

● **BREAKING** BIDEN NOMINATES RACHAEL ROLLINS AS US ATTORNEY FOR MASSACHUSETTS

BREAKING: BIDEN NOMINATES RACHAEL ROLLINS AS US ATTORNEY FOR MASSACHUSETTS

[More from Globe Opinion](#) [Submit a Letter to the Editor](#) [Meet the Editorial Board](#) [Globe Op-Talks](#)

OPINION

# There's a simple solution to the latest census fight

Trust is important for ensuring participation in all future censuses, and confidentiality of responses is essential to maintaining trust.

By **Cynthia Dwork, Ruth Greenwood, and Gary King** Updated July 26, 2021, 3:00 a.m.



IGOR VITOMIROV/ADOBE/BABAROGA - STOCK.ADOBE.COM

The political battle over the decennial US Census has begun again.

From George Washington's first presidential veto in 1792 until today, the census is a

FROM GEORGE WASHINGTON'S FIRST PRESIDENTIAL VETO IN 1792 UNTIL TODAY, THE CENSUS IS A unique combination of the science of enumeration and some of the nation's most intense internal politics. This mess is no accident: Census counts determine the allocation of [billions of dollars in federal funds](#) and the communities that will benefit from the new redistricting maps to be drawn with the new data, from congressional districts to mosquito abatement districts.

The political goals of the fight are always recognizable, while the playing field, and reasoning of the combatants, is regularly invented anew. The next dispute will start after [Aug. 16, when the redistricting numbers will be released](#) by the US Census Bureau. Alabama has already objected to the modernization of the Census Bureau's techniques used to protect confidential personally identifiable information from being inferred from census data. The data in question are to be released with the Census Bureau's new "[TopDown](#)" disclosure avoidance algorithm that relies, in part, on a technique, widely accepted in industry and academia, known as "differential privacy." It's a technique used to share information about communities within a dataset without revealing any information about individual people. Alabama claimed that the data will not be sufficiently accurate for redistricting. A three-judge panel out of the US Court of Appeals for the 11th Circuit [sided with the bureau in June](#), but expect challenges.

The dispute between Alabama and the Department of Justice [has been portrayed in the press](#) as a battle between privacy protection and data accuracy — with Alabama painted to be proposing the public release of specific demographic information that allows corporations or even malicious actors to microtarget and harass individuals, and the DOJ caricatured as so obsessed with privacy that it doesn't care whether the census data can be used for any purpose at all.

Unlike in so many cases the courts must consider, there is a way to give both sides much of what they want. It's just not what anyone has asked the court for. Yet.



Federal law requires that census results be released in ways that people's individual answers to the census takers are kept private. It is public information that the [Census Bureau from 1990 until 2010 used a method of privacy protection](#) called swapping, the details of which remain secret. Swapping [introduced biases](#) that may 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. However, these biases have never been publicly quantified, and so it was impossible for census data users to learn or correct the data. Instead, users were made to believe no biases existed.

It's therefore essential for 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, they have made much progress.

The biggest conceptual hurdle for those analyzing the new privacy protection methods has been that they forget the simple fact that the purpose of census data products is not to tell us the [age, sex, and race of the two people who live on Liberty Island](#) but to ensure that if we place those people in a ward or a district, everyone can draw accurate conclusions about the characteristics of that ward or district.

*Can valid conclusions about fair redistricting, the allocation of federal funds, minority*

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 unusual dataset the Census Bureau is planning to release, it will take 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 has not yet released? Absolutely.

The TopDown algorithm first creates what it calls the “[noisy measurements](#)” datafile — essentially, statistical tricks used to prevent individuals from being identified — which protects population counts with just enough 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 (small geographic units that, in reality, can have between zero and a few hundred people) with some populations that are slightly too small and others slightly too large.

Because this dataset isn't immediately intuitive, the TopDown algorithm then adjusts the noisy measurements file into a second dataset. Although this post-processing is quite helpful for some purposes, it introduces some 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. 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 (see the Supreme Court case [Rucho v. Common Cause](#)). More importantly, all participants will be able to evaluate the racial and partisan impact of any proposed redistricting plan, and compute any other quantity needed for the allocation of federal funds — something swapping does not permit — without violating anyone's privacy. In short, there is no legal, statistical, or privacy obstacle to the Census Bureau releasing the noisy measurements in addition to going

ahead with their plans to release the post-processed file.

The simple step of making public this “noisy-measurements” datafile, which the Census Bureau already produces as a key step in its algorithm, might then be the rare situation where Alabama and the Biden administration’s Department of Commerce 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.

We just need the Census Bureau to give the country the right data files both to ensure accuracy and protect privacy.

*Cynthia Dwork is a professor of computer science at Harvard University and an inventor of differential privacy. Ruth Greenwood is director of the Election Law Clinic at Harvard Law School and has litigated multiple redistricting cases at the Supreme Court. Gary King is university professor, director of the Institute for Quantitative Social Science at Harvard University, and an inventor of the methods for detecting racial and partisan gerrymandering.*

©2021 Boston Globe Media Partners, LLC