Privacy and the 2020 Decennial Census

Michael Hawes
Senior Advisor for Data Access and Privacy
Research and Methodology Directorate
U.S. Census Bureau

Insular and International Affairs Webinar
Department of the Interior
February 6, 2020

Shape your future START HERE >

United States Census 2020

Acknowledgements

This presentation includes work by the Census Bureau's 2020 Disclosure Avoidance System development team, Census Bureau colleagues, and our collaborators, including: John Abowd, Tammy Adams, Robert Ashmead, Craig Corl, Ryan Cummings, Jason Devine, John Fattaleh, Simson Garfinkel, Nathan Goldschlag, Michael Hawes, Michael Hay, Cynthia Hollingsworth, Michael Ikeda, Kyle Irimata, Dan Kifer, Philip Leclerc, Ashwin Machanavajjhala, Christian Martindale, Gerome Miklau, Claudia Molinar, Brett Moran, Ned Porter, Sarah Powazek, Vikram Rao, Chris Rivers, Anne Ross, Ian Schmutte, William Sexton, Rob Sienkiewicz, Matthew Spence, Tori Velkoff, Lars Vilhuber, Bei Wang, Tommy Wright, Bill Yates, and Pavel Zhurlev.

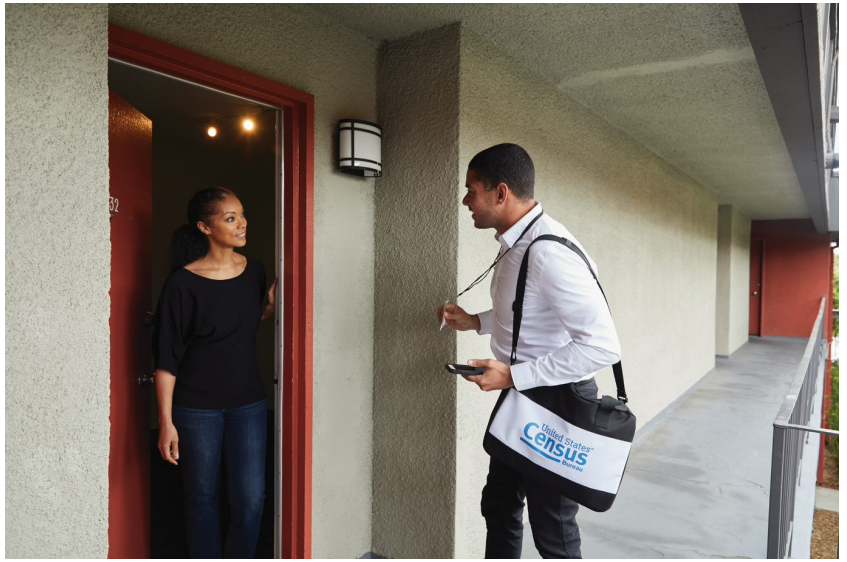
For more information and technical details relating to the issues discussed in these slides, please contact the author at michael.b.hawes@census.gov.

Any opinions and viewpoints expressed in this presentation are the author's own, and do not necessarily represent the opinions or viewpoints of the U.S. Census Bureau.

Our Commitment to Data Stewardship

Data stewardship is central to the Census Bureau's mission to produce high-quality statistics about the people and economy of the United States.

Our commitment to protect the privacy of our respondents and the confidentiality of their data is both a legal obligation and a core component of our institutional culture.



It's the Law

"To stimulate public cooperation necessary for an accurate census…Congress has provided assurances that information furnished by individuals is to be treated as confidential. Title 13 U.S.C. §§ 8(b) and 9(a) explicitly provide for nondisclosure of certain census data, and ***no discretion is provided to the Census Bureau on whether or not to disclose such data****…"* (U.S. Supreme Court, Baldrige v. Shapiro, 1982)

Title 13, Section 9 of the United State Code prohibits the Census Bureau from releasing identifiable data "furnished by any particular establishment or individual."

Census Bureau employees are sworn for life to safeguard respondents' information.

Penalties for violating these protections can include fines of up to $250,000, and/or imprisonment for up to five years!

Keeping the Public's Trust

Safeguarding the public's data is about more than just complying with the law!

The quality and accuracy of our censuses and surveys depend on our ability to keep the public's trust.

In an era of declining trust in government, increasingly common corporate data breaches, and declining response rates to surveys, we must do everything we can to keep our promise to protect the confidentiality of our respondent's data.

Upholding our Promise: Today and Tomorrow

We cannot merely consider privacy threats that exist today.

We must ensure that our disclosure avoidance methods are also sufficient to protect against the threats of tomorrow!

The Privacy Challenge

Every time you release any statistic calculated from a confidential data source you "leak" a small amount of private information.

If you release too many statistics, too accurately, you will eventually reveal the entire underlying confidential data source.

Dinur, Irit and Kobbi Nissim (2003) "Revealing Information while Preserving Privacy" PODS, June 9-12, 2003, San Diego, CA

The Growing Privacy Threat

More Data and Faster Computers!

In today's digital age, there has been a proliferation of databases that could potentially be used to attempt to undermine the privacy protections of our statistical data products.

Similarly, today's computers are able to perform complex, large-scale calculations with increasing ease.

These parallel trends represent new threats to our ability to safeguard respondents' data.

The Census Bureau's Privacy Protections Over Time

Throughout its history, the Census Bureau has been at the forefront of the design and implementation of statistical methods to safeguard respondent data.

Over the decades, as we have increased the number and detail of the data products we release, so too have we improved the statistical techniques we use to protect those data.

Reconstruction

The recreation of individual-level data from tabular or aggregate data.

If you release enough tables or statistics, eventually there will be a unique solution for what the underlying individual-level data were.

Computer algorithms can do this very easily.

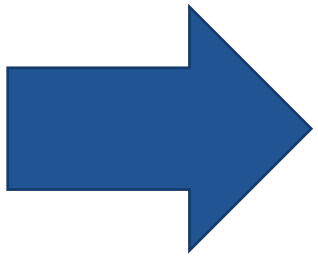
	4						2	
			7					4
1		7	8				5	
			9			3		8
5								
			6		8			
3						4		5
	8	5				1		9
		9		7	1			

Reconstruction: An Example

	Count	Median Age	Mean Age
Total	7	30	38
Female	4	30	33.5
Male	3	30	44
Black	4	51	48.5
White	3	24	24
Married	4	51	54
Black Female	3	36	36.7

Reconstruction: An Example

	Count	Median Age	Mean Age
Total	7	30	38
Female	4	30	33.5
Male	3	30	44
Black	4	51	48.5
White	3	24	24
Married	4	51	54
Black Female	3	36	36.7



Age	Sex	Race	Relationship
66	Female	Black	Married
84	Male	Black	Married
30	Male	White	Married
36	Female	Black	Married
8	Female	Black	Single
18	Male	White	Single
24	Female	White	Single

This table can be expressed by 164 equations.
Solving those equations takes 0.2 seconds on a 2013 MacBook Pro.

Re-identification

Linking public data to external data sources to re-identify specific individuals within the data.

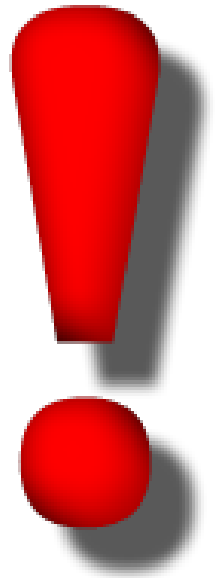
Name	Age	Sex		Age	Sex	Race	Relationship
Jane Smith	66	Female		66	Female	Black	Married
Joe Public	84	Male		84	Male	Black	Married
John Citizen	30	Male		30	Male	White	Married

External Data

Confidential Data

The Census Bureau's Decision

- Advances in computing power and the availability of external data sources make database reconstruction and re-identification increasingly likely.
- The Census Bureau recognized that its traditional disclosure avoidance methods are increasingly insufficient to counter these risks.
- To meet its continuing obligations to safeguard respondent information, the Census Bureau has committed to modernizing its approach to privacy protections.



Differential Privacy

aka “Formal Privacy”

-quantifies the precise amount of privacy risk…

-for all calculations/tables/data products produced…

-no matter what external data is available…

-now, or at any point in the future!

Precise amounts of noise

Differential privacy allows us to inject a precisely calibrated amount of noise into the data to control the privacy risk of any calculation or statistic.

Privacy vs. Accuracy

The only way to absolutely eliminate all risk of re-identification would be to never release any usable data.

Differential privacy allows you to quantify a precise level of "acceptable risk," and to precisely calibrate where on the privacy/accuracy spectrum the resulting data will be.

Data Quality | Bnae Kegouqe
Dada Qualitg | Vrkk Jzcfkdy
Data Qaality | Dncb PrhvBln
Dzte Qvality | Dncb Prtnavy
Dfha Quapyti | Tgta Ppijacy
Tgta Qucjity | Dfha Pnjvico
Dncb Qhulitn | Dzhe Njivaci
Ntue Quevdto | Dzte Privecy
Vrkk Zuhnvry | Dada Privacg
Bnaq Denorbe | Data Privacy

Establishing a Privacy-loss Budget

This measure is called the "Privacy-loss Budget" (PLB) or "Epsilon."

- **$\varepsilon=0$ (perfect privacy) would result in completely useless data**
- **$\varepsilon=\infty$ (perfect accuracy) would result in releasing the data in fully identifiable form**

Epsilon

Implications for the 2020 Decennial Census

The switch to Differential Privacy will not change the constitutional mandate to apportion the House of Representatives according to the actual enumeration.

As in 2000 and 2010, the Census Bureau will apply privacy protections to the PL94-171 redistricting data.

The switch to Differential Privacy requires us to re-evaluate the quantity of statistics and tabulations that we will release, because each additional statistic uses up a fraction of the privacy-loss budget (epsilon).

How the Census Bureau Protects Your Data

The U.S. Census Bureau is bound by law to protect your answers and keep them strictly confidential. In fact, every employee takes an oath to protect your personal information for life.

SHARE:

https://2020census.gov/en/data-protection.html

Questions?

Disclosure Avoidance and the 2020 Census Website
https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html

Michael Hawes
Senior Advisor for Data Access and Privacy
Research and Methodology Directorate
U.S. Census Bureau

301-763-1960 (Office)
michael.b.hawes@census.gov