Stakeholder Participation and Engagement in the Design and Tuning of the 2020 Census Disclosure Avoidance System

Michael B. Hawes
Senior Statistician for Scientific Communication
U.S. Census Bureau

July 29, 2024
Open Government in Action:
Emerging practices in Participatory Algorithm Design

Any opinions or viewpoints are the presenter's own and do not reflect the opinions or viewpoints of the U.S. Census Bureau

United States® Census Bureau

U.S. Census Bureau

The Census Bureau's mission is to serve as the nation’s leading provider of quality data about its people and economy.

- Decennial Census of Population and Housing
- Economic Census
- Census of Governments

…and over 100 demographic and economic surveys on a monthly, quarterly, or annual basis.

United States® Census Bureau

U.S. Census Bureau at a Glance

13 U.S. Code

Section 8(b): "…the Secretary may furnish copies of tabulations and other statistical materials **which do not disclose the information reported by, or on behalf of, any particular respondent**…"

Section 9: "[The Census Bureau may not] make any publication **whereby the data furnished by any particular establishment or individual under this title can be identified**"

The Challenge of Disclosure Avoidance

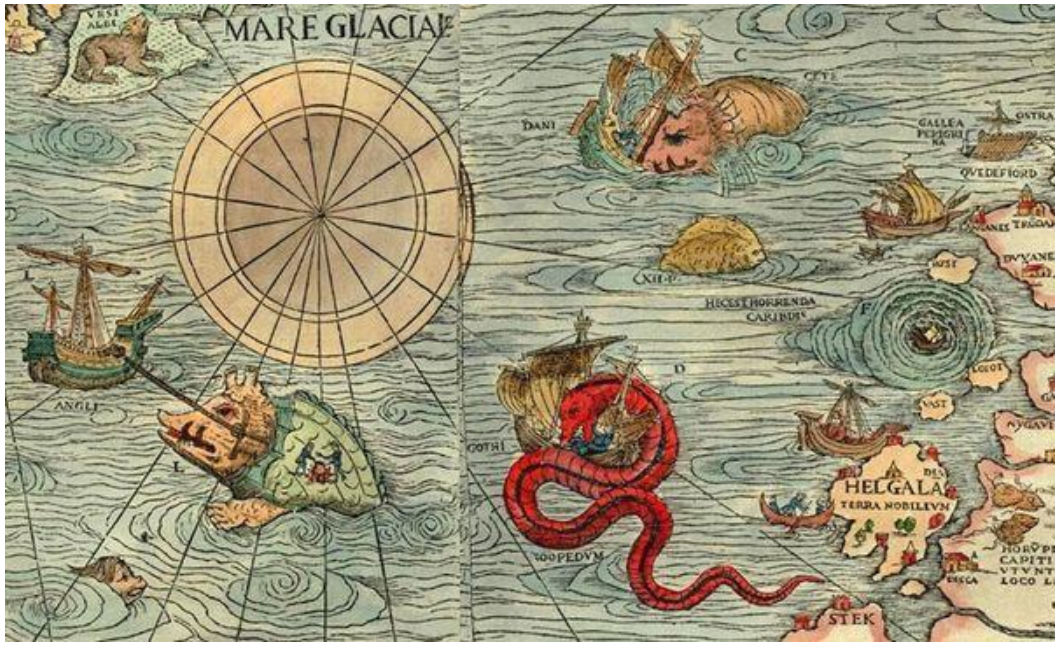
"It is the responsibility of Federal statistical agencies and recognized statistical units to produce and disseminate relevant and timely information; conduct credible, accurate, and objective statistical activities; and protect the trust of information providers by ensuring confidentiality and exclusive statistical use of their responses"

-OMB Statistical Policy Directive No.1 (2014)

"It has long been recognized that any available tabulation of the characteristics of a population is likely to narrow the range of uncertainty about the characteristics of specific individuals known to be members of that population…The release of any data usually entails at least some element of risk. A decision to eliminate all risk of disclosure would curtail statistical releases drastically, if not completely…"

-FCSM Statistical Policy Working Paper #2 (1979)

HIC SUNT DRACONES



United States® Census Bureau

The Triple Tradeoff of Official Statistics

The more statistics you publish, and the greater the granularity and accuracy of those statistics, the greater the disclosure risk.

All statistical techniques to protect confidentiality impose a tradeoff between the degree of data protection and the resulting availability and accuracy of the statistics.

You can maximize on any two dimensions, but only at profound cost to the third.

United States Census Bureau

Disclosure Avoidance for the 2020 Census

The 2020 Census improves on the noise injection methods of the 1990-2010 Censuses by employing a mathematical framework known as Differential Privacy (DP) to assess and quantify disclosure risk and confidentiality protection.

Every individual that is reflected in a particular statistic contributes towards that statistic's value.

Every statistic that you publish "leaks" a small amount of private information.

DP as a framework allows you to assess each individual's contribution to the statistic, and to measure (and thus, limit) how much information about them will leak.

United States® Census Bureau

Ensuring Fitness-for-Use

All disclosure avoidance methods, and the parameters of their implementation, impact the resulting data's fitness-for-use in different ways.

Agencies must be deliberate in their selection and implementation of disclosure avoidance methods to ensure they meet the needs of their intended data users.

Requires:

- Subject Matter Expertise
- Research and Evaluation
- Stakeholder Communication and Engagement

United States® Census Bureau

The TopDown Algorithm

For complete details see: Abowd, J., Ashmead, R., Cumings-Menon, R., Garfinkel, S., Heineck, M., Heiss, C., Johns, R., Kifer, D., Leclerc, P., Machanavajjhala, A., Moran, B., Sexton, W., Spence, M., & Zhuravlev, P. (2022). The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harvard Data Science Review*. (June) https://doi.org/10.1162/99608f92.529e3cb9

United States® Census Bureau

TDA Query Structure

TDA only takes noisy measurements for defined queries (tabulations) at particular geographic levels. Adjusting the queries asked and/or the share of privacy-loss budget (PLB) assigned to those queries determine the resulting amount of noise injected into the DHC statistics derived from those queries.

DHC-P PLB allocations by geographic level and query as reflected in the 2022-03-16 Demonstration Data Product

Global *rho*	3.325
Global *epsilon*	20.01
delta	10^{-10}

	***rho* Allocation by Geographic Level**
US	1.95%
State	27.07%
County	8.42%
Population Estimates Primitive Geography†	12.93%
Tract Subset Group‡	12.93%
Tract Subset‡	23.46%
Optimized Block Group◊	12.93%
Block	0.30%

	Per Query *rho* Allocation by Geographic Level							
Query	US	State	County	Population Estimates Primitive Geography†	Tract Subset Group‡	Tract Subset‡	Optimized Block Group◊	Block
AGE (3 bins) * HHGQ (4 Levels) (12 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
AGE (3 bins) * SEX (6 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
AGE (13 bins) * SEX (26 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
HISPANIC * SEX (4 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
SEX * HHGQ (4 levels) (8 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
HISPANIC * SEX * AGE (13 bins) * HHGQ (8 levels) * CENRACE (26,208 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
HHGQ (8 levels) * AGE (23 bins) * HISPANIC * CENRACE * SEX (46,368 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
RELGQ * AGE (23 bins) * HISPANIC * CENRACE * SEX (243,432 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%
RELGQ * SEX * AGE (116 bins) * HISPANIC * CENRACE (1,227,744 cells)	0.22%	3.01%	0.94%	1.44%	1.44%	2.61%	1.44%	0.03%

United States® Census Bureau

Questions and Discussion