12 August, 2021

Dr. Ron Jarmin
Acting Director
United States Census Bureau
4600 Silver Hill Road
Washington DC 20233

By Email: ron.s.jarmin@census.gov

Cc: Dr. John Abowd, Chief Scientist and Associate Director for Research and Methodology
Cc: Mr. James Whitehorne, Chief of the Census Redistricting and Voting Rights Data Office

Re: Request for release of "noisy measurements file" by September 30 along with redistricting data products

Dear Dr. Jarmin,

We write to ask you to publicly release, by September 30, the pre-post-processed data, i.e., what the Census Bureau calls the "noisy measurements dataset," on which the redistricting (PL 94-171) data are based.

Federal law requires that census results be released in ways that people's individual answers to the census takers are kept private. The Census Bureau itself has shown that the method it used in 2010 to protect privacy, known as swapping, is vulnerable to technical advances in data reconstruction science and is no longer protective.[1] In fact, swapping also secretly introduced biases which may even have led to the misallocation of federal funds, the drawing of unfair redistricting lines, or incorrect scholarly conclusions about the American public or public policy.[2] These biases have never been publicly quantified because of the nature of the disclosure avoidance system then in place and the secrecy of the techniques. This made it impossible for Census data users to learn about the biases, correct the data, or derive appropriate uncertainty estimates. Instead, most users merely made believe no biases existed.

It's therefore essential for the 2020 Census data to be released in a way that protects the public from biased conclusions about the American population, while also protecting every individual's private information: As privacy threats increase, the Bureau is required to respond with improved privacy protection methods. And indeed, the Bureau has made remarkable progress.

Examination of demonstration data produced by the Census Bureau's new TopDown algorithm reveals several new biases for certain quantities of interest, such as the apparent "transfer" of

[1] United States Census Bureau, *The Census Bureau's Simulated Reconstruction-Abetted Re-identification Attack on the 2010 Census* (May 7, 2021), https://www.census.gov/data/academy/webinars/2021/disclosure-avoidance-series/simulated-reconstruction-abetted-re-identification-attack-on-the-2010-census.html.
[2] Michael Hawes and Rolando A. Rodríguez, *Determining the Privacy-loss Budget* (May 25, 2021), https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf?

population from urban to rural areas.[3] This could cause problems for those engaged in redistricting, who seek to draw accurate conclusions about the characteristics of the people in newly drawn ward or district.

Can valid conclusions about fair redistricting, the allocation of federal funds, minority representation in different localities, and all the other uses of Census data be drawn from the data file the Census Bureau plans to release? Yes, because differential privacy, unlike swapping, permits us to make bias corrections and thus draw accurate conclusions. However, with only the dataset the Census Bureau is currently planning to release, it will take considerable statistical effort and expertise to measure and correct for all the biases. Can such conclusions be drawn faster and more easily from another data file the Census Bureau is creating with the identical privacy protections but is not currently planned for release? Absolutely.

In order to protect privacy, the TopDown algorithm first creates what it calls the "noisy measurements" datafile by adding carefully calibrated random noise to provide mathematical guarantees of privacy for all people in the country. The result is a datafile that lists some census blocks with some populations that are slightly too small (and some even negative) and others slightly too large. The TopDown algorithm then adjusts the noisy measurements file into a second, "post-processed" dataset. Although this post-processing is quite helpful for some purposes, it introduces biases that are difficult to correct for others.

Fortunately, making available the noisy measurements file is an easy solution and the key to ensuring that analysts can easily use the data appropriately. Without compromising privacy, release of that file will allow analysts to correct for all biases (with straightforward statistical methods easy to make available to all) and to offer accurate margins of error. Map drawers who want to gerrymander may, frustratingly, still be able to do so.[4] More importantly, all participants will be able to evaluate the racial and partisan impact of any proposed redistricting plan, and compute any other quantity required for governance or scholarly research — something swapping does not permit — without violating anyone's privacy.

The simple act of making public the "noisy measurements" datafile, which the Census Bureau already produces as a key step in its algorithm, might then be the rare situation where all parties can agree. Trust is important for ensuring participation in all future censuses, and confidentiality of responses is essential to maintaining trust. Accuracy of conclusions drawn from census data is just as essential.

There is no legal, statistical, or privacy obstacle to the Census Bureau releasing the noisy measurements. It is time to help users stop pretending there is no noise, and to empower them to address uncertainty in data – whether from disclosure avoidance or any other source – with mathematical rigor. Privacy is a first step on this path. We just need the Census Bureau to release these measurements.

Please feel free to contact Cynthia Dwork, Gary King, or Ruth Greenwood to discuss these issues in more detail.

Sincerely

---

[3] Mexican American Legal Defense and Education Fund and Asian Americans Advancing Justice, *Preliminary Report: Impact of Differential Privacy & the 2020 Census on Latinos, Asian Americans and Redistricting* (Apr. 2021), https://www.maldef.org/wp-content/uploads/2021/04/FINAL-MALDEF-AAJC-Differential-Privacy-Preliminary-Report-4.5.2021-1.pdf.

[4] *See Rucho v. Common Cause*, 139 S. Ct. 2484 (2019).

**** All affiliations are listed for identification purposes only****

Cynthia Dwork
Gordon McKay Professor of Computer Science
Harvard University
dwork@seas.harvard.edu

Gary King
Albert J. Weatherhead University Professor, and Director of the Institute for Quantitative Social Science
Harvard University
king@harvard.edu

Ruth Greenwood
Director, Election Law Clinic
Harvard Law School
rgreenwood@law.harvard.edu

William T. Adler
Senior Technologist, Elections & Democracy
Center for Democracy & Technology
wadler@cdt.org

Joel Alvarez
Deputy Director of Population Division
New York City Department of City Planning
JAlvare@planning.nyc.gov

María Ballesteros
Ph.D. Student
Harvard University
mballesteros@g.harvard.edu

Nathaniel Beck
Prof.,
Dept, of PoliticsNYU
nathaniel.beck@nyu.edu

Dan Bouk
Associate Professor of History
Colgate University
dbouk@colgate.edu

danah boyd
Partner Researcher
Microsoft Research
dmb@microsoft.com

**** All affiliations are listed for identification purposes only****

John Brehm
Professor of Political Science
University of Chicago
jjbrehm@uchicago.edu

Mark Bun
Assistant Professor of Computer Science
Boston University
mbun@bu.edu

Aloni Cohen
Postdoctoral Associate
Boston University
aloni@bu.edu

Clifford Cook
Senior Planning Information Manager
City of Cambridge, MA
ccook@cambridgema.gov

Damien Desfontaines
Privacy Engineer
Google
damien@desfontain.es

Georgina Evans
PhD Candidate
Harvard University
georginaevans@g.harvard.edu

Abraham D. Flaxman
Associate Professor
University of Washington
abie@uw.edu

Robert J. Franzese, Jr.
Professor of Political Science
University of Michigan
franzese@umich.edu

Marco Gaboardi
Associate Professor
Boston University
gaboardi@bu.edu

Roxana Geambasu
Associate Professor of Computer Science
Columbia University
roxana@cs.columbia.edu

**** All affiliations are listed for identification purposes only****

Jeff Gill
Distinguished Professor, Departments of Government, Mathematics & Statistics
American University
jgill@american.edu

Ruobin Gong
Assistant Professor of Statistics
Rutgers University
ruobin.gong@rutgers.edu

Daniel L. Goroff
Vice President and Program Director
Alfred P. Sloan Foundation
goroff@sloan.org

Erica L. Groshen
Senior Economic Advisor and Former Commissioner of US Bureau of Labor Statistics
Cornell University--ILR School
Erica.Groshen@Cornell.edu

Evelynn M. Hammonds
Professor
Harvard University
evelynn_hammonds@harvard.edu

Kosuke Imai
Professor of Government and of Statistics
Harvard University
imai@harvard.edu

Jonathan N. Katz
Kay Sugahara Professor of Social Sciences and Statistics
Caltech
jkatz@caltech.edu

Nitin Kohli
PhD Candidate
UC Berkeley School of Information
nitin.kohli@berkeley.edu

Aleksandra Korolova
WiSE Gabilan Assistant Professor of Computer Science
University of Southern California
korolova@usc.edu

**** All affiliations are listed for identification purposes only****

Frauke Kreuter
Professor
University of Maryland
fkreuter@umd.edu

Tracy Kugler
Research Scientist
IPUMS, University of Minnesota
takugler@umn.edu

Shiro Kuriwaki
Postdoctoral Fellow
Stanford University
shirokuriwaki@stanford.edu

David Lazer
University Distinguished Professor
Northeastern University
d.lazer@northeastern.edu

Jing Lei
Associate Professor of Statistics and Data Science
Carnegie Mellon University
jinglei@andrew.cmu.edu

Margaret Levenstein
Director
ICPSR
MaggieL@umich.edu

Jeffrey B. Lewis
Professor of Political Science
University of California Los Angeles
jblewis@ucla.edu

Katrina Ligett
Associate Professor of Computer Science
The Hebrew University of Jerusalem
katrina@cs.huji.ac.il

John Londregan
Professor of Politics and International Affairs
Princeton University
jbl@princeton.edu

**** All affiliations are listed for identification purposes only****

Cory McCartan
Ph.D. Candidate
Department of Statistics, Harvard University
cmccartan@g.harvard.edu

Michael McDonald
Professor
University of Florida
michael.mcdonald@ufl.edu

Walter Mebane
Professor Political Science and Statistics
University of Michigan
wmebane@umich.edu

Martha Minow
Professor
Harvard Law School
minow@law.harvard.edu

Kobbi Nissim
Professor
Department of Computer Science, Georgetown University
kobbi.nissim@georgetown.edu

Jules Polonetsky
CEO
Future of Privacy Forum
julespol@fpf.org

Kevin M. Quinn
Professor of Political Science
University of Michigan
kmq@umich.edu

Evan T. R. Rosenman
Postdoctoral Fellow
Harvard Data Science Initiative
erosenm@fas.harvard.edu

Aaron Roth
Professor
University of Pennsylvania
aaroth@cis.upenn.edu

Guy Rothblum
Associate Professor
Weizmann Institute of Science
rothblum@alum.mit.edu

**** All affiliations are listed for identification purposes only****

Joseph J Salvo
Senior Advisor/Former NYC Chief Demographer
National Conference on Citizenship
joe@ncoc.org

Aleksandra Slavkovic
Professor of Statistics
Penn State University
abs12@psu.edu

Adam Smith
Professor of Computer Science
Boston University
ads22@bu.edu

Nicholas Stephanopoulos
Kirkland & Ellis Professor of Law
Harvard Law School
nstephanopoulos@law.harvard.edu

Weijie Su
Assistant Professor of Statistics and Data Science
University of Pennsylvania
suw@wharton.upenn.edu

Kunal Talwar
Research Scientist
Apple
kunal@kunaltalwar.org

Abhradeep Guha Thakurta
guhathakurta.abhradeep@gmail.com

Rocio Titiunik
Professor of Politics
Princeton University
titiunik@princeton.edu

Salil Vadhan
Vicky Joseph Professor of Computer Science & Applied Mathematics
Harvard University
salil_vadhan@harvard.edu

David Van Riper
Director of Spatial Analysis
IPUMS
vanriper@umn.edu

**** All affiliations are listed for identification purposes only****

Larry Wasserman
UPMC Professor of Statistics
Carnegie Mellon University
larry@cmu.edu

Linjun Zhang
Assistant Professor
Rutgers University
linjun.zhang@rutgers.edu