabilities of the swapping mechanism used for the 2010 Census, we considered a range of options for the 2020 Census. The three leading options were differential privacy, an enhanced version of data swapping, and a return to whole-table suppression. But to achieve the necessary level of privacy protection, both enhanced data swapping and suppression had severely deleterious effects on data quality and availability. With its enhanced privacy protections and precision control over the tuning of privacy/accuracy tradeoff, the Census Bureau determined that differential privacy was the only viable solution for the 2020 Census.

52. **Differential privacy can be fine-tuned to strike a balance between privacy and accuracy.** The Data Stewardship Executive Policy Committee made the preliminary decision to pursue differential privacy on May 10-11, 2017. Since that decision was announced, the Census Bureau has worked extensively with our advisory committees, federal agency partners, American Indian and Alaska Native tribal leaders, the Committee on National Statistics, professional associations, data user groups, and many others at the national, state, and local levels to understand how they use decennial census data and to ensure that our implementation of differential privacy will preserve the value of the decennial census as a national resource. The Census also released sets of demonstrative data to allow the public and end-users to provide feedback that allowed us to fine-tune and tweak how we will ultimately implement differential privacy.[48]

---

[48] U.S. Census Bureau "Developing the DAS: Demonstration Data and Progress Metrics" https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-development.html.

53. **The need to modernize our privacy protections has been confirmed by external experts.** The Census Bureau's ongoing partnerships with scientific and academic experts from around the country helped us conduct the internal evaluation of the disclosure risk of the 2010 Census swapping methodology and confirmed the need to modernize our privacy protections. To supplement this ongoing work and to get external expert confirmation of the conclusions that we have drawn from it, the Census Bureau also commissioned an independent expert review by JASON, an independent group of elite scientists that advise the federal government on science and technology. The JASON report confirmed our findings regarding the re-identification risk inherent to the 2010 Census swapping methodology.[49]

IMPLEMENTING DIFFERENTIAL PRIVACY FOR THE 2020 CENSUS: INVARIANTS

54. Census announced that it planned to use Differential Privacy for the 2020 Census in a few different venues: (1) August 3, 2018, 2020 Census Program Management Review; (2) December 6, 2018, Census Scientific Advisory Committee Meeting; and (3) May 2, 2019, Census National Advisory Committee meeting.

55. The Bureau has engaged in a years-long campaign to educate the user community and solicit their views about how differential privacy should be implemented. Census Bureau staff have made hundreds of public presentations, held dozens of webinars, held formal consultations with American Indian and Alaska Native tribal leaders, created an extensive website with plain English blog posts, and conducted regular outreach with dozens of stakeholder groups. We have made presentations to our

[49] JASON (2020). "Formal Privacy Methods for the 2020 Census" JASON Report JSR-19-2F. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/privacy-methods-2020-census.pdf.

scientific advisory committees and provided substantial information to oversight entities such as the Government Accountability Office and the Office of the Inspector General.

56. Differential privacy is a hugely complex and technical statistical process; an explanation of all of its components is beyond the scope of this declaration. But one aspect of differential privacy is critical to understand for purposes of this litigation: invariants. Invariants are data held constant when applying statistical noise. Invariants were not well understood either theoretically or empirically in 2016 when the Census Bureau began its research on differential privacy for decennial census data, but we now understand that invariants defeat privacy protections and must be limited in order to protect the integrity of the system as a whole.

57. In designing the requirements for the 2020 Census Disclosure Avoidance System (DAS) we set certain numbers as invariant, meaning we would report the number unobscured. The invariants for the 2020 Census are the state level population totals (the "apportionment totals" reported to the President as required by 13 U.S.C. § 141(b)), the block level housing unit counts, and block-level occupied group quarters counts by type. Neither housing unit nor the group quarters counts include population data; they are counts of addresses. The Census Bureau did NOT set as invariant any other totals and forcing us to produce such numbers unredacted now would severely compromise and weaken the confidentiality protections of the DAS, which would have cascading effects on the Census Bureau's ability to meets its confidentiality obligations under Title 13.

58. Unlike traditional approaches to disclosure avoidance, differentially private noise infusion offers quantifiable and provable privacy guarantees. These guarantees, reflected in the global privacy-loss budget[50] and its allocation to each statistic, serve as a promise to data subjects that there is an inviolable upper bound to the risk that an attacker can learn or infer something about those data subjects through publicly released data products. While that upper bound is ultimately a policy decision, and may be low or high depending on the balancing of the countervailing obligations to produce accurate data and to protect respondent confidentiality, the level of the global privacy-loss budget is central to the ability of the approach to protect the data. Invariants are, by their very nature, the equivalent of assigning infinite privacy-loss budget to particular statistics, which fundamentally violates the central promise of differentially private solutions to controlling disclosure risk. By excluding the accuracy of invariant data elements from the control of the privacy-loss budget, invariants exclude the disclosure risk and potential inferences that can be drawn from those data elements from the formal privacy guarantees. Thus, instead of being able to promise data subjects that the publication of data products will limit an attacker to being able to infer, at most, a certain amount about them (with that amount being determined by the size of the privacy-loss budget and its allocation to each characteristic), the inclusion of one or more invariants fundamentally excludes attacker inferences about the invariant characteristic(s) from the very nature of that promise. The qualifications and

[50] The global privacy-loss budget and its allocation to each statistic produced by the TopDown Algorithm are the tools that differential privacy uses to keep track of the overall risk of confidentiality breaches. Larger global privacy-loss budgets imply increased vulnerability because statistical accuracy increases as the privacy-loss budget increases. The vulnerability of releasing statistics that are too accurate, and thus pose a confidentiality risk, is controlled by allocating the privacy-loss budget over the geographic hierarchy and across all the statistics computed at each level of the hierarchy.

exclusions to the privacy guarantee weaken the strength of the approach and make communicating the resulting level of protection substantially more difficult.

59. The federal government and the broader statistical disclosure limitation field have long acknowledged the necessity of considering all releases of related data when making decisions regarding disclosure risk.[51] These disclosure risk assessments for the 2020 Census have already been made and implemented for the 2020 Census P.L. 94-171 Redistricting Data Summary Files, based on the previously approved list of invariants.

60. The Census Bureau has already evaluated the impact of the existing invariants on the stability of the DAS and the resulting confidentiality of the data. The privacy loss accounting reflected in the approved DAS settings and privacy-loss budget allocation for production of the 2020 Census P.L. 94-171 Redistricting Data Summary Files takes these impacts into account. The inclusion of additional invariants (publication of additional data without privacy protections) would invalidate this accounting, would render the resulting privacy guarantee represented by the privacy-loss budget allocation meaningless, and would subject Census respondents to unquantified additional privacy risk.

61. The Census Bureau has subjected its differential privacy mechanisms, programming code, and system architecture to thorough outside peer review. We have also committed to publicly releasing the entire production code base. We have already released the full suite of implementation settings and parameters for the production code base. Many traditional disclosure avoidance methods, most notably swapping techniques,

---

[51] Office of Federal Statistical Policy and Standards (1978) Statistical Policy Working Paper #2 "Report on Statistical Disclosure and Disclosure Avoidance Techniques" p. 14, available at https://nces.ed.gov/FCSM/pdf/spwp2.pdf#:~:text=Policy%20and%20Standards%20Statisti-cal%20Policy%20Working%20Paper%202,Economist%20Office%20of%20Federal%20Statistical%20Policy%20and%20Standards. See also: Cox (1976) and Fellegi (1972).

must be implemented in a "black box." Implementation parameters for these legacy disclosure avoidance methods, especially swapping rates, are often some of the most tightly guarded secrets that the Census Bureau protects. But differential privacy does not rely on the obfuscation of its implementation as a means of protecting the data. The Census Bureau's transparency will allow any interested party to review exactly how the algorithm was applied to the 2020 Census data, and to independently verify that there was no improper or partisan manipulation of the data.

Application of Confidentiality Provisions to the Data at Issue in this Lawsuit

62. The Census Bureau produced 988 pages of responsive information to Plaintiff in late May 2021. This information consisted of material considered in assessing the need to make processing adjustments because of anomalies arising from the disrupted production schedule for the 2020 Census. These assessments were made by a group at the Census Bureau called the Data Quality EGG, or Executive Governance Group. The Data Quality EGG consists of Census Bureau subject matter experts and senior executives charged with ensuring the quality of the information produced in the 2020 Census. The Data Quality EGG is co-chaired by Deborah Stempowski (Assistant Director for Decennial Programs), Victoria Velkoff (Associate Director for Demographic Programs and Chief Demographer) and me. It was specifically chartered by Ron Jarmin (Deputy Director and Chief Operating Officer) in April 2020 following the pandemic-induced suspension of regular 2020 Census operations "to provide direction and approvals about: quality assessments of changes to the operation plans and quality assessments of the 2020 Census data during and post data collection."[52] In late

[52] From the Charter of the 2020 Data Quality Executive Governance Group supplied as Appendix H.

2020, the Data Quality EGG reviewed various production data relating to the Group Quarters population.[53]

63. Group quarters are defined as "places where people live or stay in a group living arrangement that is owned or managed by an organization providing housing and/or services for the residents. Group quarters differ from typical household living arrangements because the people living in them are usually not related to one another. Group quarters include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, prisons, and worker dormitories."[54]

64. The EGG reviewed statistical summaries for certain specific group quarters facilities and totals by state. The review indicated anomalies that prompted the EGG to direct the Decennial Statistical Studies Division (the office, led by Patrick Cantwell, responsible for researching and developing the data processing methodologies used for the decennial census) to develop a method to correct those anomalies, a method that came to be called "Group Quarters Count Imputation," or GQCI.

65. Count imputation is a commonly used technique in censuses and surveys for addressing the problem of missing or contradictory data. Missing and contradictory data during enumeration has been a recurring problem for the decennial census since 1790, and the Census Bureau has routinely used various forms of count imputation to address these challenges for census apportionment data since the 1960 Census. Most notably, imputation for census apportionment counts has historically filled in housing unit status (occupied, vacant, or non-existent) and household size (number of persons

---

[53] Presentations to the Data Quality EGG are deemed "Title 13 sensitive," meaning that their public release is governed by the Data Security Executive Policy Committee.

[54] See: U.S. Census Bureau (March 2021) "2020 Census Group Quarters", available at https://www.census.gov/newsroom/blogs/random-samplings/2021/03/2020-census-group-quarters.html (cited on July 19, 2021).

in the household).[55] The use of these count imputation methods for apportionment purposes was upheld by the U.S. Supreme Court in *Utah v. Evans,* 526 U.S. 425 (2002).

66. To address identified deficiencies in group quarters data for the 2020 Census, GQCI used information from the group quarters enumeration records, group quarters advance contact records, and administrative data to determine whether records were double counted, appropriately counted, or missing. The GQCI resolved the status of group quarters addresses for frame eligibility (occupied or not; unoccupied group quarters are deleted from the census frame) and, if occupied, the status of persons residing in the group quarters—eliminating duplicates and imputing missing persons.

67. The redactions challenged by Plaintiff relate to data the Census Bureau reviewed when developing the GQCI. The Disclosure Review Board reviewed these materials in connection with a document production in another lawsuit and, following standard Census Bureau procedure, applied the necessary disclosure avoidance measures, including redactions, to allow the documents to be made public.

68. The data withheld in the previously-released documents relate to specific facilities, such as the group quarters address and population counts for specific colleges and dormitories. Others are state-level numbers reflecting the group quarters address and population totals enumerated for that state compared with benchmarks. Prior to release (first done in a separate lawsuit), these data were either rounded or redacted to ensure that the released information cannot be used, in combination with other available or published information, to re-calculate specific information about the individuals residing in those group quarters facilities. This process of protecting against

---

[55] Cantwell, P.J., Hogan, H. and Styles, K. (2005). *Imputation, Apportionment, and Statistical Methods in the U.S. Census: Issues Surrounding Utah v. Evans.* U.S. Census Bureau Research Report Series (Statistics 2005-01), p.11, available at: https://www.census.gov/content/dam/Census/library/working-papers/2005/adrm/rrs2005-01.pdf.

indirect disclosure of personally identifiable information through the use of complementary disclosure avoidance methods is required under 13 U.S.C. §§ 8(b) and 9 to protect against disclosures of individual Census responses, and has been recognized as a necessary cornerstone of responsible statistical disclosure limitation since Ivan Fellegi's modernization of the discipline in 1972.[56] The risk of re-identification when complementary disclosure avoidance is not applied has more recently been called the "mosaic effect," whereby an attacker can piece together disparate information from multiple sources to recover confidential information.[57] Under the Office of Management and Budget's Memorandum M-13-13, federal agencies are required to consider the risks of the mosaic effect when performing their disclosure reviews: "Before disclosing potential PII or other potentially sensitive information, agencies must consider other publicly available data – in any medium and from any source – to determine whether some combination of existing data and the data intended to be publicly released could allow for the identification of an individual or pose another security concern."[58]

69. The disclosure avoidance performed on the data previously released to Plaintiff was performed in accordance with the Census Bureau's established disclosure avoidance

[56] Fellegi, I. P. (1972). On the question of statistical confidentiality. Journal of the American Statistical Association, 67(337), 7-18.

[57] For a definition and examples of the "mosaic effect" see OMB Memorandum M-13-13 available at: https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf pp. 4-5 (cited on July 21, 2021).

[58] Ibid.

rules for the release of summary statistics[59] and cleared for public release by the Census Bureau's Disclosure Review Board (DRB) under clearance number CBDRB-FY21-DSEP-002. Pursuant to the disclosure avoidance rules established by the DRB, the number of unweighted record counts, or counts by category, may be reported if they are rounded, with the coarseness of rounding contingent on the underlying number of records.[60] Means for unweighted count data may be reported with up to four significant digits, though decimals must often be redacted as they can be used to calculate the underlying number of counts used as the denominator for calculation of the mean. Quartile distributions, maxima, and minima for unweighted counts are generally suppressed, as are statistics calculated from those counts (e.g., a range, which is calculated by subtracting the minimum from the maximum of the distribution; alternatively, ranges can be reported if they are appropriately rounded). In certain cases, additional redactions may also be required, depending on the characteristics or geographic detail of the data being summarized. Conversely, the Disclosure Avoidance Officers performing the disclosure reviews may, depending on the characteristics of the data being summarized, use their expert judgement to identify alternative disclo-

[59] See "Federal Statistical Research Data Center Disclosure Avoidance Methods: A Handbook for Researchers" for a summary of these types of disclosure avoidance mechanisms that have been approved by the Census Bureau's Disclosure Review Board.

[60] The standard Census Bureau rules for rounding unweighted counts are:

- If N is less than 15, report N < 15
- If N is between 15 and 99, round to the nearest 10
- If N is between 100-999, round to the nearest 50
- If N is between 1,000-9,999, round to the nearest 100
- If N is between 10,000-99,999, round to the nearest 500
- If N is between 100,000-999,999, round to the nearest 1,000
- If N is 1,000,000 or more, round to four significant digits.

sure avoidance mechanisms to apply. For example, while ranges are typically redacted (e.g., ECF No. 8-10, Page 47 of 118) according to DRB rules, some ranges could be alternatively protected through rounding (e.g., ECF No. 8-10, Page 57 of 118).

70. I also understand that Plaintiff is seeking production of emails related to the state-by-state totals that reflect the number of individuals added to the 2020 Census totals by the GQCI program. We consider these numbers to be clearly Title 13 protected. As information protected by Title 13, Census Bureau policy prohibits this information from being sent though regular email.[61] Instead, the requested information is stored on a secure file space and only transmitted over secure encrypted networks when necessary.

71. The Census Bureau can identify the requested information in the secure file space, but it is withholding such information in full because it is protected by Title 13. If these numbers were not redacted or rounded, they would have to be considered invariant by the 2020 DAS. But, as explained above, the Census Bureau did not set group quarters population data as invariant. The disclosure of such information without redaction or rounding would, therefore, significantly weaken the privacy protections of the 2020 Census Disclosure Avoidance System, compromise the confidentiality protections used for the redistricting data, and undermine the Census Bureau's efforts to fulfill its duties under Title 13's confidentiality provisions for future 2020 Census data releases.

72. During the extensive stakeholder engagement performed to design, improve, and tune the DAS, the Census Bureau received substantial feedback from our data user community highlighting concerns related to group quarters data. One of those concerns was the accuracy of group quarters population counts by type at the block and

[61] U.S. Census Bureau Data Stewardship Policy (DS007) "Safeguarding and Managing Information" available at https://www2.census.gov/foia/ds_policies/ds007.pdf

block group levels. To address this concern, the DAS geographic spine was modified to isolate group quarters of the same type in their own custom block groups. Then, substantial privacy-loss budget was allocated to the block-group population totals in the production redistricting data. Releasing further group quarters population data that have not been processed through the DAS, such as the information requested by Plaintiff, would greatly compromise the confidentiality for all respondents living in the block groups containing GQs (both those respondents residing in GQs and those in non-GQ housing units). The release of unintended exact information that has not been accounted for by the DAS —the data requested by Plaintiff—provides information about these populations above and beyond the controlled statistics produced by the DAS. Even release of state-level summaries can compromise these protections, most easily in the case of small states or for less common types of group quarters facilities. For example, if there were only one of particular type of group quarters facility within a geographic area (e.g., a single military/maritime vessel within a state), then unprotected state-level GQCI statistics for that type of group quarters could easily be leveraged to undermine the disclosure protections afforded to the tabulated Census data for that GQ in the published census data products, thus exposing the personal information of the facility's residents. Unprotected GQCI statistics for larger numbers of GQs within a state can similarly be disclosive, though the calculations to leverage these data in a privacy attack would require a bit more effort. This is why the Disclosure Review Board, acting on instructions from the Data Stewardship Executive Policy Committee, applied its unweighted count rounding rules to the state-level summaries.

SETTING OF THE PRIVACY LOSS BUDGET AND TIMELINE MOVING FORWARD

73. On June 9, 2021,[62] the U.S. Census Bureau's Data Stewardship Executive Policy Committee announced it had selected the settings and parameters for the Disclosure Avoidance System (DAS) for the 2020 Census redistricting data (PL-94-171). After reviewing feedback from the data user community, the committee approved a revised algorithm that ensures the accuracy of data necessary for redistricting and Voting Rights Act enforcement. Our Disclosure Avoidance team used these parameters when processing the 2020 Census data and related quality assurance checks. The data was run and quality checked multiple times.
74. As part of an agreement in another litigation, the Census Bureau has committed to producing the redistricting data products by August 16, 2021.

HARM SHOULD THIS COURT MANDATE DISCLOSURE

75. Were the Court to order that we release the redacted information in the records already produced, or produce the unredacted state-by-state GQCI totals, the Census would be faced with hard choices. The result would likely be a significant delay in delivery of the already-delayed redistricting data and diminished accuracy.
76. The effect on the schedule for delivering redistricting data would likely be substantial. The Census Bureau cannot ascertain the length of the delay, but to account for the addition of another invariant, the Census Bureau's Data Stewardship Executive Policy Committee would need to review and evaluate the impact of the new invariant on the privacy-loss accounting, test the stability of the new invariant on DAS processing, assess the impact of such a decision on associated products, and potentially determine

---

[62] U.S. Census Bureau (June 2021) "Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results" available at https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates/2021-06-09.html

a new privacy-loss budget for the redistricting data product that factors in the additional privacy risk resulting from the invariant, re-tune the DAS algorithms to ensure fitness-for-use for the identified priority use cases, then apply those settings in production and subject the results to expert subject matter review prior to production of data. Even if this process were to be expedited and no algorithmic stability issues were to occur as a result of the inclusion of the new invariant, this process would likely delay the release of the redistricting data – potentially for as long as six months beyond the court-mandated August 16, 2021 deadline.

77. Disclosure of the redacted information after release of the redistricting data on August 16 would similarly harm the protections provided by the DAS, as well as harming public trust in the Census Bureau's promises for future censuses and surveys, as discussed above in "Importance of Confidentiality" (paragraphs 10-15).

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

DATED and SIGNED:

JOHN ABOWD Digitally signed by JOHN ABOWD
Date: 2021.07.25 20:45:38 -04'00'

John M. Abowd

Chief Scientist and Associate Director for Research and Methodology

United States Bureau of the Census

Appendix A

John M. Abowd

Home | Professional Information | Courses | Recent News | Special Materials

Professional Information

[Updated July 21, 2021]

CONTACT INFORMATION

U.S. Census Bureau

Please do not send physical mail or packages to the U.S. Census Bureau. Use my email contact to get a mailing address.

HQ 8H120 ATTN: Sara Sullivan
4600 Silver Hill Road
Private delivery services (FedEx, UPS, etc.) physical location: Suitland, MD 20746
USPS mail only: Washington, DC 20233
Voice: +1.301.763.5880
Mobile: +1.202.591.0766
Fax: +1.301.763.8360
Executive assistant Sara Sullivan: +1.301.763.5116
E-mail: john.maron.abowd@census.gov

ILR School

Please do not send physical mail or packages to Cornell University. Use my email contact to get a mailing address.

Ives Hall
Cornell University
Ithaca, New York 14853-3901
Assistant: LDI@cornell.edu
E-mail: john.abowd@cornell.edu

Webpage: https://blogs.cornell.edu/abowd/ or https://www.johnabowd.com

Twitter: @john_abowd (opinions are my own)

Short biography in PDF format

CURRENT POSITIONS

Chief Scientist and Associate Director for Research and Methodology, U. S. Census Bureau, IPA June 1, 2016 – March 27, 2020; Career Senior Executive Service March 29, 2020 —

Edmund Ezra Day Professor Emeritus of Economics, Statistics and Data Science, July 1, 2021 —

Member of the Graduate Fields of Economics, Industrial and Labor Relations, Information Science, and Statistics

Search … SEARCH

INSTITUTIONS

- U.S. Census Bureau
- Cornell Economics
- Labor Dynamics Institute
- NCRN node at Cornell
- CISER

OTHER INFORMATION

- Google Scholar
- ORCID
- RePEC/Ideas
- SSRN

Research Associate, National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, Massachusetts 02138, September 1983 – (on leave while serving at the U.S. Census Bureau)

Research Affiliate, Centre de Recherche en Economie et Statistique/INSEE, 15, bd Gabriel Péri, 92245 Malakoff Cedex France, November 1997 –

Research Fellow, IZA (Institute for the Study of Labor), P.O. Box 7240 D-53072 Bonn, Germany, June 2002 —

Research Fellow, IAB (Institut für Arbeitsmarkt-und Berufsforschung), Dienstgebäude Weddigenstraβe 20-22, 90478 Nürnberg, Germany, January 2013 —

President and Principal, ACES-Research, LLC, john@aces-research.com, July 2007 —

Editor, Journal of Privacy and Confidentiality Online journal

PREVIOUS AND VISITING POSITIONS

Edmund Ezra Day Professor, Department of Economics, Cornell University, July 2011 – June 2021

Director, Labor Dynamics Institute, Cornell University, October 2011 – June 30, 2020

Founding member and Professor of Information Science (by courtesy), Faculty of Computing and Information Science, July 2000 – April 2021

Professor of Statistics and Data Science, September 2013 – April 2021

Distinguished Senior Research Fellow, United States Census Bureau, September 1998 – May 2016

Associate Chair, Department of Economics, Cornell University, August 2015 – May 2016

Visiting Professor, Center for Labor Economics, University of California-Berkeley, August 2014 – July 2015

Director of Graduate Studies, Economics, July 2010 – June 2014

Professor of Economics and Econometrics, University of Notre Dame, January 2008 – May 2008.

Director, Cornell Institute for Social and Economic Research (CISER), July 1999 – December 2007

Associate Director, Cornell Theory Center (became Cornell University Center for Advanced Computing), September, 2006 – August 2007.

Professor of Labor Economics, Cornell University, January 1990 – October 2001.

Edmund Ezra Day Professor, School of Industrial and Labor Relations, November 2001 —

Associate Director, Cornell Institute for Social and Economic Research (CISER), July 1998 – June 1999.

Chair, Department of Labor Economics, Cornell University, September 1992 – June 1998.

Acting Director, CISER, January 1998-June 1998.

Professeur invité, Laboratoire de Microéconomie Appliquée-Theorie Et Applications en Microéconomie et macroéconomie (LAMIA-TEAM), Université de Paris-I (Panthéon-Sorbonne), May 1998.

Consultant, Centre de Recherche en Economie et Statistique (CREST), Institut National de la Statistique et des Etudes Economiques (INSEE), February 1997.

Professeur invité, ERMES (Equipe de Recherche sur les Marchés, l'Emploi et la Simulation) Université Panthéon-Assas (Paris II), October 1995 – July 1996 (part time).

Professor, Samuel Curtis Johnson Graduate School of Management, Cornell University (adjunct appointment), August 1987 – July 1995.

Chercheur étranger, Institut National de la Statistique et des Etudes Economiques (INSEE), Paris, Department of Research, August 1991 – July 1992, January 1993, January 1994.

Professeur visitant, HEC (Hautes Etudes Commerciales, Paris) Department of Finance and Economics, September 1991 – July 1992 and January 1993, December 1993 – January 1994.

Professeur visitant, CREST (Centre de Recherche en Statistiqu et Economie, Paris), September 1991 – July 1992, July 1993.

Associate Professor with tenure, Cornell University, August 1987 – December 1989.

Research Associate, Industrial Relations Section, Department of Economics, Princeton University, September 1986 – August 1987.

Visiting Associate Professor of Economics, Department of Economics, Massachusetts Institute of Technology, September 1985 – August 1986.

Associate Professor of Econometrics and Industrial Relations, Graduate School of Business, University of Chicago, September 1982 – August 1986. Assistant Professor, September 1979 – August 1982. Visiting Assistant Professor, September 1978 – August 1979.

Senior Study Director/Research Associate, NORC/Economics Research Center, 6030 Ellis Avenue, Chicago, Illinois 60637, September 1978 – August 1986.

Academic Consultant, Centre for Labour Economics, London School of Economics, January 1979 – April 1979.

Assistant Professor of Economics, Department of Economics, Princeton University, September 1977 – August 1979 (on leave September 1978 – August 1979). Lecturer in Economics, September 1976 – August 1977.

Associate Editor, *Journal of Business and Economic Statistics*, 1983 – 1989.

Editorial Board, *Journal of Applied Econometrics*, 1987 – 1989.

Associate Editor, *Journal of Econometrics*, 1987 – 1989.

EDUCATION

Ph.D. Department of Economics, University of Chicago, December 1977.
Thesis: An Econometric Model of the U.S. Market for Higher Education

M.A. Department of Economics, University of Chicago, March 1976.

A.B. Department of Economics (with highest honors), University of Notre Dame, May 1973.

LANGUAGES

English (native), French

HONORS AND FELLOWSHIPS

Fellow, American Association for the Advancement of Science (elected October 2020)

Julius Shiskin Award, American Statistical Association, Business and Economic Statistics Section (2016)

Cornell University, Graduate and Professional Student Assembly Award for Excellence in Teaching, Advising, and Mentoring (May 2015)

Fellow, Econometric Society (elected November 2014)

Roger Herriot Award, American Statistical Association, Government and Social Statistics Sections (2014)

Elected member, International Statistical Institute (March 2012)

Council of Sections (2014-2016), Chair (2013) Business and Economic Statistics Section (Chair-elect 2012), American Statistical Association

President (2014-2015), Society of Labor Economists, President-elect (2013-2014), Vice President (2011-2013)

Fellow, The American Statistical Association (elected August 2009)

Fellow, Society of Labor Economists (elected November 2006)

La bourse de haut niveau du Ministère de la Recherche et de la Technologie, fellowship for research at the Institut National de la Statistique et des Etudes Economiques (INSEE) awarded by the French Government, September 1991 – February 1992.

National Institute of mental Health postdoctoral fellow at NORC, September 1978 – August 1980.

National Institute of Mental Health pre-doctoral fellow at the University of Chicago, September 1973 – June 1976.

TEACHING EXPERIENCE

Graduate:

Microeconometrics using Linked Employer-Employee Data (CREST-ENSAE)
Understanding Social and Economic Data (Cornell, co-instructor: Lars Vilhuber)
Third-year Research Seminar I and II (Cornell)
Seminar in Labor Economics I, II, and III (Cornell)
Microéconometrie des Données Appariées (CREST-GENES, in French)
Microéconomie et Microéconometrie du Travail (Université de Paris I, in French)
Economie du Travail (Université de Paris II, in French)
Economics of Compensation and Organization (Cornell)
International Human Resource Management (Cornell)
Corporate Finance (Hautes Etudes Commerciales, Paris)
International Human Resource Management (HEC, Paris)
Workshop in Labor Economics (Cornell)
Economics of Collective Bargaining (Cornell)
Executive Compensation (Cornell)
Labor Economics (MIT)
Labor and Public Policy (MIT)
Applied Econometrics I, II (Chicago)
Introduction to Industrial Relations (Chicago)
Econometric Theory I (Chicago)
Industrial Relations and International Business (Chicago)

Workshop in Economics and Econometrics (Chicago)
Econometric Analysis of Time Series (Princeton)
Mathematics for Economists (Princeton)

Undergraduate:

Understanding Social and Economic Data (Cornell, co-instructor: Lars Vilhuber)
Introductory Microeconomics (Cornell)
Economics of Employee Benefits (Cornell)
Economics of Wages and Employment (Cornell)
Corporate Finance (Cornell)
Introduction to Econometrics (Princeton)
Microeconomics (Princeton)

BIBLIOGRAPHY

Books

1. Abowd, John M. and Francis Kramarz (eds.) *The Microeconometrics of Human Resource Management*, special issue of *Annales d'économie et de statistique* 41/42 (Paris: ADRES, January/June 1996).
2. Abowd, John M. and Richard B. Freeman (eds.) *Immigration, Trade and the Labor Market* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 1991).

Articles

1. McKinney, Kevin L., John M. Abowd, and John Sabelhaus, "United States Earnings Dynamics: Inequality, Mobility, and Volatility," In Raj Chetty, John N. Friedman, Janet C. Gornick, Barry Johnson, and Arthur Kennickel, eds., *Measuring the Distribution and Mobility of Income and Wealth*, (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2021), forthcoming. [download preprint] [download chapter (open access)]
2. Abowd, John M. "Official Statistics at the Crossroads: Data Quality and Access in an Era of Heightened Privacy Risk," *The Survey Statistician*, Vol. 83 (January 2021):23-26. [download (open access)]
3. McKinney, Kevin L., Andrew S. Green, Lars Vilhuber and John M. Abowd "Total Error and Variability Measures for the Quarterly Workforce Indicators and LEHD Origin-Destination Employment Statistics in OnTheMap" *Journal of Survey Statistics and Methodology* (November 2020). [download arxiv preprint], DOI: https://doi.org/10.1093/jssam/smaa029, supplemental online materials DOI: https://doi.org/10.5281/zenodo.3951670
4. Abowd, John M., Ian M. Schmutte, William Sexton, and Lars Vilhuber "Why the Economics Profession Must Actively Participate in the Privacy Protection Debate," *American Economic Association: Papers and Proceedings*, Vol. 109 (May 2019): 397-402, DOI:10.1257/pandp.20191106. [download preprint]
5. Abowd, John M. and Ian M. Schmutte "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review,* Vol. 109, No. 1 (January 2019):171-202, DOI:10.1257/aer.20170627. [AER, ArXiv preprint, Replication information]
6. Weinberg, Daniel H., John M. Abowd, Robert F. Belli, Noel Cressie, David C. Folch, Scott H. Holan, Margaret C. Levenstein, Kristen M. Olson, Jerome P. Reiter, Matthew D. Shapiro, Jolene Smyth, Leen-Kiat Soh, Bruce D. Spencer, Seth E. Spielman, Lars Vilhuber, and Christopher K. Wikle "Effects of a Government-Academic Partnership: Has the NSF-Census Bureau Research Network Helped Secure the Future of the Federal Statistical System?" *Journal of Survey Statistics and Methodology* (2018) DOI:10.1093/jssam/smy023. [download, open access] [download preprint]
7. Abowd, John M., Ian M. Schmutte and Lars Vilhuber "Disclosure Limitation and Confidentiality Protection in Linked Data," in A.Y. Chun, M. Larson, J. Reiter, and G. Durrant (eds.) *Administrative Records for Survey Methodology* (New York: Wiley, forthcoming). [download preprint]
8. Abowd, John M., Kevin L. McKinney and Ian M. Schmutte "Modeling Endogenous Mobility in Earnings Determination," *Journal of Business and Economic Statistics* Vol. 37, Issue 3

(2019):405-418. DOI: 10.1080/07350015.2017.1356727. [download preprint] [JBES]

9. Abowd, John M., Francis Kramarz, Sébastien Perez-Duarte, and Ian Schmutte "Sorting between and within Industries: A Testable Model of Assortative Matching," *Annals of Economics and Statistics* 129 (March 2018): 1-32. NBER WP-20472. [download preprint] [programs] [data]
10. Abowd, John M., Kevin L. McKinney and Nellie Zhao "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data," *Journal of Labor Economics* 36, S1 (January 2018):S183-S300 DOI: 10.1086/694104. [download, not copyrighted] [download preprint]
11. Abowd, John M. "How Will Statistical Agencies Operate When All Data Are Private?" *Journal of Privacy and Confidentiality*, Vol. 7, Issue 3, Article 1 (2017). [download, open journal]
12. Haney, Samuel, Ashwin Machanavajjhala, John M. Abowd, Matthew Graham, Mark Kutzbach, and Lars Vilhuber "Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics," ACM SIGMOD 2017, DOI: 10.1145/3035918.3035940. [download]
13. Abowd, John M. and Kevin L. McKinney "Noise Infusion as a Confidentiality Protection Measure for Graph-based Statistics" *Statistical Journal of the International Association for Official Statistics* (2016) Vol. 32, No. 1, pp. 127-135, DOI: 10.3233/SJI-160958. [download article, open access] [download preprint]
14. Abowd, John M. and Ian Schmutte "Economic Analysis and Statistical Disclosure Limitation" *Brookings Panel on Economic Activity* (Spring 2015): 221-267. [download article and discussion, open access] [download preprint]
15. Schneider, Matthew J. and John M. Abowd "A New Method for Protecting Interrelated Time Series with Bayesian Prior Distributions and Synthetic Data," *Journal of the Royal Statistical Society*, Series A (2015) DOI:10.1111/rssa.12100. [download preprint]
16. Lagoze, Carl, William C. Block, Jeremy Williams, Lars Vilhuber, and John M. Abowd "Data Management of Confidential Data." *International Journal of Digital Curation* 8, no. 1 (2013): 265–278. doi:10.2218/ijdc.v8i1.259. [download preprint]
17. Abowd, John M. and Martha H. Stinson "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *Review of Economics and Statistics*, Vol. 95, No. 5 (December 2013): 1451-1467. doi:10.1162/REST_a_00352. [download, not copyrighted]
18. Abowd, John M., Matthew J. Schneider and Lars Vilhuber "Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation," *Journal of Privacy and Confidentiality*: Vol. 5: Iss. 1 (2013): Article 4. [download, open access]
19. Abowd, John M., Francis Kramarz, Paul Lengermann, Kevin L. McKinney, and Sébastien Roux "Persistent Inter-Industry Wage Differences: Rent Sharing and Opportunity Costs," *IZA Journal of Labor Economics*, 2012, 1:7, doi:10.1186/2193-8997-1-7. [download, open access] [online Appendix]
20. Abowd, John M., Lars Vilhuber and William Block "A Proposed Solution to the Archiving and Curation of Confidential Scientific Inputs," in J. Domingo-Ferrer and I. Tinnirello, eds., *Privacy in Statistical Databases* 2012, LNCS 7556, pp. 216–225, (2012). [download, open access]
21. Abowd, John M. and Lars Vilhuber "Did the Housing Price Bubble Clobber Local Labor Markets When It Burst?" *American Economic Review Papers and Proceedings* Vol. 102, No. 3 (May 2012): 589-93, doi:pdfplus/10.1257/aer.102.3.589. [download preprint] [online Appendix] [data Readme] [data]
22. Abowd, John M., R. Kaj Gittings, Kevin L. McKinney, Bryce E. Stephens, Lars Vilhuber, and Simon Woodcock "Dynamically Consistent Noise Infusion and Partially Synthetic Data As Confidentiality Protection Measures for Related Time-series," Federal Committee on Statistical Methodology, Office of Management and Budget, 2012 Research Conference Papers. [download, open access, cited on May 21, 2012] [download archival copy].
23. Abowd, John M. and Matthew Schneider "An Application of Differentially Private Linear Mixed Modeling," ICDMW, pp. 614-619, 2011 IEEE 11th International Conference on Data Mining Workshops, 2011. [download, open access]
24. Kinney, Satkartar K. , Jerome P. Reiter, Arnold P. Reznek, Javier Miranda, Ron S. Jarmin, and John M. Abowd "Towards Unrestricted Public Use Business Micro-data: The Synthetic Longitudinal Business Database," *International Statistical Review*, Vol. 79, No. 2 (December 2011):362-84, doi:10.1111/j.1751-5823.2011.00153.x. [download, subscription required] [download preprint]
25. Abowd, John M. and Lars Vilhuber "National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail," *Journal of*

Econometrics, Vol. 161 (March 2011): 82-99, doi: 10.1016/j.jeconom.2010.09.008. [download preprint] [data]

26. Abowd, John M., Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators" in T. Dunne, J.B. Jensen and M.J. Roberts, eds., *Producer Dynamics: New Evidence from Micro Data* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2009), pp. 149-230. [download, not copyrighted] [archival copy]
27. Abowd, John M., Kevin McKinney and Lars Vilhuber "The Link between Human Capital, Mass Layoffs, and Firm Deaths" in T. Dunne, J.B. Jensen and M.J. Roberts, eds., *Producer Dynamics: New Evidence from Micro Data* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2009), pp. 447-472. [download, not copyrighted] [archival copy]
28. Abowd, John M. and Lars Vilhuber "How Protective are Synthetic Data," in J. Domingo-Ferrer and Y. Saygun, eds., *Privacy in Statistical Databases*, (Berlin: Springer-Verlag, 2008), pp. 239-246. [download preprint]
29. Abowd, John M., Francis Kramarz and Simon Woodcock "Econometric Analyses of Linked Employer-Employee Data," in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data* (The Netherlands: Springer, 2008), pp. 727-760. [download preprint]
30. Abowd, John M., John Haltiwanger and Julia Lane "Wage Structure and Labor Mobility in the United States," in E. P. Lazear and K. L. Shaw, eds., *Wage Structure, Raises, and Mobility: International Comparisons of the Structure of Wages within and Across Firms* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2008), pp. 81-100. [download] [download preprint]
31. Machanavajjhala Ashwin, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber "Privacy: Theory Meets Practice on the Map," International Conference on Data Engineering (ICDE) 2008: 277-286, doi:10.1109/ICDE.2008.4497436. [download preprint])
32. Abowd, John M. and Francis Kramarz "Human Capital and Worker Productivity: Direct Evidence from Linked Employer-Employee Data," *Annales d'Economie et de Statistique*, No. 79/80, (Juillet/Décembre 2005): 323-338. [download preprint]
33. Torra, V. J.M. Abowd and J. Domingo-Ferrer "Using Mahalanobis Distance-Based Record Linkage for Disclosure Risk Assessment," in J. Domingo-Ferrer and Luisa Franconi (eds.) *Privacy in Statistical Databases* (Berlin: Springer-Verlag, 2006), pp. 233-242. [download preprint]
34. Abowd, John M., Francis Kramarz and Sébastien Roux "Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data," *Economic Journal*, Vol. 116, (June 2006): F245–F285. [download preprint]
35. Abowd, John M., Francis Kramarz and Sébastien Roux "Heterogeneity in Firms' Wages and Mobility Policies," in H. Bunzel, B.J. Christensen, G.R. Neumann and J-M. Robin, eds., *Structural Models of Wage and Employment Dynamics*, (Amsterdam: Elsevier Science, 2006), pp. 237-268. [download preprint]
36. Abowd, John M. and Lars Vilhuber "The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers," *Journal of Business and Economics Statistics*, Vol. 23, No. 2 (April 2005): 133-152, *JBES* Joint Statistical Meetingsinvited paper with discussion and "Rejoinder" (April 2005): 162-165. [download preprint].
37. Abowd, John M., John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, and Kristin Sandusky "The Relation among Human Capital, Productivity and Market Value: Building Up from Micro Evidence," in *Measuring Capital in the New Economy*, C. Corrado, J. Haltiwanger, and D. Sichel (eds.), (Chicago: University of Chicago Press for the NBER, 2005), Chapter 5, pp. 153-198. [download, not copyrighted] [download preprint]
38. Abowd, John M. and Simon Woodcock "Multiply-Imputing Confidential Characteristics and File Links in Longitudinal Linked Data," in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases* (Berlin: Springer-Verlag, 2004), pp. 290-297. [download preprint]
39. Abowd, John M., John Haltiwanger and Julia Lane "Integrated Longitudinal Employee-Employer Data for the United States," *American Economic Review Papers and Proceedings*, Vol. 94, No. 2 (May 2004): 224-229. [download preprint]
40. Abowd, John M. and Julia Lane "New Approaches to Confidentiality Protection: Synthetic Data, Remote Access and Research Data Centers," in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases* (Berlin: Springer-Verlag, 2004), pp. 282-289. [download preprint]

41. Abowd, John M. and Francis Kramarz "The Costs of Hiring and Separations," *Labour Economics*, Vol. 10, Issue 5 (October 2003): 499-530. [download preprint]
42. Abowd, John M. "Unlocking the Information in Integrated Social Data," *New Zealand Economic Papers*, 0077-9954, Vol. 36, No. 1 (June 2002): 9-31.[download preprint]
43. Abowd, John M. and Orley Ashenfelter "Using Price Indices and Sale Rates to Assess Short Run Changes in the Market for Impressionist and Contemporary Paintings" in *The Economics of Art Auctions*, G. Mosetto and M. Vecco (eds.), (Milan: F. Angeli Press, 2002). [download preprint] [access book]
44. Abowd, John M. and Simon Woodcock "Disclosure Limitation in Longitudinal Linked Data," in *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz (eds.), (Amsterdam: North Holland, 2001), 215-277. [download preprint]
45. Abowd, John M., Bruno Crépon and Francis Kramarz "Moment Estimation with Attrition: An Application to Economic Models," *Journal of the American Statistical Association,* 96, No. 456 (December 2001): 1223-1231. [download preprint]
46. Abowd, John M., Francis Kramarz, David Margolis, and Kenneth Troske "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," *Journal of the Japanese and International Economies*. Vol. 15, No. 4, (December 2001): 419-436. [download preprint]
47. Abowd, John M. Julia Lane and Ronald Prevost "Design and Conceptual Issues in Realizing Analytical Enhancements through Data Linkages of Employer and Employee Data" in the *Proceedings of the Federal Committee on Statistical Methodology*, November 2000. [download preprint]
48. Abowd, John M., Francis Kramarz, David Margolis and Kenneth Troske "Politiques salariales et performances des entreprises : une comparaison France/Etats-Unis," *Economie et Statistique*, No. 332-333 (2000): 27-38. [Corporate Wage Policies and Performances: Comparing France with the United States] [download preprint]
49. Abowd, John M. and David Kaplan "Executive Compensation: Six Questions That Need Answering," *Journal of Economic Perspectives,* 13 (1999): 145-168. [Preprint and supplementary materials available at http://hdl.handle.net/1813/56585]
50. Abowd, John M., Patrick Corbel and Francis Kramarz "The Entry and Exit of Workers and the Growth of Employment: An Analysis of French Establishments" *Review of Economics and Statistics*, 81(2), (May 1999): 170-187. [download preprint]
51. Abowd, John M. and Francis Kramarz "Econometric Analysis of Linked Employer-Employee Data," *Labour Economics,* 6(March 1999): 53-74. [download preprint]
52. Abowd, John M., Hampton Finer and Francis Kramarz "Individual and Firm Heterogeneity in Compensation: An Analysis of Matched Longitudinal Employer-Employee Data for the State of Washington" in J. Haltiwanger *et al.* (eds.) *The Creation and Analysis of Employer-Employee Matched Data*, (Amsterdam: North Holland, 1999), pp. 3-24. [download preprint]
53. Abowd, John M. and Francis Kramarz "The Analysis of Labor Markets Using Matched Employer-Employee Data," in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3(B), Chapter 40 (Amsterdam: North Holland, 1999), pp. 2629-2710. [download preprint]
54. Abowd, John M. Francis Kramarz and David Margolis "High Wage Workers and High Wage Firms," *Econometrica,* 67(2) (March 1999): 251-333. [download preprint]
55. Abowd, John M. Francis Kramarz, Thomas Lemieux, and David Margolis "Minimum Wages and Youth Employment in France and the United States," in D. Blanchflower and R. Freeman (eds.) *Youth Employment and Joblessness in Advanced Countries* (Chicago: University of Chicago Press, 1999), pp. 427-472. [download] [download preprint]
56. Abowd, John M. and Francis Kramarz "Internal and External Labor Markets: An Analysis of Matched Longitudinal Employer-Employee Data" in J. Haltiwanger, M. Manser, and R. Topel (eds.) *Labor Statistics and Measurement Issues* (Chicago: University of Chicago Press, 1998), pp. 357-370. [download] [download preprint]
57. Abowd, John M., Francis Kramarz, David Margolis and Kenneth Troske "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," in *Comparaisons internationales de salaires* (Paris: Ministère du travail et des affaires sociales and INSEE, 1996), pp. 315-327.
58. Abowd, John M. and Laurence Allain "Compensation Structure and Product Market Competition," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 207-217. [download preprint]

59. Abowd, John M., Francis Kramarz and Antoine Moreau "Product Quality and Worker Quality," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 300-322. [download]
60. Abowd, John M. and Francis Kramarz "The Microeconometrics of Human Resource Management: International Studies of Firm Practices, Introduction and Overview," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 1-9 (French), 11-19 (English).
61. Abowd, John M. and Francis Kramarz "Les Politiques Salariales : Individus et Entreprises" (Compensation Policies: Individuals and Firms), *Revue Economique* 47 (May 1996): 611-622. [download preprint]
62. Abowd, John M. and Francis Kramarz "The Economic Analysis of Compensation Systems: Collective and Individual" in Norman Bowes and Alex Grey, eds. *Job Creation and Loss: Analysis, Policy and Data Development* (Paris: OECD, 1996), pp. 47-54.
63. Abowd, John M. and Michael Bognanno "International Differences in Executive and Managerial Compensation" in R.B. Freeman and L. Katz, eds. *Differences and Changes in Wage Structures* (Chicago: NBER, 1995), pp. 67-103. [download]
64. Abowd, John M. and Thomas Lemieux "The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada," *Quarterly Journal of Economics* 108 (November 1993): 983-1014.
65. Abowd, John M. and Francis Kramarz "A Test of Negotiation and Incentive Compensation Models Using Longitudinal French Enterprise Data," in J.C. van Ours, G.A. Pfann and G. Ridder, eds. *Labour Demand and Equilibrium Wage Formation* Contributions to Economic Analysis (Amsterdam: North-Holland, 1993), pp. 111-46. [download preprint]
66. Abowd, John M. and Richard B. Freeman "Introduction and Summary" in J.M. Abowd and R.B. Freeman, eds. *Immigration, Trade and the Labor Market* (Chicago: NBER, 1991), pp. 1-25. [download]
67. Abowd, John M. and Thomas Lemieux "The Effects of International Competition on Collective Bargaining Outcomes: A Comparison of the United States and Canada," in J.M. Abowd and R.B. Freeman, eds. *Immigration, Trade and the Labor Market* (Chicago: NBER, 1991), pp. 343-67. [download]
68. Abowd, John M. "The NBER Trade and Immigration Data Files," in J.M. Abowd and R.B. Freeman, eds. *Immigration, Trade and the Labor Market* (Chicago: NBER, 1991), pp. 407-21. [download]
69. Abowd, John M. "Does Performance-based Compensation Affect Corporate Performance?" *Industrial and Labor Relations Review* 43:3 (February 1990): 52S-73S. Reprinted in *Do Compensation Policies Matter?* R.G. Ehrenberg, ed. (Ithaca, NY: ILR Press, 1990), pp. 52-73.
70. Abowd, John M., George Milkovich and John Hannon "The Effects of Human Resource Management Decisions on Shareholder Value," *Industrial and Labor Relations Review* 43:3 (February 1990): 203S-236S. Reprinted in *Do Compensation Policies Matter?* R.G. Ehrenberg, ed. (Ithaca, NY: ILR Press, 1990), pp. 203-236.
71. Abowd, John M. "The Effect of Wage Bargains on the Stock Market Value of the Firm," *American Economic Review* 79:4 (September 1989): 774-800. (working paper title: "Collective Bargaining and the Division of the Value of the Enterprise.")
72. Abowd, John M. and Joseph Tracy "Market Structure, Strike Activity, and Union Wage Settlements," *Industrial Relations* 57:2 (Spring 1989): 227-50.
73. Abowd, John M. and David Card "On the Covariance Structure of Earnings and Hours Changes," *Econometrica* 57:2 (March, 1989): 411-45.
74. Vroman, Wayne and John M. Abowd "Disaggregated Wage Developments," *Brookings Papers on Economic Activity* (1:1988): 313-46.
75. Abowd, John M. and David Card "Intertemporal Labor Supply and Long Term Employment Contracts," *American Economic Review* 77:1 (March 1987): 50-68.
76. Abowd, John M. "New Development in Longitudinal Data Collection for Labor Market Analysis: Collective Bargaining Data," *American Statistical Association 1985 Proceedings of the Business and Economic Statistics Section* (Washington, DC: ASA, 1985). (invited paper)
77. Abowd, John M. and Arnold Zellner "Estimating Gross Labor Force Flows," *Journal of Business and Economic Statistics 3* (July 1985): 254-283.
78. Abowd, John M. and Arnold Zellner "Application of Adjustment Techniques to U.S. Gross Flow Data," *Gross Flows in Labor Force Statistics*, edited by Paul Flaim and Carma Hogue, Bureau of the Census/Bureau of Labor Statistics Conference Volume (Washington, DC: GPO, 1985).

79. Abowd, John M. and Mark Killingsworth "Employment, Wages, and Earnings of Hispanics in the Federal and Nonfederal Sectors: Methodological Issues and Their Empirical Consequences," in *Hispanics in the U.S. Economy*, edited by G. Borjas and M. Tienda (New York: Academic Press, 1985), pp. 77-125.
80. Abowd, John M. "Economic and Statistical Analysis of Discrimination in Job Assignment," *Industrial Relations Research Association Proceedings of the Thirty-Sixth Annual Meetings* (Madison, WI: IRRA, 1984), pp. 34-47. (invited paper)
81. Abowd, John M. and Mark Killingsworth "Do Minority/White Unemployment Differences Really Exist," *Journal of Business and Economic Statistics* 2 (January 1984): 64-72.
82. Abowd, John M. and Arnold Zellner "Estimating Gross Labor Force Flows," *American Statistical Association 1983 Proceedings of the Business and Economic Statistics Section* (Washington, DC: ASA, 1983), pp. 162-67.
83. Abowd, John M. and Mark Killingsworth "Sex Discrimination, Atrophy and the Male-Female Wage Differential," *Industrial Relations* 22 (Fall 1983): 387-402.
84. Abowd, John M. and Henry S. Farber "Job Queues and the Union Status of Workers," *Industrial and Labor Relations Review* 35 (April 1982): 354-67. [download]
85. Abowd, John M. and Orley Ashenfelter "Anticipated Unemployment, Temporary Layoffs and Compensating Wage Differentials," in *Studies in Labor Markets*, edited by S. Rosen (Chicago: University of Chicago Press for the NBER, 1981), pp. 141-170. [download]
86. Abowd, John M. "An Econometric Model of Higher Education," in *Managing Higher Education: Economic Perspectives*, A Monograph of the Center for the Management of Public and Nonprofit Enterprises (Chicago: University of Chicago Press, 1981), pp. 1-56.
87. Mulvey, Charles and John M. Abowd "Estimating the Union/Nonunion Wage Differential: A Statistical Issue," *Economica,* 47 (February 1980): 73-79.
88. Abowd, John M. and T. James Trussell "Teenage Mothers, Labor Force Participation, and Wage Rates," *Canadian Studies in Population* (1980): 33-48.

Monographs

1. Abowd, John M., William R. Bell, J. David Brown, Michael B. Hawes , Misty L. Heggeness, Andrew D. Keller, Vincent T. Mule Jr., Joseph L. Schafer, Matthew Spence, Lawrence Warren, and Moises Yi *Determination of the 2020 U.S. Citizen Voting Age Population (CVAP) Using Administrative Records and Statistical Methodology Technical Report*, U.S. Census Bureau, Center for Economic Studies, October 2020 [census.gov] [repec.org stable url]
2. Abowd, John M., Martha H. Stinson and Gary Benedetto *Final Report to the Social Security Administration on the SIPP/SSA/IRS Public Use File Project*, November 2006. [download archival copy and Excel tables at http://hdl.handle.net/1813/43929]
3. Abowd, John M. and Michael Bognanno "The Center for Advanced Human Resource Studies Managerial Compensation Database: User's Guide," March 1991.
4. Abowd, John M. and Michael Bognanno "The Center for Advanced Human Resource Studies Managerial Compensation Database: Technical Guide," March 1991.
5. Abowd, John M. *An Econometric Model of the U.S. Market for Higher Education* (New York: Garland Press, 1984).
6. Abowd, John M. and Mark Killingsworth "Employment, Wages, and Earnings of Hispanics in the federal and Nonfederal Sectors," in *Hispanics in the Labor Force: A Conference Report*, edited by G. Borjas and M. Tienda. Final Report to the National Employment Policy Commission (Washington, DC: GPO, 1982).
7. Abowd, John M. "Program Evaluation: New Panel Data Methods for Evaluating Training Effects," in *Program Evaluation Final Report to the U.S. Department of Labor* (Contract No. 23-17-80-01) (Washington, DC: NTIS, 1983).
8. Abowd, John M. and Mark Killingsworth "Employment, Wages, and Earnings of Hispanics in the federal and Nonfederal Sectors," in *Hispanics in the Labor Force: A Conference Report*, edited by G. Borjas and M. Tienda. Final Report to the National Employment Policy Commission (Washington, DC: GPO, 1982).
9. Abowd, John M. "Minority Unemployment, Compensating Differentials and the Effectiveness of the EEOC," in *Issues in Minority and Youth Unemployment final Report to the U.S. Department of Labor* (Contract No. 20-17-80-44) (Washington, DC: NTIS, 1982)
10. Abowd, John M. and Mark Killingsworth "Structural Models of the Effects of Minimum Wages on Employment by Age Groups," *Final Report of the Minimum Wage Study Commission*, Volume 5 (Washington, DC: GPO, 1981).

11. Abowd, John M. and Mark Killingsworth "An Analysis of Hispanic Employment, Earnings and Wages with Special Reference to Puerto Ricans," *Final Report to the U.S. Department of Labor*(Grant 21-36-78-61) (Washington, DC: NTIS, 1981).

Miscellany

1. Abowd, John M., Ian M. Schmutte, William Sexton, and Lars Vilhuber, Introductory Readings in Formal Privacy for Economists (May 8, 2019, updated regularly). [read, download]
2. Abowd, John M., "The Census Bureau Tries to Be a Good Data Steward in the 21st Century" International Conference on Machine Learning (ICML) 2019 keynote address. [video, start at minute 18:00] [slides]
3. Garfinkel, Simson L., John M. Abowd, and Christian Martindale, "Understanding Database Reconstruction Attacks on Public Data," *ACMQueue*, Vol. 16, No. 5 (September/October 2018): 28-53. [download, not copyrighted]
4. Garfinkel, Simson L., John M. Abowd and Sarah Powazek "Issues Encountered Deploying Differential Privacy," *WPES'18 Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, Ontario, CA (October 2018): 133-137, DOI:10.1145/3267323.3268949. [ArXiv preprint]
5. Abowd, John M. "The U.S. Census Bureau Adopts Differential Privacy," KDD '18 Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK (August 2018): 2867, DOI:10.1145/3219819.3226070. [download, subscription required], [archival copy] [video]
6. Abowd, John M., Lorenzo Alvisi, Cynthia Dwork, Sampath Kannan, Ashwin Machanavajjhala, and Jerome Reiter "Privacy-Preserving Data Analysis for Federal Statistical Agencies," Computing Community Consortium White Papers (January 2017). [CCC white paper archive; ArXiv preprint]
7. Abowd, John M. "Why Statistical Agencies Need to Take Privacy-loss Budgets Seriously, and What It Means When They Do," presented to the Federal Committee on Statistical Methodology, Policy Conference, December 7-8, 2016. [download]
8. Vilhuber, Lars, John M. Abowd and Jerome P. Reiter "Synthetic Establishment Microdata around the World," *Statistical Journal of the International Association for Official Statistics*, Vol. 32 (2016): 65-68. [download, open access] [download preprint]
9. Abowd, John M. "Synthetic Establishment Data: Origins and Introduction to Current Research," *Statistical Journal of the International Association for Official Statistics*, Vol. 30, No. 2 (Summer 2014): 113-115. [download, subscription required] [download preprint]
10. Benedetto, Gary, Martha H. Stinson and John M. Abowd "The Creation and Use of the SIPP Synthetic Beta," U.S. Census Bureau Technical Paper (April 2013). [download]
11. Abowd, John M. and Lars Vilhuber "Science, Confidentiality, and the Public Interest," *Chance*, Vol. 24, No. 3 (Fall 2011): 58-62. [download]
12. Abowd, John M. "OnTheMap: Block-level Job Estimates Based on Longitudinally Integrated Employer-Employee Micro-data," *Association of Public Data Users Newsletter* Vol. 33, No. 2 (March/April 2010): 10-19. [download]
13. Abowd, John M. Kobbi Nissim and Chris Skinner "First Issue Editorial" *Journal of Privacy and Confidentiality*, Vol. 1, No. 1 (2009): 1-6. [download]
14. Abowd, John M. "Comments on "Regional difference-in-differences in France using the German annexation of Alsace-Moselle in 1870-1918" by Matthieu Chemin and Etienne Wasmer" *NBER International Seminar on Macroeconomics* (2008): 306-309. [download]
15. Abowd, John M. and Julia Lane "The Economics of Data Confidentiality," *ICP Bulletin*, Volume 4, No. 2 (August 2007):18-21. [download preprint]
16. Abowd, John M. "Rapporteur comments: International Symposium on Linked Employer-Employee Data, Econometric Issues" *Monthly Labor Review* 121:7 (July, 1998): 52-53.
17. Abowd, John M. "Discussion of 'How much do immigration and trade affect labor market outcomes' by Geroge J. Borjas, Richard B. Freeman and Lawrence F. Katz." *Brookings Papers in Economic Activity* (1997:I): 76-82.
18. Abowd, John M. "Discussion of Gross Worker and Job Flows in Europe by M. Burda and C. Wyplosz." *European Economic Review* (1994): 1316-1320.
19. Abowd, John M. "Discussion of 'The Quality Dimension in Army Retention' by Charles Brown." in A. Meltzer (ed.) *The Carnegie-Rochester Conference on Public Policy* 33 (1990).
20. Abowd, John M. "Immigration, Trade, and Labor Markets in Australia and Canada," in *Immigration, Trade, and the Labor Market*, edited by R.B. Freeman (Cambridge, Mass: NBER,

1988), pp. 29-34.

21. Abowd, John M. "Discussion of 'Public Sector Union Growth and Bargaining Laws: A Proportional Hazards Approach with Time-Varying Treatments' by c. Ichniowski." in *Public Sector Unionism*, edited by R. Freeman (Chicago: University of Chicago Press for the NBER, 1988).
22. Abowd, John M., Ross Stolzenberg and Roseann Giarusso "Abandoning the Myth of the Modern MBA Student," Selections The Magazine of the Graduate Management Admission Council (Autumn 1986): 9-21.
23. Abowd, John M., Brent Moulton and Arnold Zellner "The Bayesian Regression Analysis Package: BRAP User's Manual Version 2.0," H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago, 1985.
24. Abowd, John M. and Mark R. Killingsworth "The Minimum Wage Law Winners and Losers," *The Wall Street Journal* (August 1981).

Working and Unpublished Papers

1. McKinney, Kevin L. and John M. Abowd, "Male Earnings Volatility in LEHD before, during, and after the Great Recession," (August 2020). [download preprint]
2. Abowd, John M., Gary L. Benedetto, Simson L. Garfinkel et al. "The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau," (August 2020). [download preprint]
3. Abowd, John M., Ian M. Schmutte, William Sexton, and Lars Vilhuber "Suboptimal Provision of Privacy and Statistical Accuracy When They are Public Goods," (June 2019). [download preprint]
4. Abowd, John M., Joelle Abramowitz, Margaret C. Levenstein, Kristin McCue, Dhiren Patki, Trivellore Raghunathan, Ann M. Rodgers, Matthew D. Shapiro, Nada Wasi, 2019. "Optimal Probabilistic Record Linkage: Best Practice for Linking Employers in Survey and Administrative Data," Working Papers 19-08, Center for Economic Studies, U.S. Census Bureau, handle: RePEc:cen:wpaper:19-08. [download preprint]
5. McKinney, Kevin L. Andrew Green, Lars Vilhuber, and John M. Abowd "Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in On The Map" (December 2017). [download preprint]
6. Abowd, John M. and Ian Schmutte "Revisiting the Economics of Privacy: Population Statistics and Confidentiality Protection as Public Goods" (April 2017), [download preprint], published as Abowd, John M. and Ian M. Schmutte "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review,* Vol. 109, No. 1 (January 2019):171-202, DOI:10.1257/aer.20170627. [AER, ArXiv preprint, Replication information]
7. Abowd, John M. "Where Have All the (Good) Jobs Gone? (May 2014) Society of Labor Economists Presidential Address. [download preprint] [accompanying audio]
8. Abowd, John M., John Haltiwanger, Julia Lane, Kevin McKinney and Kristin Sandusky "Technology and Skill: An Analysis of Within and Between Firm Differences" (March 2007) NBER WP-13043. [download preprint]
9. Abowd, John M., Francis Kramarz, David N. Margolis, and Thomas Philippon "Minimum Wages and Employment in France and the United States" (February 2006). [archival download]
10. Abowd, John M., Paul Lengermann and Kevin L. McKinney "The Measurement of Human Capital in the U.S. Economy," (March 2003) [download Census, cited on September 1, 2015] [archival download]
11. Abowd, John M., Robert Creecy and Francis Kramarz "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," (March 2002). [download Census, cited on September 1, 2015] [archival download] [Fortran source] [Support files] [VirtualRDC archive]

MAJOR GRANTS AND RESEARCH CONTRACTS

1. Associate Director for Research and Methodology and Chief Scientist U.S. Census Bureau, Intergovernmental Personnel Act (IPA) with Cornell University, June 1, 2016—March 27, 2020.
2. Research and Methodology Support Services, U.S. Census Bureau contract with Cornell University, June 1, 2015—May 31, 2016, $268,897.

3. The Economics of Socially Efficient Privacy and Confidentiality Management for Statistical Agencies, Alfred P. Sloan Foundation awarded to Cornell University, April 1, 2015—March 31, 2019, $535,970. (co-PIs Lars Vilhuber and Ian Schmutte)
4. RCN: Coordination of the NSF-Census Research Network, National Science Foundation SES 1237602 awarded to the National Institute of Statistical Sciences, July 15, 2012—June 30, 2017, transferred to Cornell University, September 2014, $748,577. (PI Lars Vilhuber, other co-PIs Alan Karr, Jerome Reiter)
5. NCRN-MN: Cornell Census-NSF Research Node: Integrated Research Support, Training and Data Documentation, National Science Foundation Grant SES 1131848 awarded to Cornell University, October 1, 2011—September 30, 2016, $2,999,614. (with William Block, Ping Li, and Lars Vilhuber)
6. A Census-Enhanced Health and Retirement Study: A Proposal to Create and Analyze an HRS Dataset Enhanced with Characteristics of Employers, Alfred P. Sloan Foundation grant awarded to the Institute for Social Research, University of Michigan with a subcontract to Cornell University, September 1, 2011—August 31, 2016, Cornell component $349,608. (PI: Margaret Levenstein; other co-PIs: Matthew Shapiro, Kristin McCue and David Weir)
7. Synthetic Data User Testing and Dissemination, National Science Foundation Grant SES 1042181 awarded to Cornell University, September 15, 2010 to September 14, 2013, $197,170. (Co-PI Lars Vilhuber)
8. CDI-Type II: Collaborative Research: Integrating Statistical and Computational Approaches to Privacy, National Science Foundation Grant BCS 0941226 awarded to Cornell University, September 1, 2010—August 31, 2014, $409,296. (Other PIs: Aleksandra B Slavkovic, Stephen E. Fienberg, Sofya Raskhodnikova, and Adam Smith)
9. TC:Large: Collaborative Research: Practical Privacy: Metrics and Methods for Protecting Record-level and Relational Data, National Science Foundation Grant TC 1012593 awarded to Cornell University, July 15, 2010 to July 14, 2015, $1,326,660. (Other PIs: Johannes Gehrke, Gerome Miklau, and Jerome Reiter)
10. The Longitudinal Employer-Household Dynamics Program, U.S. Bureau of the Census, Interagency Personnel Act (IPA) with Cornell University, September 18, 1998 – September 17, 2000, $260,000; renewed September 14, 2000—September 13, 2002, $320,000; contract renewed as consultant September 14, 2002—September 13, 2003 ($120,000); renewed as IPA September 15, 2003 – September 14, 2005 ($384,590); renewed as IPA September 15, 2005—September 14, 2007 ($425,215); new September 15, 2008—September 14, 2010 (497,897); renewed September 15, 2010—September 14, 2012 (532,893); continued as a contract with ACES-Research, LLC (September 17, 2012–September 16, 2013); re-established as IPA October 1, 2013—September 30, 2014 ($231,757); re-established as IPA November 14, 2014—May 31, 2015 ($229,095).
11. Social Science Gateway to TeraGrid, National Science Foundation Grant SES 0922005 awarded to Cornell University, July 1, 2009 to June 30, 2012, $393,523. (Co-PI Lars Vihuber) [Cornell Chronicle Article] [ILR News Release]
12. Joint NSF-Census-IRS Workshop on Synthetic Data and Confidentiality Protection, July 2009 Washington, DC, National Science Foundation Grant SES 0922494 awarded to Cornell University, July 1, 2009 to June 30, 2010, $18,480. (Co-PIs Lars Vilhuber, Jerome Reiter, and Ron Jarmin)
13. The Economics of Mass Layoffs: Displaced Workers, Displacing Firms, Causes and Consequences, National Science Foundation Grant SES-0820349 awarded to Cornell University, October 1, 2008 to September 30, 2010, $245,950. (Co-PI Lars Vilhuber)
14. LEHD Developmental and Confidentiality Research, Census Bureau Contract to Abt Associates with subcontract awarded to Cornell University, August 1, 2007 to September 30, 2008, $358,270.
15. CT-T: Collaborative Research: Preserving Utility While Ensuring Privacy for Linked Data, National Science Foundation Grant CNS-0627680 awarded to Cornell University, September 5, 2006 to August 31, 2009, $488,950. (PI Johannes Gehrke)
16. LEHD Confidentiality Research, Census Bureau Contract to Abt Associates with subcontract awarded to Cornell University, October 1, 2004 to September 30, 2005, $230,155.
17. ITR-(ECS+ASE)-(dmc+int): Info Tech Challenges for Secure Access to Confidential Social Science Data, National Science Foundation Grant SES-0427889 awarded to Cornell University, October 1, 2004 to September 30, 2007, $2,938,000. (Co-PIs Matthew D. Shapiro, Ronald Jarmin, Stephen F. Roehrig, and Trivellore Raghunathan) [Cornell Chronicle article]

18. EITM: Developing the Tools to Understand Human Performance: An Empirical Infrastructure to Foster Research Collaboration, National Science Foundation Grant SES-0339191 awarded to Cornell University, October 1, 2004 to September 30, 2007, $337,455 (Co-PIs John Haltiwanger and Ron Jarmin)
19. The New York Research Data Center, National Science Foundation Grant SES-0322902 awarded to the NBER, August 1, 2003 to July 31, 2004, $300,000. (PI Neil G. Bennett, Other co-PIs Bart Hobijn, Erica L. Groshen, Robert E. Lipsey)
20. Workshop on Confidentiality Research, National Science Foundation Grant SES-0328395 awarded to the Urban Institute, June 1, 2003 – May 31, 2004, $43,602. (Co-PI Julia Lane)
21. Firms, Workers and Workforce Quality: Implications for Earnings Inequality and Economic Growth, Alfred P. Sloan Foundation Grant 22319-000-00 awarded to the Urban Institute, January 2003—January 2006, $1,400,000. (Co-PIs John Haltiwanger, Julia Lane, J. Bradford Jensen, Fredrick Knickerbocker, and Ronald Prevost)
22. The Demand for Older Workers: Using Linked Employer-Employee Data for Aging Research, National Institute on Aging, R01-AG18854-01 to Cornell University, July 1, 2002 – April 30, 2007, $1,753,637. (Co-PIs John Haltiwanger, Andrew Hildreth, and Julia Lane)
23. Workers and Firms in the Low-wage Labor Market: Interactions and Long Run Dynamics, Russell Sage Foundation, Rockefeller Foundation, and Department of Health and Human Services (ASPE) to the Urban Institute $700,000, September 1, 2001 August 31, 2003. (Co-PIs John Haltiwanger, Harry Holzer, and Julia Lane)
24. From Workshop Floor to Workforce Clusters: A New View of the Firm, Alfred P. Sloan Foundation, 99-12-12 to the Urban Institute, March 1, 2000 – March 31, 2002, $314,604. (Co-PIs John Haltiwanger and Julia Lane)
25. Dynamic Employer-Household Data and the Social Data Infrastructure, National Science Foundation, SES-9978093 to Cornell University, September 28, 1999 – September 27, 2005, $4,084,634. (Co-PIs John Haltiwanger and Julia Lane)
26. The Longitudinal Employer-Household Dynamics Program, National Institute on Aging, interagency funding to the United States Census Bureau, September, 1999 – August, 2001, $490,000. Renewed September 2001– August 2004, $750,000 (Co-PIs John Haltiwanger and Julia Lane) [Cornell Chronicle article]
27. Individual and Firm Heterogeneity in Labor Markets: Studies of Matched Employee-Employer Data, National Science Foundation SBER 9618111 to the NBER, March 15, 1997 – February 28, 2002, $243,361.
28. Creation of an Employer Identification Link File and Addition of Employer Information to the National Longitudinal Survey of Youth 1979 Cohort, Bureau of Labor Statistics (subcontracted by NORC, University of Chicago, Chicago, IL 60637), July 1, 1995 – December 31, 1997, $82,946.
29. Employment and Compensation Policies: Studies of American and French Labor Markets Using Matched Employer-Employee Data, National Science Foundation SBR 9321053 to the NBER, July 1, 1994 – June 31, 1997, $ 185,257. (Co-PIs David Margolis and Kenneth Troske)
30. Compensation System Design, Employment and Firm Performance: An Analysis of French Microdata and a Comparison to the United States, National Science Foundation, SBR 9111186 to Cornell University, July 1, 1991 – December 30, 1994, $174,565.
31. The Effects of Collective Bargaining and Threats of Unionization on Firm Investment Policy, Return on Investment, and Stock Valuation, National Science Foundation, SES 8813847 to the NBER, July 1, 1988 – June 30, 1990, $81,107.
32. Improving the Scientific Research Utility of Labor Force Gross Flow Data, National Science Foundation, SES 85-13700 to the NBER, April 15, 1986 – March 31, 1988, $69,993.
33. Program Evaluation: New Panel Data Methods for Evaluating Training Effects, U.S. Department of Labor Contract 23-17-80-01 to NORC at the University of Chicago, 1983.
34. Minority Unemployment, Compensating Differentials and the Effectiveness of the EEOC, U.S. Department of Labor Contract 20-17-80-44 to NORC at the University of Chicago, 1982.
35. An Analysis of Hispanic Employment, Earnings and Wages with Special Reference to Puerto Ricans, U.S. Department of Labor Grant 21-36-78-61, 1981.

PROFESSIONAL SERVICE, SURVEYS, AND DATA COLLECTION

1. Canadian Research Data Centre Network Inaugural Board 2017-2019.
2. American Economic Association, Committee on Economic Statistics (AEAWeb) 2013-2018.

3. National Academy of Sciences, Committee on National Statistics (CNSTAT) 2010-2013; reappointed 2013-2016.
4. National Academy of Sciences, CNSTAT, Panel on Measuring and Collecting Pay Information from U.S. Employers by Gender, Race, and National Origin, (Chair) 2011-2012.
5. National Academy of Sciences, CNSTAT, Panel on Measuring Business Formation, Dynamics and Performance, 2004-2007.
6. National Academy of Sciences, CNSTAT, Panel on Data Access for Research Purposes, 2002-2005.
7. Executive Committee, Conference on Research in Income and Wealth 2002-.
8. Distinguished Senior Research Fellow, LEHD Program, U.S. Census Bureau 1998-2016.
9. Social Science and Humanities Research Council (Canada), Major Collaborative Research Initiatives review panel, 1997, 1998.
10. Technical Advisory Board for the National Longitudinal Surveys of the Bureau of Labor Statistics, 1988-1990, 1992-2001, Chair 1999-2001.
11. National Science Foundation, Economics Panel, 1990-91, 1992-93; KDI Panel 1999; Infrastructure Panel 2000; CDI Panel 2008; CDI Panel 2009.
12. Principal Investigator for The Center for Advanced Human Resource Studies Managerial Compensation Data Base. sponsored by the Cornell University Center for Advanced Human Resource Studies, 1989-1994.
13. Principal Investigator for A Longitudinal Data Base of Collective Bargaining Agreements. Sponsored by the Bureau of National Affairs and the University of Chicago Graduate School of Business, 1985.

PROFESSIONAL ORGANIZATIONS

1. American Economic Association
2. American Statistical Association
3. Econometric Society
4. Society of Labor Economists
5. International Statistical Institute
6. International Association for Official Statistics
7. National Association for Business Economics
8. American Association of Wine Economists
9. American Association for Public Opinion Research
10. Association for Computing Machinery
11. American Association for the Advancement of Science

PERSONAL INFORMATION

United States citizen
Personal email: john.abowd@gmail.com

Hosted by CampusPress

Appendix B

APPENDIX B — 2010 RECONSTRUCTION-ABETTED RE-IDENTIFICATION SIMULATED ATTACK

1. This appendix provides a high-level summary of the reconstruction-abetted re-identification attack simulation that the Census Bureau conducted on the released 2010 Census data. To assess the risk of a reconstruction-abetted re-identification attack, the Census Bureau conducted a series of statistical exercises to quantify the contemporaneous and future risk that individual responses could be disclosed. The Census Bureau has completed two simulated attacks that address the re-identification risk of a 100% microdata file (a file with detailed, individual-level records for every person enumerated in the census) reconstructed from the published Summary File 1 data. The 2010 Summary File 1, usually called SF1, includes the 2010 P.L. 94-171 Redistricting Data Summary File, the 2010 Advanced Group Quarters Data Summary File, and the bulk of the demographic and housing characteristics released from the 2010 Census in tabular format.[1] The fundamental structure of these simulations is as follows.

SIMULATED RECONSTRUCTION ATTACK

2. Database reconstruction is the process of statistically re-creating the individual-level records from which a set of published tabulations was originally calculated. That is, database reconstruction attempts to "reverse engineer" the confidential input data used in a statistical tabulation system.
3. The Census Bureau released over 150 billion statistics as part of the 2010 Census. The simulated reconstruction attack used as its input a small fraction of those statistics — approximately 6.2 billion statistics contained in the following published SF1 tables from the 2010 Census:

    P001 (Total Population by Block)
    P006 (Total Races Tallied by Block)
    P007 (Hispanic or Latino Origin by Race by Block)

[1] See the technical documents in Summary File 1 Dataset (census.gov).

P009 (Hispanic or Latino, and Not Hispanic or Latino by Race by Block)
P011 (Hispanic or Latino, and Not Hispanic or Latino by Race for the Population 18 Years and Over by Block)
P012 (Sex by Age by Block)
P012A-I (Sex by Age by Block, iterated by Race)
P014 (Sex by Single-year-of-age for the Population under 20 Years by Block)
PCT012A-N (Sex by Single-year-of-age by Tract, iterated by Race)

4. The reconstruction of the 2010 Census microdata for the sex, age, race, Hispanic/Latino ethnicity, and census block variables was carried out by constructing a system of equations consistent with the published tables listed above that, once solved, could then be converted into microdata. This system of equations was solved using commercial mixed-integer linear programming software (Gurobi).
5. Because the parameters of the 2010 Census swapping methodology included invariants on total population and voting age population at the block level, the reconstruction was able to exactly reconstruct all 308,745,538 million records with correct block location and voting age (18+). Then, leveraging the race (63 categories), Hispanic/Latino origin, sex, and age (in years) data from the specified tables, the simulated attack was able to further reconstruct those variables on the individual-level records.
6. To assess the accuracy of these reconstructed individual-level records, the team performed exact record linkage of the five variables in the reconstructed microdata to the same five variables in the Census Edited File (CEF, the confidential data) and Hundred-percent Detail File (HDF, the confidential swapped individual-level data before tabulation). The results are summarized in Table 1. The "left" file of the record linkage is in the first column. The "right" file is the reconstructed microdata from SF1.

Table 1
Agreement Rates between the Reconstructed Microdata
and the 2010 Census Edited File and Hundred-percent Detail File

	Record Counts		Agreement Rates		
Left file	In Left	In Reconstructed	Exact	Fuzzy Age	One error
CEF	308,745,538	308,745,538	46.48%	70.98%	78.31%
HDF	308,745,538	308,745,538	48.34%	73.33%	80.39%
DRB clearance number CBDRB-FY21-DSEP-003					

7. The agreement rates shown in Table 1 include block (which was never wrong), sex, age (in years), race (63 OMB categories), and Hispanic ethnicity and are computed as a percentage of the total population. Exact agreement means all five variables agreed precisely bit for bit. Fuzzy-age agreement means that block, sex, race, and Hispanic ethnicity agreed exactly, but age agreed only +/- 1 year (e.g., age 25 on the CEF is in fuzzy-age agreement with ages 24, 25, and 26 on the reconstructed data). The one-error agreement rate allows one variable—sex, age (outside +/- one year), race or ethnicity to be wrong.
8. Most errors in the reconstructed file are that the age variable is off by +/- 2 years rather than +/- 1 year. This error is the balance of the width of the 5-year categories used in the block-level summaries. Hence, even though the disclosure avoidance requirement for the 2010 Census SF1 tabular summaries specified block-level aggregation to 5-year bins for those age 20 and over, the effective aggregation was far less.
9. Figure 1 shows the distribution of agreement rates by block size. Agreement rates are only substantially lower than the population averages shown in Table 1 for blocks with populations between 0 and 9 people, which is where the Census Bureau has said it concentrated the swaps.[2] However, uniqueness on sex, age, race, and ethnicity is

[2] McKenna, L. (2018), "Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing," https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf,p. 8.

not limited to small population blocks. *This is one of the principal failures of the 2010 tabular disclosure avoidance methodology – swapping provided protection for households deemed "at risk," primarily those in blocks with small populations, whereas for the for the entire 2010 Census a full 57% of the persons are population uniques on the basis of block, sex, age (in years), race (OMB 63 categories), and ethnicity. Furthermore, 44% are population uniques on block, age and sex.*[3]

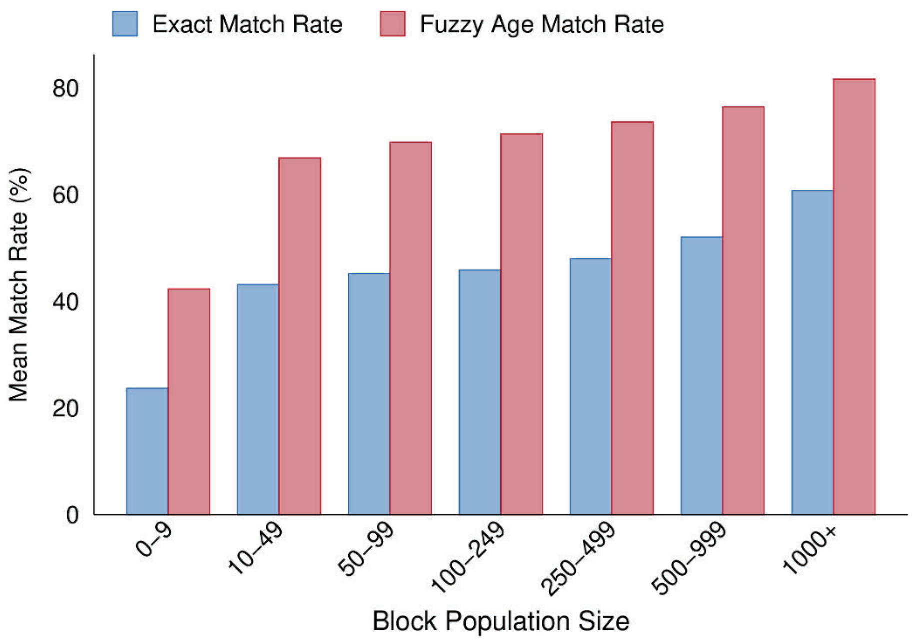

Figure 1 Block-level agreement rates between the reconstructed 2010 Census micro-data and the 2010 Census Edited File by population in the block
DRB clearance number CBDRB-FY21-DSEP-003.

10. Although there are no recent re-identification studies for decennial Public Use Micro-data Samples (PUMS) with geography coded to the Public Use Microdata Area

[3] The statistics in this paragraph are cleared for public release by the Census Bureau Disclosure Review Board (CBDRB-FY21-DSEP-003).

(PUMA), the Census Bureau continues to use 100,000 persons as the minimum population threshold for such areas and has coded geography on the 2010 PUMS and all American Community Survey (ACS) PUMS using these PUMAs. Since sex and age (single years) are population uniques at the tract level for only 0.18% of persons, this may still be justifiable for a 10% sample of 2010 Census records, but the potential reidentification rate for a 100% public-use microdata file geocoded to the block level is certainly quite large.

11. The reconstruction experiment demonstrated that existing technology can convert the Census Bureau's traditional tabular summaries of Census data which was released in 2010 into a 100% coverage microdata file geocoded to the block level with very limited noise which was not released in 2010. This microdata file contains so much detail that it would have been deemed "unreleasable" if it had been proposed in conjunction with the original 2010 Census data products.

12. The ability to reconstruct the microdata means that there is now a significant disclosure risk for the 2010 Census Summary Files 1 and 2 (SF1, SF2) and the American Indian Alaska Native Summary File (AIANSF) data. There are approximately 150 billion statistics in the SF1, SF2, and AIANSF summaries (recall that the 2010 P.L. 94-171 Redistricting Data Summary File and the 2010 Advanced Groups Quarters Summary File are part of SF1). Because of the features noted above, releasing this many very accurate statistics made the ensemble of those publications equivalent to releasing the 2010 Hundred-percent Detail File (HDF), the swapped version of and the 2010 Census Edited File (CEF). There can be no uncertainty about this: *the 2010 Census tabular publications were equivalent to releasing every tabulation variable in the 2010 HDF in universe public-use microdata files without the hierarchical structure--person and household records can be fully reconstructed, but not directly linked to each other*. The team that demonstrated this vulnerability stopped after reconstructing person-level records for block, sex, age

(in years), race (63 OMB categories), and Hispanic ethnicity because the vulnerability had been fully exposed mathematically and demonstrated empirically.

13. There are 308,745,538 (U.S. only) person records and 131,704,730 housing unit records in both the 2010 HDF and CEF, linked in their correct hierarchy. For the unswapped records in HDF, the images are identical to their CEF counterparts. For the swapped household records, the block identifier, household size, adult (age 18+) household size, occupancy, and tenure variables are identical to their unswapped counterparts and on the person record the voting-age variable is identical to the unswapped counterpart.

14. A public-use microdata file containing the 308,745,538 person records in the HDF including only the five tabulation variables block, sex, age (in years), race (63 OMB categories), and Hispanic ethnicity is so disclosive that it would not have passed the disclosure avoidance criteria used for the 2010 Census Public-Use Microdata Sample.[4] Furthermore, the same file would not have passed the disclosure avoidance criteria applied to SF1 itself.[5] The official 2010 PUMS had a geographic population threshold of 100,000, collapsed categories to national population thresholds of 10,000, used partially synthetic data for the group quarters population, and "topcoding, bottom-coding, and noise infusion for large households." The PUMS was sampled from the swapped version of the 2010 HDF, not the Census Edited File.

[4] McKenna, L. (2019a) "Disclosure Avoidance Techniques Used for the 1960 Through 2010 Decennial Censuses of Population and Housing Public Use Microdata Samples," Research and Methodology Technical Report available at Disclosure Avoidance Techniques Used for the 1960 Through 2010 Census.; McKenna, L. (2018)

[5] McKenna, L. (2018)

15. The additional disclosure avoidance methods used for the 2010 PUMS are explicitly noted on pages 2-1 and 2-2 of its technical documentation. The definition of a Public Use Microdata Area also explicitly references its confidentiality protection purpose:

> "The Public Use Microdata Sample (PUMS) files contain geographic units known as Public Use Microdata Areas (PUMAs). To maintain the confidentiality of the PUMS data, a minimum population threshold of 100,000 is set for PUMAs. Each state is separately identified and may be comprised of one or more PUMAs. PUMAs do not cross state lines. (page 1-2, emphasis added)"

16. This failure to apply microdata disclosure avoidance matters because the reconstructed 2010 microdata for block, sex, age (in years), race (63 OMB categories), and Hispanic ethnicity are a very accurate image of the HDF, and the HDF is a very accurate image of the CEF, which is the reason that it is also confidential. Consequently, the new technology-enabled possibility of accurately re-constructing HDF microdata from the published tabular summaries and the fact that those reconstructed data do not meet the disclosure avoidance standards established at the time for microdata products derived from the HDF demonstrate that the swapping methodology as implemented for the 2010 Census no longer meets the acceptable disclosure risk standards established when that swapping mechanism was selected for the 2010 Census.
17. Having demonstrated that a 100% microdata file can be successfully reconstructed from the published 2010 Census tabulations, the Census Bureau proceeded to use these reconstructed microdata to simulate a re-identification attack on those data.

De-identification attack simulation

18. The simulated re-identification attack proceeds as follows. Identify a person-level data source file that contains name, address, sex, and birthdate (e.g., commercially available data). Convert the names and addresses to their corresponding Census Bureau Protected Identification Key (PIK). Identify the corresponding census block for

every address in the source file. Then, looping through all the records in the reconstructed microdata file produced from the reconstruction, find the first record in the source file that matches exactly on block, sex, and age. Once this step is completed, run through the remaining unmatched records from the reconstructed microdata and find the first unmatched record from the source file that matches exactly on block and sex, and matches on age plus or minus 1 year.

19. When both steps have been completed, output the records with successful matches from these two passes. These are called *putative re-identifications* because they appear to link the reconstructed microdata to a real name and address associated with the block, sex, age, race, and ethnicity on the reconstructed microdata. These are the records the hypothetical attacker thinks are re-identified.

20. Putative re-identifications are not necessarily correct. An external attacker would have to do extra field work to estimate the *confirmation rate*—the percentage of putative re-identifications that are correct. An external attacker might estimate the confirmation rate by contacting a sample of the putative re-identifications to confirm the name and address. An external attacker might also perform more sophisticated verification using multiple source files to select the name and address most consistent with all source files and the reconstructed microdata.

21. At the Census Bureau we usually estimate the confirmation rate as a percentage of the total population, not as a percentage of the putative re-identifications, by performing a similar record linkage exercise of the putative re-identifications against the CEF, looking for exact matches on all variables (including PIK, block, sex, age, race, and ethnicity), followed by a second pass looking for exact matches except age, which is allowed to vary by plus or minus 1 year. Once these two passes have been completed, the matched records are the confirmed re-identifications, using exact match on PIK, block, sex, race (63 OMB categories), and ethnicity and match on age +/- 1 year as the

definition of correct. The remaining unmatched records from the putative re-identifications of the reconstructed data are the unconfirmed re-identifications.

22. Table 2 shows the results of two such re-identification confirmation exercises. The first of these uses the combined commercial databases from Experian Marketing Solutions Incorporated, Infogroup Incorporated, Melissa Data Corporation, Targus Information Corporation, and VSGI LLC as the source file for name, address, sex, and age. This exercise simulates data quality circa 2010 for an external attacker relying on the consumer information in these databases. These results are in the row labeled "Commercial." This re-identification experiment was the basis for the statistics released at the American Association for the Advancement of Science 2019 annual meeting. Putative re-identifications were 138 million (45% of the 2010 Census resident population of the U.S.). Confirmed re-identifications were 52 million (17% of the same population).

23. Using the commercial data as the source for name, address, sex, and age is, as discussed in the main declaration, a best-case assumption. We know that these data exist and were available circa 2010 because that is when the Census Bureau acquired them. An external attacker, using the versions that the Census Bureau acquired and the relatively straightforward methodology above, would succeed at least as often as we did. This means that at least 52 million persons enumerated during the 2010 Census could be correctly re-identified using the attack strategy outlined here.

24. Suppose the external attacker had name, address, sex, and age of much better quality than the five commercial sources above. How much better could that attacker do using exactly the same strategy? This question can be answered by substituting the name, address, sex, and age from the 2010 CEF as the source file in the putative re-identification simulation. This is not cheating because no extra information in the CEF such as race, ethnicity or household structure is used for the source file. Hence, it is a proper worst-case scenario, and the one historically used by the Census Bureau in

assessing microdata re-identification risk.[6] If the external data on name, address, sex, and age are comparable to the 2010 Census, then the attacker will putatively re-identify 238 million persons (77% of the 2010 Census resident U.S. population). Confirmed re-identifications will be 179 million (58% of the same population). This means that with the best quality external data, relative to the 2010 Census, as many as 179 million persons could be correctly re-identified using the attack strategy outlined here.

Table 2
Record Linkage Summary from Commercial and CEF Record Sources

PIK, Block, Age, Sex Record Linkage **Source**	Available Records	Records with PIK, Block, Sex, and Age	Putative Re-identifications using **Source**	Confirmed Re-identifications
Commercial	413,137,184	286,671,152	137,709,807	52,038,366
CEF	308,745,538	279,179,329	238,175,305	178,958,726
DRB clearance number CBDRB-FY21-DSEP-003.				

25. The record linkage results reported in Table 2 can be interpreted using two additional statistical quality measures: the *recall rate* and the *precision rate*. Taken together, these measures assess how successful an attacker can be at re-identifying records and how confident the attacker would be in those re-identifications.

26. *Recall rate.* The recall rate is the percentage of available source records that are correctly re-identified. Its numerator is the same as the confirmation rate, but its denominator is the number of records in the source file with sufficient information to perform the putative re-identification record linkage. For the two source files analyzed in these

[6] McKenna, L. (2019b). "U.S. Census Bureau Reidentification Studies," available at https://www.census.gov/library/working-papers/2019/adrm/2019-04-ReidentificationStudies.html.

experiments, Table 2 shows the denominators for the recall rate in the column "Records with PIK, Block, Sex, and Age," which gives the count of records with sufficient information to generate a putative match. Table 3 shows the recall rates for the two experiments. Both are greater than the respective confirmation rate because both the commercial data and the CEF have fewer usable records than the U.S. resident population. A critical result is the recall rate of 64% when the CEF is used as the source file. This result means that an external attacker with high quality name, address, sex, and age information succeeds in re-identification almost two times in three.

Table 3
Confirmation and Recall Rates

Source	Percentage of U.S. Resident Population (Confirmation Rate)	Percentage of Complete Data Population (Recall Rate)
Commercial	16.85%	18.15%
CEF	57.96%	64.10%
DRB clearance number CBDRB-FY21-DSEP-003.		

27. *Precision rate.* Precision is the ratio of confirmed to putative re-identifications. It answers the question "How often is the attacker's claimed re-identification correct as a percentage of the names the attacker attached to reconstructed census microdata?" Table 4 summarizes the precision rates for the two experiments. The precision of the experiment reported in February 2019 was 38% (first row of Table 4). The precision of the worst-case experiment is 75% (second row of Table 4). *This result means that an attacker using high-quality name, address, sex, and age data is correct three times out four.*

Table 4 Precision Rates	
Source	Confirmed Percentage of Putative Re-identification (Precision Rate)
Commercial	37.79%
CEF	75.14%
DRB Clearance number CBDRB-FY21-DSEP-003.	

28. To be successful, an attacker does not have to be a commercial entity, nor does a successful attack need to use commercially available data. Many agencies of federal, state and local governments in the U.S. now possess high-quality data on name, address, sex, and age. When preparing public-use microdata files that contain variables that other agencies can access exactly, it has long been the practice to coarsen such data to prevent non-statistical uses by other agencies.[7] Applying such precautions to decennial census data products would imply severe limitations on the variables published at the block level, even in the presence of swapping.

29. Since the influential work of Duncan and Lambert, [8] the risk of identity disclosure for a microdata record has been measured by the probability that the record is a population unique on key variables that can be used for record linkage to external data. Population uniques have a combination of key characteristics that occurs exactly once in the entire population. The most basic set of key variables is location, sex and age. A more extensive set is location, sex, age, race and ethnicity.

[7] see McKenna 2019b.

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

30. Skinner and Shlomo[9] use population census data from the United Kingdom to demonstrate how to estimate the risk that a record in a sample corresponds to a population unique in the census and, therefore, requires active disclosure limitation. In all disclosure limitation systems designed since Fellegi[10] invented the discipline, records containing population uniques on key variables are the highest risk records for re-identification and receive direct disclosure avoidance protection: suppression, coarsening categories to eliminate uniqueness, noise infusion or some combination of these. Skinner and Shlomo had to predict the probability that a sample record was a population unique because, depending on the sampling rate, records that are unique on the variables in the sample may have many duplicates in the population. They used the UK census to validate their prediction model.

31. Fifty-seven percent of the 308,745,538 person records in the confidential 2010 Census Edited File, the definitive source for all 2010 Census tabulations, were unique on their block location, sex, age (in years), race (any combination of the 6 OMB-approved race categories, 63 possibilities in all) and Hispanic/Latino ethnicity.[11] In the case of the reconstructed 2010 Census microdata, we know the probability that a record is unique—no estimation is necessary. As shown in Table 5, 44% of all persons in the reconstructed data are population uniques on three key linkage variables: location (census block code), sex, and age (in years). The high proportion of population uniques on these three variables make the reconstructed data vulnerable to a classic

---

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

[10] Fellegi, I. P. (1972). On the question of statistical confidentiality. Journal of the American Statistical Association, 67(337).

[11] This previously confidential statistic was approved for publication with DRB clearance number CBDRB-FY21-DSEP-003

record linkage attack identical to the one modeled by Duncan and Lambert and by Skinner and Shlomo resulting in a re-identification, when the attacker knows the name of the person associated with the location, sex and age. This is exactly the definition of a re-identification used by McClure and Reiter[12] and by Wasserman and Zhou.[13] This risk assessment is derived from conventional statistical disclosure limitation methods, not differential privacy accounting.

Table 5
Distribution of Population and Population Uniques by Block Population Size

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

DRB clearance number CBDRB-FY21-DSEP-003

32. Table 5 is based on the actual 2010 Census, not simulated data like some of the cited studies use. It uses the exact distribution of block populations found in the official 2010 Census data and the actual responses on the 2010 Census. The table shows the

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

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

distribution of the population by the size of the block where the person resides. Only 2.61% of the population lives in blocks with 1 to 9 persons. This is significant because these very small blocks are the ones most likely to be protected by the 2010 Census swapping method. 21.89% of the population live in blocks with 10 to 49 residents, and 22.37% live in blocks with 50 to 99 persons. Fully 46.88% of the population lives in a block with fewer than 100 residents. The column labeled "Percent of (block, sex, age) Uniques in Bin" shows the percentage of the residents of the block who are unique in their census block, sex and age (in years) values. This percentage ranges from almost everyone (95.06%) in the least populous blocks to very few (1.12%) in the most populous blocks. There are no simulated or reconstructed data used in this table. These are characteristics of the 2010 Census resident population as they appear in the 2010 Census Edited File (CEF). [14]

33. At the Census Bureau, the existence of documented population uniques, even one—not to mention 135 million—triggers mandatory active disclosure limitation, as documented in McKenna.[15] If presented with a proposed public-use microdata file containing the variables: census block, sex, age (in years), race (OMB-designated coding), and ethnicity (OMB-designated coding) in 1990, 2000, 2010, or 2020, the Census Bureau Disclosure Review Board (or its predecessor) would have insisted on aggregation of the census block codes into more populous geographic areas and would have im-

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

[15] McKenna (2019b).

posed minimum population sizes (at least 100,000) and minimum population thresholds for the race and ethnicity coding. It would also have insisted on sampling, as documented in McKenna.[16]

34. Table 5 shows that the reconstructed data would be subject to Census Bureau Disclosure Review Board regulation because they contain known population unique identifiers (the combination of census block, sex and age in years). They were produced using tabulations from a confidential Census Bureau data file—the swapped version of the CEF. And they are in record-level format with one record for every person enumerated in the 2010 Census. In their present form, they would not have been certified for release in 2011, when the other 2010 Census data products were released, nor were they certified for release in 2019, when the Census Bureau performed the full reconstruction—even though any person anywhere in the world can perform the same reconstruction because the tables *were* approved for release. The reconstructed 2010 Census data present a clear and present disclosure risk based on the in-place standards of the Census Bureau, which predate differential privacy by several decades. They also present a clear and present disclosure risk using the traditional methods of assessing such risks, as initiated by Duncan and Lambert, refined by Skinner and Shlomo, and analyzed by the methods used in McClure and Reiter.
35. The traditional standard for applying disclosure limitation methods to microdata is based on the *existence of known unique identifier combinations* in the tabulation variables—census block, sex and age in years, in this case—*not their efficacy in abetting re-identification*. Statistical agencies are expected to document the uniqueness of the identifier—that is in Table 5 above—and to continually assess the adequacy of the proposed disclosure limitation methods. Such assessments often involve re-identification

[16] McKenna (2019a).

studies. Such studies inform the strength of the traditional disclosure limitations applied.

36. Armed with the knowledge that the reconstructed data contain population uniques, an external agent can also conduct fieldwork or reference multiple commercially available data sources to verify that re-identifications based on these uniques are valid. But even if the external agent could not, it is the agency's duty to protect the confidentiality of the microdata and therefore it must, just as in cybersecurity, assume that attackers are clever enough to gather information that confirms the efficacy of their attacks.

37. Table 6 shows that the reconstruction-abetted re-identification attack simulated by the Census Bureau has very high precision precisely in the blocks that are most vulnerable to such an attack, whether one uses the best-case or worst-case analysis. In blocks with populations between 1 and 9 persons, the re-identification attack has a precision of 72.24% when using commercial data available in 2010.[17] The exact block population was public information following the release of the 2010 Census data. That means an attacker has a clean, public predictor of the success of the re-identification attack. Fieldwork in sparsely populated blocks can confirm this precision, as can sophisticated Bayesian methods like entity resolution without field work.[18] If the attacker has better quality name, address, sex and age data than were available in 2010, certainly a plausible assumption, then the worst-case analysis for blocks with populations of 1 to 9 is precision of 96.98%--more precise than the 95% confidence interval test often

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

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

used in statistics. Again, this can be confirmed by fieldwork or Bayesian entity resolution. The situation is only a little better for the 68 million people who live in blocks with populations of 10 to 49. The precision of the 2010-era commercial data is 53.61%--correct more than half the time, and the precision with high-quality external data is 91.68%. Although the best-case precisions for block populations of 50 or more are less than one-half, the worst-case precision, even in the most populous blocks, is always greater than one-half—*an attacker with high quality external data is always more likely to be correct than wrong.* With high-quality data, the attacker is correct on average three times out of four regardless of the number of persons who live in the block.

Table 6
Disclosure Risk Assessment of Population Uniques by Block Population Size

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

DRB clearance number CBDRB-FY21-DSEP-003.

38. The Data Stewardship Executive Policy Committee (DSEP) determined that the simulated attack success rates in Table 6 were unacceptable for the 2020 Census. Decennial census data protected by the 2010 disclosure avoidance software is no longer safe to release.

39. In conclusion, the Census Bureau's simulated reconstruction-abetted re-identification attack definitively established that the tabular summaries from the 2010 Census could

be used to reconstruct individual record-level data containing the tabulation variables with their most granular definitions. Such microdata violated the disclosure avoidance rules that the Data Stewardship Executive Policy Committee had established for the 2010 Census and would not have been released had they been proposed as an official product because they posed too great a disclosure risk. The disclosure risk presumed by the 2010 standards recognized the excessive risk of re-identification if block geographic identifiers were placed on a 100% enumeration microdata file along with age (in years) and sex. The Census Bureau believed in 2010 that the minimum population that the geographic identifier could represent in such microdata is 100,000 persons—the minimum size of a Public-Use Microdata Area. That belief was strongly confirmed by the simulated re-identification attack. Somewhere between 52 and 179 million person who responded to the 2010 Census can be correctly re-identified from the re-constructed microdata, depending upon the quality of the external name, address, birth date, and sex information.

Appendix C

FINAL 2/16/10

DSEP Meeting Record

Topic: Updates

Meeting Date: 1/14/10

Attendees:

Position	*Attending for Position*
Deputy Director (Chair)	Tom Mesenbourg
AD, Administration	Ted Johnson
AD, Decennial	David Whitford
AD, Demographic	Cheryl Landman
AD, Economic	Harvey Monk
AD, Field	Marilia Matos
AD, IT	Brian McGrath
AD, Strategic Planning	Nancy Gordon
Rep. for Statistical Methodology	Tommy Wright
Senior Advisor for Data Management	Teresa Angueira
Chief, ITSO	
Chief, OAES	Kathleen Styles
Chief Privacy Officer	Mary Frazier
Also Attending:	Carol Comisarow, Ron Jarmin, David Raglin, Sharon Stern, Laura Zayatz, Michael Hawes

UPDATES

Background

[redacted]

Disclosure Review Board

- The DRB has been using enhanced disclosure avoidance procedures and methods for projects involving sensitive topics and/or sensitive populations. These procedures were implemented in response to an August 2004 memo from Director Kincannon. Though the memo was superseded by the Custom Tabulations policy, the DRB was not informed of this, and has not changed it procedures for sensitive topics and populations.
- Laura Zayatz also voiced the DRB's concern about planning for the 2020 Census and continuing to release data at the block level, as block populations continue to decrease (e.g. 40% of blocks in North Dakota have only 1 household in them)

[redacted]

Action Items

1. The DRB will develop recommendations for addressing the issue of disclosure review for sensitive populations. They will present their recommendations to DSEP once they have been vetted at the staff level.
2. DSEP agrees that the problem of block population size and disclosure avoidance is real, and that it deserves attention in the context of 2020 planning.

[redacted]

Appendix D

DSEP Meeting Record

Topics: [redacted]

Public Use File Reidentification Threats Update

Meeting Date: February 5, 2015

Attendees:

Position	*Attending for Position*
Deputy Director (Chair)	*absent*
AD, Administration	*absent*
AD, Decennial	Lisa Blumerman
AD, Demographic	Enrique Lamas
AD, Economic	*absent*
AD, Field	Jay Keller
AD, IT	Avi Bender
AD, Research and Methodology	Tom Louis
AD, 2020 Census	Lisa Blumerman
AD, Communications	Kim Collier
AD, Performance Improvement	Susan Reeves
Chief, PCO/ Chief Privacy Officer	Robin Bachman
Chief Demographer	Howard Hogan
Chief Information Security Officer	Tim Ruland
Asst. Director, Research and Methodology	Ron Jarmin
Also Attending:	Barbara Downs, Randy Becker, Byron Crenshaw, Laura McKenna, Heather Madray, Raj Dwivedy, Julie Atwell, Mike Castro

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Public Use File Reidentification Threats Update

Background and Discussion:
On December 11, 2014, DSEP discussed a reidentification issue that occurred involving a Public Use File (PUF) produced as part of the New York City Housing Vacancy Survey (NYCHVS). At that meeting, DSEP commissioned a team to pursue the recommendations presented to DSEP in July 2014 in the paper titled "PUMS File Re-identification Threats and Potential Solutions for Mitigating those Threats."

After discussing the logistics with some key stakeholders, and the difficulties of managing so many different angles on one team, DSEP approved a two-pronged approach to pursuing the paper's recommendations.

The Center for Disclosure Avoidance Research (CDAR) has recently received authorization to hire new staff to focus primarily on synthetic data and reidentification research. This group is preparing a research proposal that focuses on the disclosure avoidance side of the PUF reidentification issue.

In addition to these efforts, the Demographic Programs Directorate (ADDP) will charter a team that focuses on the broader future of PUFs as well as some of the non-technical means of disclosure avoidance discussed in the July 2014 paper. This team will discuss Terms of Use for PUFs, restricted access, and other methods. This team will have representation from all of the impacted directorates. DSEP also recommended the team engage with external researchers on some of these ideas, and address their concerns.

Action Items:
- CDAR will prepare a research proposal to outline future Census Bureau efforts in Synthetic Data and Reidentification Research.
- ADDP will charter a team to evaluate the future of PUFs and explore some of the non-technical solutions outlined in the July 2014 paper.

Appendix E

DSEP Meeting Record

Topics: Initial Request for DSEP Determination on Disclosure Avoidance for the 2018 End-to-End Test of the 2020 Census of Population and Housing (John Abowd, ADRM)

Record-level Re-identification Linkages for Evaluating the 2010 and 2020 Census Disclosure Avoidance Systems (John Abowd, ADRM)

[redacted]

Meeting Date: May 10, 2017

Position	*Attending for Position*
Deputy Director (Chair)	Laura Furgione
CAO	David Ziaya
CFO	Joanne Crane
AD, Decennial	Lisa Blumerman
AD, Demographic	Karen Battle
AD, Economic	Ron Jarmin
AD, Field	Joan Hill
AD, IT	Nitin Naik
AD, Research and Methodology	John Abowd
AD, 2020 Census	Lisa Blumerman
AD, Communications	Stephen Buckner
AD, Performance Improvement	Ted Johnson
Chief, PCO/ Chief Privacy Officer	Robin Bachman
Chief Demographer	Howard Hogan
Senior Advisor Designee from the Director's Office	*absent*
Chief Information Security Officer	*absent*

Asst. Director, Research and Methodology	John Eltinge
Also Attending:	Simson Garfinkel, Byron Crenshaw, Eloise Parker, Ashley Landreth, Mike Castro, Harold Saintelien, Janean Darden, Julie Atwell

Initial Request for DSEP Determination on Disclosure Avoidance for the 2018 End-to-End Test of the 2020 Census of Population and Housing

Background:

The Census Bureau's Research and Methodology Directorate (ADRM) is researching and developing disclosure avoidance methods and systems to replace those used for Census 2000 and the 2010 that were not designed to protect against database reconstruction attacks. ADRM is establishing the 2020 Disclosure Avoidance System (DAS), a formally private system based on the theoretical model known as differential privacy. This is the available technology for controlling reconstruction attacks.

The 2020 DAS team is working to establish adjustable formal privacy parameters for the 2018 End-to-End test. They are seeking DSEP concurrence with the Disclosure Review Board's (DRB's) April 10, 2017 determination that six data elements of PL 94-171 can continue to be published as enumerated. The team will test methods and systems with these elements published as enumerated for the 2018 End-to-End with the goal of making sound recommendations to DSEP for the full 2020 DAS. These elements to be published as enumerated are:

- the number of occupied housing units per block,
- the number of vacant housing units per block,
- the number of households per block,
- the number of adults (age 18+) per block (where the definition of an adult is inferred from the structure of the PL94-171 age categories),
- the number of children (age less than 18) per block (where the definition of a child is also inferred from the structure of the PL94-171 age categories),
- and the number of persons per block.

ADRM expects to perform follow-up analyses of the test products developed for the End-to-End Test. Because there is no national sample in 2018, some aspects of the differentially private system cannot be implemented in the End-to-End Test. They will have to be simulated from the 2010 Census data. This means that the demonstration data from the test can be made as noisy as DSEP wishes. However, there is only time to implement algorithms that maintain confidentiality with the six data elements in the 2010 PL94-171 redistricting data. There will be both policy and disclosure avoidance issues surrounding how broadly those products can be disseminated. Those issues will be brought to the DRB in a timely fashion.

ADRM also notes that DSEP will be asked to assume a formal policy consultant role for setting the confidentiality protection parameters for the final 2018 End-to-End Test and the 2020 DAS. The charter for DSEP currently delegates the authority to set disclosure avoidance standards to the DRB, with review by DSEP if necessary. However, these parameters now must be public in a formal privacy system. Furthermore, they, like any other operational decision need to be discussed and set in a manner consistent with their importance in the publication of results from the 2020 Census. The privacy-loss setting recommended by DRB and DSEP, and accepted by the Director, will be implemented in the production system.

Requests to DSEP:

Request 1: Concurrence with the DRB's decision on the PL 94-171 file items that can be published as enumerated.

In order to meet the timeline for the 2018 End-to-End Test, the version of the DAS under development for the test is limited in scope to the PL94-171 redistricting data. ADRM will not have time to experiment with a suite of potential implementations. And, in particular, ADRM will not have time to modify certain implementation decisions. They will be put back on the table for the full 2020 Disclosure Avoidance System and the decision on these six specific items may be revisited.

Request 2: Concurrence with Change to DRB Operating Principles Related to 2020 Census

The second request is for DSEP concurrence on a change in the operating principles of the DRB for issues related to disclosure avoidance in the 2020 Census of Population. Because the differentially private disclosure avoidance methods operate on the ensemble of proposed publications, DSEP is asked to concur that any disclosure avoidance request for publications from 2020 Census data be routed to the 2020 DAS team first. Those requests should not be considered by the DRB until the 2020 DAS team supplies a memo stating that the requested publication can or cannot be incorporated into the total privacy-loss accounting.

This is not a request for a moratorium on approvals for decennial data releases or design. The privacy-loss budget itself and its allocation to various components of the publication system are policy decisions that the 2020 Disclosure Avoidance System team will not make. Those decisions will ultimately be made in a manner consistent with the charters of the DRB and DSEP, and defended by the Director.

There is very little historical guidance for this process. We need to develop practical use cases that illustrate the consequences of publication decisions under alternative privacy-loss scenarios. We need to document the extent to which a best-effort reconstruction of the 2010 Hundred-percent Detail File (HDF) is correlated with the actual HDF. This is going to take some time. In the interim, ADRM is asking the DRB to take a leadership role in making these important choices by enabling the development of technologies better adapted to global risk management.

Discussion:

DSEP recognized the value in ADRM's efforts to assemble a skilled team of experts in an effort to modernize Census Bureau disclosure avoidance techniques using formal privacy methods.

This is essential in light of research that demonstrates that we must protect against database reconstructions that could lead to re-identification.

DSEP discussed the details of the six data elements from PL 94-171 and considered the necessity of including all of these in the proposed 2020 DAS research. ADRM requested that all elements remain available for the 2018 test research with a reconsideration for the full 2020 DAS, once the Census Bureau understand the outcomes. Conversations with the Department of Justice for Voting Rights Acts requirements with PL 94-171 will also play a part in future decisions about published enumerations.

DSEP recognized the need to develop ways to communicate with state stakeholders and the public about data protections that based on 2020 DAS methods. Our messaging will have to provide some simpler description of how the methods make changes to the attributes of the people in block counts, but still provide accurate and usable data.

DSEP noted that The National Conference of State Legislators (NCSL) will be expecting updates from Decennial based on 2018 testing outcomes in anticipation of 2020 releases of PL 94-171. It will be important to engage NCSL in discussions about 2020 DAS methods.

DSEP acknowledged that this and other details from ADRM's research were scheduled for discussion at the May 10, 2017 meeting of the 2020 Census Portfolio Management Governing Board (PMGB). DSEP postponed further discussion on this project and requests, pending any feedback from the presentation on this topic to the 2020 PMGB.

Post Meeting Notes:

DSEP revisited this topic at the beginning of the May 11, 2017 meeting.

Regarding issues of surrounding Voting Rights Acts Requirements, DSEP recognized that Decennial would need to talk to Justice if we were to alter any of the 6 constraints from PL 94-171 for 2020.

DSEP noted that the 2020 PMGB is supportive of the efforts of the 2020 DAS to optimize output noise infusion methods while publishing the most accurate data possible. There was unanimous support from 2020 PMGB for DRB's determination that the six data elements from PL 94-171 should be published as enumerated and form the base for the 2018 End-to-End testing research with the 2020 DAS.

DSEP agreed that the DRB should require that any request for disclosure avoidance of proposed publications for the 2020 Census be routed to the 2020 DAS team before going to the DRB.

Decision:

Request 1: DSEP approves publication of the six data elements from PL 94-171 as enumerated for the 2018 End-to-End test. Based on lessons learned, the use of these constraints for the PL 94-171 will be revisited for 2020.

Request 2: DSEP agreed that the DRB should require that any request for disclosure avoidance of proposed publications for the 2020 Census be routed to the 2020 DAS team before going to the DRB.

Record-level Re-identification Linkages for Evaluating the 2010 and 2020 Census Disclosure Avoidance Systems

Background:

The DAS team is attempting a database reconstruction using data from the 2010 PL94-171 and SF1 tabulations. The next step is to link those reconstructed microdata to commercial name and address files obtained in support of post-2010 research meant to represent the type of publically available file an attacker might potentially acquire. These files include Experian, InfoGroup, Melissa, Targus, TransUnion, and VSGI. This linkage involves the use of name and address data.

The final step is to compare the fully reconstructed microdata, including the commercially supplied names and address, to the name and address data on the 2010 Census Unedited File (CUF). Following accepted disclosure avoidance evaluation practices on re-identification, the 2020 DAS team would report to DRB and DSEP the putative re-identification rate (percentage of the records in the reconstructed microdata that could be linked to name and address information in the commercial files) and the proportion of putative re-identifications that were correct (proportion of reconstructed data records with putative re-identifications that were correctly linked to 2010 Census responses, including name and address).

Discussion:

DSEP recognized that the project proposal meets Data Linkage Policy requirements and involves sensitive but critical work that will allow the 2020 DAS subteam to understand the degree of risk of re-identification and database reconstruction with Census files.

DSEP noted that the subteam assembled for this research is composed of federal employees and one SSS individual.

Decision:

DSEP approved this project.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Appendix F

DSEP Meeting Record

Topics: 2020 Decennial Record Linkage Test (Ned Porter, CSRM)

[redacted]

[redacted]

Meeting Date: May 11, 2017

Position	*Attending for Position*
Deputy Director (Chair)	Ron Jarmin
CAO	David Ziaya
CFO	Joanne Crane
AD, Decennial	Al Fontenot
AD, Demographic	Karen Battle
AD, Economic	Ron Jarmin
AD, Field	Joan Hill
AD, IT	Nitin Naik
AD, Research and Methodology	John Abowd
AD, 2020 Census	Al Fontenot
AD, Communications	Stephen Buckner
AD, Performance Improvement	Ted Johnson
Chief, PCO/ Chief Privacy Officer	Robin Bachman
Chief Demographer	Howard Hogan
Senior Advisor Designee from the Director's Office	*absent*
Chief Information Security Officer	Tim Ruland
Asst. Director, Research and Methodology	John Eltinge

Also Attending:	Simson Garfinkle, Tommy Wright, Eloise Parker, Ned Porter, Bill Winkler, Christa Jones, Letitia McKoy, Melissa Creech, Hampton Wilson, Ashley Landreth, Mike Castro, Janean Darden, Julie Atwell

Administrative Notes:

At the beginning of the meeting, DSEP resumed their discussion and made a final decision on the topic: *Initial Request for DSEP Determination on Disclosure Avoidance for the 2018 End-to-End Test of the 2020 Census of Population and Housing*. The summary of that discussion and decision is in the May 10, 2017 meeting record.

2020 Decennial Record Linkage Test

Background:

Identifying duplicate records in the decennial census is critical to providing a more accurate count. One of the areas of research for improving the Decennial Matching methodology is improving the computer matching in the Duplicate Person Identification (DPI) process. This research will use the 2010 Census Unedited File (CUF) as well as data from Census Coverage Measurement (CCM). In addition, the research will determine if it is possible to increase the proportion of records receiving Personal Identification Keys (PIKs).

This research requires access to PIKs and complete name data on the files. This access will be limited to only five Census Bureau researchers as well as the Center for Statistical Research and Methodology's Data Steward. The data will be restricted to only authorized clusters.

Discussion:

DSEP acknowledged that research into deduplication methods is a routine and critical part of Census operations. DSEP further acknowledged that while this research project will use new technology and methods, it is fundamentally the same as research that happened in previous censuses.

Decision:

DSEP approved the project.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Appendix G

DSEP Meeting Record

Topics: [redacted]

Database Reconstruction Issue Mitigation (John Abowd, ADRM)

Meeting Date: February 15, 2018

Position	*Attending for Position*
COO (Chair)	Enrique Lamas
ADDC	Albert Fontenot
ADDP	Karen Battle
ADEP	Nick Orsini and Ron Jarmin
ADFO	Tim Olson
ADITCIO	Nitin Naik
ADRM	John Abowd
ADCOM	Stephen Buckner
CAO	David Ziaya
CFO	Joanne Crane
Asst. DRM	John Eltinge
Chief PCO/ Chief Privacy Officer	Robin Bachman
S.A. Director's Office	Douglas Clift
CISO	*Absent*
At-Large	Howard Hogan
At-Large	Frank Vitrano
Also Attending:	William Samples, David Waddington, Burton Reist, Victoria Velkoff, Robert Sienkiewicz, Jim Treat, Cynthia Hollingsworth, Clifford Jordan, Julia Naum, Jim Dinwiddie, Simson Garfinkel, Melissa Creech, Pat Cantwell, Byron Crenshaw, Hampton Wilson, Ashley Landreth, Mike Castro, Julie Atwell, Michael Snow

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

Database Reconstruction Issue Mitigation

<u>Background</u>
The Census Bureau's Operating Committee (OPCOM), serving as the Enterprise Risk Review Board, elevated the enterprise risk of database reconstruction to an enterprise issue based on the results of a database reconstruction attack research effort the Census Bureau launched to understand that risk better. When an enterprise risk is elevated to an enterprise issue, the risk owner must implement an active mitigation plan to mitigate the risk. To that end, the Research and Methodology Directorate presented six recommendations to help manage the Census Bureau's publication strategy in ways that will protect its databases from reconstruction attacks.

NOTE: presenters and DSEP recognized that implementing several of the recommendations will require decisions on budget and staffing resources and that those decisions would need to be handled by other bodies at the Census Bureau. DSEP confined its discussion to establishing policy in response to the recommendations.

The following 6 recommendations were presented to DSEP:

1. **Suspension until September 30, 2019 of ad hoc releases of sub-state geography from any confidential source unless vetted differential privacy tools, or a DRB-approved noise-infusion alternative, have been used to produce the publication. This applies to all research projects whether they are external or internal. It does not apply to scheduled publications from sponsored survey clients for whom there is already an approved DRB protocol. Those clients should be put on notice for subsequent contracts. The complete list of approved exceptions, including sponsored survey products, is provided in 20180215b-External_Internal_Substate_Geography.xlsx. The suspension will be reviewed prior to September 30, 2019.**

   NOTE: This suspension does not apply to state and national publications. It also does not apply to already scheduled publications from regular production activities. Program areas provided ADRM a list of those scheduled publications that should be exempted from the suspension. ADRM proposed ending those exemptions by September 30, 2019 even for those publications if they were not being produced using formally private systems by that point.

   Discussion: DSEP recognized the need to modernize the Census Bureau's disclosure avoidance systems. DSEP acknowledged that by approving a list of exemptions they are agreeing to hold elevated levels of risk of database reconstruction associated with all of these data products. However, DSEP acknowledged the Census Bureau is obligated to provide the data the public needs for decision making and some of the release dates are required by law.

   DSEP also acknowledged the need to set a target date for making these changes. While the ultimate goal is to make the publications of all of our programs formally private, that likely will not happen by September, 2019. However, in the meantime significantly improved noise infusion methods will be put in place to mitigate reconstruction risk.

   DSEP members expressed concern that the list of already scheduled publications presented might be incomplete and asked for additional time for program areas to review the list and submit updates. DSEP agreed that the Center for Disclosure Avoidance Research (CDAR) should continue to accept submissions and finalize the list in advance of the next DSEP meeting. DSEP will formally approve the list at that point.

   Decision: DSEP will finalize their approval of this recommendation at the March 15 DSEP meeting once the list of excepted publications has been finalized.

   Action Items: Program areas will send updates on the table of exempted data releases to the Chief of CDAR by February 23. The Chief of CDAR will redistribute the combined list to all contributors by February 28. CDAR will finalize the list of approved exceptions for distribution before DSEP's meeting on March 15.

2. **Suspension of all proposed tables in Summary File 1 and Summary File 2 for the 2020 Census at the block, block-group, tract, and county level except for the PL94-171 tables, as announced in Federal Register Notice 170824806–7806–01 (November 8, 2017, pp. 51805-6). To add a summary file table at any level of geography, racial/ethnic subpopulation other than OMB aggregate categories as specified in the 1997 standard (Federal Register October 30, 1997, pp. 58782-90), or group quarters type below the 2010 P42 seven categories, an affirmative case must be made for that table, use cases identified, and suitability for use standards developed. In addition, we recommend that the voting-age invariant in PL94-171 be removed, so that voting-age would be protected. DSEP will be asked to approve the SF1 and SF2 table specifications once they have cleared 2020 governance.**

   NOTE: The PL94-17 tables from the 2018 End-to-End Census Test have been designed with a formally private system already and will be published, with the voting-age invariant, as planned.

   Discussion: DSEP recognized that the SF1 and SF2 involved a very detailed set of tables that had been created to suit a wide set of data users. These tables were created, as a rule, to produce as much highly accurate data as possible within the existing disclosure avoidance framework. However, DSEP acknowledged that these data in many cases were accurate to a level that was not supported by the actual uses of those data, and such an approach is simply untenable in a formally private system.

   DSEP acknowledged a fundamental need to take stock of what data the Census Bureau is required to publish, both by statute and the needs of our data users, and at what level of accuracy. This is not an activity that should be done by our Disclosure Review Board. Program areas have to make the case of what the data will be used for, and the actual minimum level of accuracy needed for those uses, so that CDAR and the DRB can build the system to allocate the privacy-loss budget according to those use cases.

   A redesign of SF1 and SF2 based on formally articulated use cases will take a tremendous amount of effort but cannot be done in a vacuum. Program areas will have to reach out to data-user communities on developing the use cases for the needed data accuracy and levels of geography.

   NOTE: DSEP discussed but tabled until later any decision on changing the voting-age invariant for the PL94-171 table produced as part of the 2020 Census.

Decision: DSEP approved this recommendation. For the 2020 Census, SF1 and SF2 will be rebuilt based on use cases.

Action Items: DCMD, POP, and ADDC divisions will work with the relevant program management governing board (PMGB) to establish a plan to execute this redesign.

3. **Immediate review of all sub-state geography scheduled publications from the American Community Survey (ACS) to determine which ones can be delayed until there is a formally private publishing system for ACS.**

Discussion: DSEP acknowledged that many of the ACS tables are already in production and that production needs to move forward. DSEP acknowledged that there are likely no publications currently suitable for delay, however they emphasized that ACSO needs to ensure that all exceptions are added to the list.

Decision: DSEP approved this recommendation.

Action Items: ACSO will verify that they have included all of the necessary publications on the list of exempted data releases.

4. **Consideration of postponing ACS PUMS releases indefinitely.**

NOTE: DSEP recognized that all of the publication systems and methods for the Census of Island Areas are identical to the ACS. DSEP emphasized that any changes made to the ACS should also reflect consideration of the needs of the Island Areas.

Discussion: DSEP acknowledged that while the threat of database reconstruction and reidentification attacks applies to all of the Census Bureau's data products, should the ACS data be subject to a reidentification attack, from a public perception standpoint, our continued publication of the ACS PUMS files would appear to be an egregious mistake.

However, DSEP also acknowledged that the ACS PUMS is a heavily used dataset for research and recognized that discontinuing this publication could generate a great deal of traffic for the FSRDCs. DSEP acknowledged that, before the Census Bureau restricts use the ACS PUMS to the FSRDCs, it needs to verify that the they can handle the increased workload. Additionally, at present there are no FSRDCs that are readily accessible from the Island Areas.

DSEP recognized that immediate suspension of the ACS PUMS would cause a great deal of concern among data users and others. DSEP discussed the need to work on messaging around any suspension and to brief the Department of Commerce before the Census Bureau implements the suspension.

Decision: DSEP deferred for one month any decisions to suspend release of the ACS PUMS pending further consideration of the ability of the FSRDC network to support increased demand, the impact on the data needs of the Island Areas, and development of a messaging plan.

Action Items: ADRM will prepare an assessment of the potential increased demand on the FSRDC network, and Decennial will prepare an assessment of the impact of suspending this publication on the Island Areas. ADCOM will work on a messaging plan.

5. **Mandate for the 2022 Economic Censuses to use formally private publication systems for all tables.**

Discussion: DSEP recognized that it is too late to begin creating a formally private system for data releases from the 2017 Economic Census. DSEP additionally discussed how modernizing disclosure avoidance systems will involve much more than just budgeting extra funds. It also will require having the adequate number of people with the right skills to do the work.

DSEP recognized that program areas will have to involve their PMGB in setting resources, budgets, and timelines and that it should be feasible to put formally private systems in place in time for the 2022 Economic Census.

Decision: DSEP approved this recommendation. The Census Bureau will move forward with designing and implementing formally private systems for the 2022 Economic Census.

6. **Mandate to the Demographics Directorate to begin negotiations with survey clients for increased use of restricted-access microdata protocols and formally private table publication systems.**

POST MEETING NOTE: a member in attendance recommended that there should also be outreach to reimbursable clients for the Economic Directorate.

Discussion: DSEP recognized the need to begin discussions with sponsors of Census Bureau surveys but determined that the Census Bureau should have a communications plan in place before mandating that the Demographic Directorate speak to sponsors.

Decision: DSEP will reconsider in one month whether to mandate conversations with survey and report sponsors.

Consolidated Action items:

- Program areas will send updates on the table of exempted data releases to the Chief of CDAR by February 23.
- The Chief of CDAR will redistribute the combined list to all contributors by February 28.
- DCMD, POP, and the ADDC will work with the relevant PMGBs to establish a plan to execute the redesign of SF1 and SF2 based on use cases.
- ACSO will work to determine that all ACS data releases in production are listed on the spreadsheet of exceptions to the suspension.
- ADRM will prepare an assessment of the potential increased demand on the FSRDC network from suspension of the ACS PUMS.
- ADCOM will work on a messaging plan related to the suspension of the ACS PUMS.
- Decennial will prepare an assessment of the impact of suspending publication of the ACS PUMS on the Island Areas.

Appendix H

April 24, 2020

CHARTER OF THE 2020 DATA QUALITY EXECUTIVE GOVERNANCE GROUP

Background

The operational design of the 2020 Census has been transformed due to the ongoing COVID-19 pandemic. The start of the nonresponse followup (NRFU) operation has been delayed and,in addition, there are now delays for other field operations such as group quarters enumeration, service-based enumeration and transitory locations. Field work has been extended until October 31. These changes have a number of cascading effects on 2020 Census operations.

The Census Bureau remains committed to delivering a quality count of the nation's population. The agency will build upon existing quality assurance efforts to ensure changes to the 2020 Census operational design and their consequences are documented.

A series of work groups, comprised of staff from across the Census Bureau with the requisite expertise, will be established to develop proposals for enhanced use of administrative records, modifications to field operations, and quality assessments of changes to the operational plans.

2020 Data Quality Executive Governance Group

The 2020 Data Quality Executive Governance Group (EGG) was constituted by the Deputy Director to provide guidance and vet of statements about the quality of the 2020 Census data. The EGG draws upon expertise within the Census Bureau in the fields of census operations, statistical methodology, acquisition and utilization of administrative records, and in the social, economic, and housing subject areas.

The EGG membership is as follows:

- Leads:
  - John Abowd, Associate Director for Research and Methodology and Chief Scientist
  - Tori Velkoff, Associate Director for Demographic Programs
  - Deb Stempowski, Assistant Director for Decennial Programs, Operations and Schedule Management
- Members:
  - Ron Jarmin, Deputy Director and Chief Operating Officer
  - Enrique Lamas, Senior Advisor, Office of the Deputy Director
  - Christa Jones, Chief of Staff, Office of the Deputy Director
  - Al Fontenot, Associate Director for Decennial Programs
  - Pat Cantwell, Chief of the Decennial Statistical Studies Division

    - Jamey Christy, Assistant Director for Field Operations
    - Ben Page, Chief Financial Officer
- Coordinator: Jennifer Ortman, Population Division

Mission of the EGG

The mission of the EGG is to provide direction and approvals about:

- Quality assessments of changes to the operational plans.
- Quality assessments of the 2020 Census data during and post data collection.

It is expected that the EGG will:

- Be informed on modifications to the 2020 Census field operations including enhanced use of statistical procedures.
- Follow the 2020 Census governance structure.
- Set forth requirements for documentation of changes to the 2020 Census operational design as they impact the quality of the census data products.
- Provide guidance to the expert work groups.
- Engage internal and external stakeholders to inform relevant audiences and solicit feedback as work progresses.

Work Groups

The EGG will create work groups as necessary to study particular issues, organize workshops or seminars, or provide a forum for discussion of particular topics. These work groups are intended to provide support and expert resources for the operational coordinators. They do not serve as additional oversight for census operations and quality assessments.

The work groups report to the EGG. This interaction will be facilitated by the EGG Coordinator.

Communication

A secure, shared directory will be created and maintained to facilitate communication among the EGG and work groups. The EGG will hold regular meetings and use email to stay up-to-date on the activities of the 2020 Data Quality work groups.

Amendments and Revisions to this Charter

Should the activities of the EGG evolve over time, this charter will be updated/revised to reflect those changes in activities. However, the principles that underlie those activities will remain – namely that the work groups will work closely with the EGG on determining their priorities, and will report out their research and findings. Amendments and revisions to this charter must be approved by the EGG.

Amendment and Revision History

Date approved	Description of amendment or revision
April 24, 2020	Initial release.

Signatures

Name and Title	Signature	Date
Ron Jarmin Deputy Director and Chief Operating Officer	RON JARMIN Digitally signed by RON JARMIN Date: 2020.04.28 10:55:17 -04'00'	
John Abowd Associate Director for Research and Methodology and Chief Scientist	JOHN ABOWD Digitally signed by JOHN ABOWD Date: 2020.04.24 11:24:58 -04'00'	
Victoria Velkoff Associate Director for Demographic Programs	VICTORIA VELKOFF Digitally signed by VICTORIA VELKOFF Date: 2020.04.24 11:32:14 -04'00'	
Albert Fontenot Associate Director for Decennial Programs	ALBERT FONTENOT Digitally signed by ALBERT FONTENOT Date: 2020.04.28 10:48:29 -04'00'	
Deborah Stempowski Assistant Director for Decennial Programs, Operations and Schedule Management	Deborah Stempowski Digitally signed by Deborah Stempowski Date: 2020.04.24 11:54:05 -04'00'	
Enrique Lamas Senior Advisor Office of the Deputy Director	ENRIQUE LAMAS Digitally signed by ENRIQUE LAMAS Date: 2020.04.24 12:34:05 -04'00'	
Christa Jones Chief of Staff Office of the Deputy Director	CHRISTA JONES Digitally signed by CHRISTA JONES Date: 2020.05.01 09:52:59 -04'00'	
Patrick Cantwell Chief of the Decennial Statistical Studies Division	Patrick J. Cantwell Digitally signed by Patrick J. Cantwell Date: 2020.04.27 11:38:31 -04'00'	
James Christy Assistant Director for Field Operations	James Christy Digitally signed by James Christy Date: 2020.04.27 12:01:34 -04'00'	
Ben Page Chief Financial Officer	BENJAMIN PAGE Digitally signed by BENJAMIN PAGE Date: 2020.04.27 14:05:37 -04'00'	