



# County of Fairfax, Virginia

To protect and enrich the quality of life for the people, neighborhoods and diverse communities of Fairfax County

September 11, 2019

Census Scientific Advisory Committee

Re: Public Comments to Census Scientific Advisory Committee Fall Meeting

Dear Committee Members:

We, at the Economic, Demographic and Statistical Research (EDSR) of Fairfax County, Virginia are writing to you with serious concerns and considerable confusion about the disclosure avoidance measure, "differential privacy" (DP), to be implemented with the 2020 Census. We are calling for a transparency of the methodology that highlights the impact on redistricting process and any subsequent census data products/programs. **Census Bureau's mission is to serve as the nation's leading provider of quality data about its people and economy.** 2020 Census needs to provide accurate data that meet local governments' and public services' data needs. We urge a carefully mapped-out adoption procedure preceded by rigorous and prudent scientific evaluation from diverse local data users.

We have outlined our concerns and confusion below.

## Redistricting

Under the U.S. Constitution and the 14th Amendment, redistricting is clearly associated with population equality (i.e., Apportionment Clause of Article I, Section 2) and anti-discrimination (i.e., plans are prohibited to intentionally or inadvertently discriminate on the basis of race, which could dilute the minority vote; Voting Rights Act of 1965, Section 2). Getting accurate data from the 2020 Census will be the first step towards a fair redistricting process. Currently, 21 states explicitly require the use of census data for redistricting, and 17 states have an implied basis or in-practice reliance on using the census for redistricting<sup>1</sup>. This essentially requires 2020 Decennial Census data to be accurate enough at small area such as block and block group levels for redistricting purpose.

We urge Census Bureau demonstrate how the accuracy will be improved from the dress rehearsal (Abowd 2019) and clearly communicate it to local jurisdictions prior to the 2020 Census.

---

<sup>1</sup> <http://www.ncsl.org/research/redistricting/redistricting-criteria.aspx>

## Local Governments Rights and Public Services

We are extremely concerned to learn from news articles after Census announced plans to “report less than accurate data, to reduce privacy risks <sup>2</sup>.” The reduced accuracy will have the most significant ramification at the small area, such as census blocks, census block groups and census tracts; and we, the local government are the entity that will be affected the most. Not having accurate data impairs our core function as a local government to provide essential targeted services to our residents. Without properly evaluating the impacts of altered disclosure avoidance procedure, we are extremely worried that 2020 census data could mislead us in creating policy or service delivery at small geographies.

As a century-long participant in the census, and the keeper of local administrative and survey data, we completely understand the importance of protecting confidentiality. We appreciate Census detecting the vulnerability of public data exposure in reconstructing and facing the emerging challenges in this big data era. **However, the Census Bureau needs to find a level of injected noise that balances the utility of local area data with the requirement that individual responses remain confidential.**

Fairfax County, as well as thousands of other local governments, rely on small area statistics from the Census to guide policymaking and local planning in order to provide public services critical to residents’ safety and well-being. Most of the policies and programs at the local level are at neighborhood or smaller geographies. The local government is where the rubber meets the road. Together with other jurisdictions in Northern Virginia, Fairfax County champions every decennial census. Fairfax County allocates financial, material and human resources to actively support many Census technical programs such as Local Update of Census Addresses Operation (LUCA) and Participant Statistical Areas Program (PSAP) in addition to carrying out community outreach for the Complete Count Committee. These are all vital components to ensure the success of 2020 Census by ‘Counting Everyone Once, Only Once, in the Right Place’.

Therefore, we have an obligation to our taxpayers to use information to allocate resources appropriately. Census Bureau also has an obligation to provide accurate and useful data.

### **Necessity of Sufficient Scientific Evaluation to Ensure Data Utility**

Since the differential privacy concept appeared 13 years ago (Dwork 2006), it has been mainly utilized by tech firms. These business organizations usually have clearly defined analytical purposes with their data and operate on completely different business models from the field of Public Administration. With mission to “**serve as the nation’s leading provider of quality data about its**

---

<sup>2</sup> <https://www.nytimes.com/2018/12/05/upshot/to-reduce-privacy-risks-the-census-plans-to-report-less-accurate-data.html>

**people and economy”,** the Census Bureau is tasked with providing data to serve a wide range of data users. However, to date the DP’s potential impact on the data usability is still weakly tested.

Investigating the potential implication of differential privacy to public and social services with census data, 12 articles were found in scientific database with “Differential Privacy” and “Census” as key words to date (Web of Science 2019). Among them, five did not evaluate census data usability (Abowd 2018; Bilogrevic et al. 2014; Garfinkel et al. 2018; Salas et al. 2018; Torra and Navarro-Arribas, 2014); three suggests it could be a favorable approach if the methodology is developed and implemented in a way that allows high accuracy (Haney et al. 2017; Rinott et al. 2018; Schneider and Abowd 2015); three cautioned the implication of differential privacy from a technical perspective by revealing “serious questions regarding the efficacy of using differential privacy-based masking mechanism for numerical data” (Muralidhar and Sarathy 2010) and “adding more noise does not always increase the real privacy” (Sramka 2010, 2012); and one warned about the use of the differential privacy approach “could make the release of scientifically useful microdata impossible and severely limit the utility of tabular small-area data” (Ruggles et al. 2019).

Clearly, further investigation and methodology development are needed prior to 2020 Census to ensure adequate accuracy for data utility in small-area statistics. We urge Census Bureau to work closely with the data user communities in determining the noise levels. We earnestly ask the CSAC to help ensure that the Census Bureau provide local government data users with the data needed to make a proper evaluation prior to 2020 Census and to provide feedback on how to best allocate the privacy budget.

We therefore call for a carefully mapped-out DP adoption procedure that is preceded by rigorous and prudent scientific evaluation and with greater transparency in this process.

Thank you for giving us this opportunity to voice our concerns on this very important subject matter.

Sincerely,



Fatima Khaja, Manager  
Economic, Demographic and Statistical Research

<https://www.fairfaxcounty.gov/demographics/>

## References

- Abowd J.M. 2018. The US Census Bureau Adopts Differential Privacy. *Proceedings of the 24<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*: 2867. <https://doi.org/10.1145/3219819.3226070>
- Abowd J. M. 2019. Stepping-up: The Census Bureau Sets an Example of How to Be a Good Data Steward in the 21st Century. Presentation in *Data Privacy: From Foundations to Applications Seminar*, Simons Institute, University of California at Berkeley, [https://youtu.be/yUyCYC6rb\\_4](https://youtu.be/yUyCYC6rb_4)
- Bilogrevic I., Freudiger J., De Cristofaro E., Uzun E. 2014. What's the Gist? Privacy-Preserving Aggregation of User Profiles. In Kutylowski M.; Vaidya J. ed. *Computer Security- ESORICS 2014, PT II. Book Series: Lecture Notes in Computer Science*. 8713: 128-145.
- Dwork C. 2006. Differential Privacy. In Bugliesi M., Preneel B, Sassone V. Wegener I. eds. *Automata, Languages and Programming: 33<sup>rd</sup> International Colloquium*: 1-12
- Garfinkel S. L., Abowd J. M., Powazek S. 2018. Issues Encountered Deploying Differential Privacy. *Proceedings of the 2018 Workshop on Privacy in the Electronic Society (WPES'18)*: 133-137. <https://doi.org/10.1145/3267323.3268949>
- Haney S., Machanavajjhala A., Abowd J. M., Graham M., Kutzbach M., Vilhuber L. 2017. Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics. *Proceedings of the 2017 ACM International Conference on Management of Data (SIGMOD'17)*: 1339-1354. <https://doi.org/10.1145/3035918.3035940>
- Muralidhar K., Sarathy R. 2010. Does Differential Privacy Protect Terry Gross' Privacy? In DomingoFerrer J., Magkos E. ed. *Privacy in Statistical Databases*. Book Series: Lecture Notes in Computer Science. 6344: 200-209.
- Rinott Y., O'Keefe C.M., Shlomo N., Skinner C. 2018. Confidentiality and Differential Privacy in the Dissemination of Frequency Tables. *Statistical Science*. 33(3): 358-385
- Ruggles S., Fitch C., Magnuson D., Schroeder J. 2019. Differential Privacy and Census Data: Implications for Social and Economic Research. *AEA Papers and Proceedings*. 109: 403-408. <https://doi.org/10.1257/pandp.20191107>
- Salas J., Domingo-Ferrer, J. 2018. Some Basics on Privacy Techniques, Anonymization and their Big Data Challenges. *Mathematics in Computer Science*. 12(3): 263-274
- Schneider M. J., Abowd J. M. 2015. A new method for protecting interrelated time series with Bayesian prior distributions and synthetic data. *Journal of the Royal Statistical Society Series A-Statistics in Society*. 178(4): 963-975. <https://doi.org/10.1111/rssa.12100>

Sramka, M. 2012. Breaching Privacy Using Data Mining: Removing Noise from Perturbed Data. In Elizondo D.A., Solanas A. MartinezBalleste A. ed. *Computational Intelligence for Privacy and Security. Book Series: Studies in Computational Intelligence*. 394: 135-157

Sramka, M. 2010. A Privacy Attack That Removes the Majority of the Noise from Perturbed Data. 2010 International Joint Conference on Neural Networks (IJCNN). IEEE.

Torra V., Navarro-Arribas G. 2014. Data privacy. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*. 4 (4): 269-280. <https://doi.org/10.1002/widm.1129>

Web of Science. Database. Search done on Aug 23, 2019. [www.webofknowledge.com](http://www.webofknowledge.com)