Data Disclosure and the ACS

Looking Forward Using Lessons from the Past

ACS Data Users Conference
May 16, 2023

"You have to know the past to understand the present."
— Carl Sagan

Joseph Salvo
Institute Fellow
University of Virginia Biocomplexity Institute

UVA | BIOCOMPLEXITY INSTITUTE

Lessons from the Past

"Perhaps one of the greatest lessons learned from the Census Bureau's experience transitioning to DP is the **active role that data users can and should play in shaping the disclosure limitation system.**"

John M. Abowd and Michael B. Hawes
"Confidentiality Protection in the 2020 US Census of Population and Housing"
Annual Review of Statistics and Its Application 2023 10:1, pp 23-24.
https://doi.org/10.1146/annurev-statistics-010422-034226

More Lessons from the Past

"…the Census Bureau's external stakeholders were justifiably disappointed and concerned about the timing, pace, and mechanisms of the agency's public engagement regarding the development of the 2020 DAS… the urgency of the need to transition to a new disclosure avoidance mechanism both precluded some stakeholder engagement that would have been valuable and condensed the time-frame for the engagement that could occur. **Going forward, however, the Census Bureau must explore additional approaches to engaging stakeholders."**

John M. Abowd and Michael B. Hawes
"Confidentiality Protection in the 2020 US Census of Population and Housing"
Annual Review of Statistics and Its Application 2023 10:1, pp 23-24.
https://doi.org/10.1146/annurev-statistics-010422-034226

User Data Needs First

User-Centric Approach

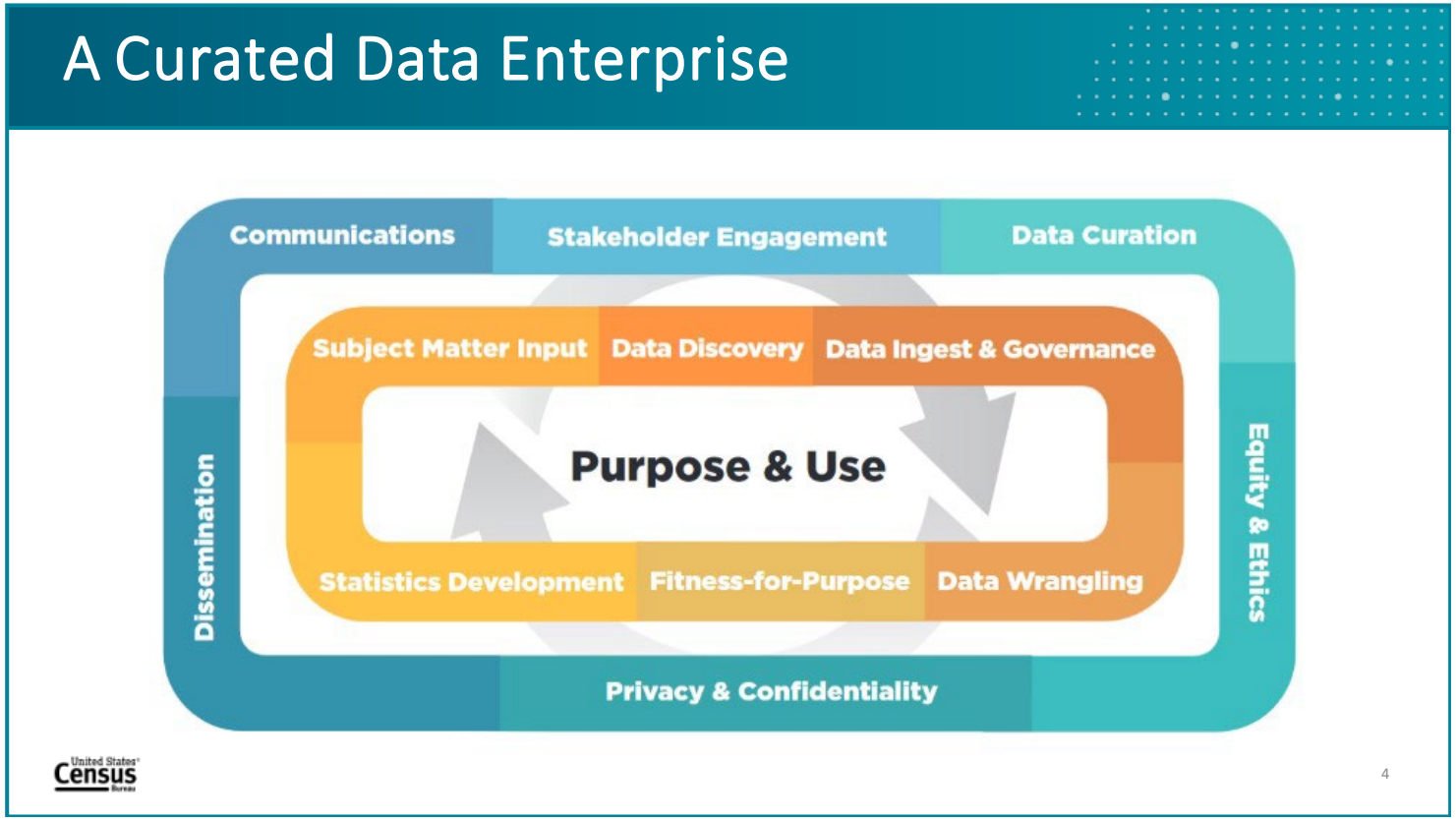
Based on **collaboration:** with the Data User Community, defined as:

“work jointly with others or together especially in an intellectual endeavor”
Merriam-Webster’s Dictionary

The Census Bureau's New Enterprise Framework

Purpose and Use drive the creation of statistical products

Stakeholder needs as a key part of the process – **on par with** privacy and confidentiality considerations



The Best Laid Plans…?

CDE "Statistical Product First" framework

How do you know if the synthetic data approach for the Public Use Microdata Sample (PUMS) will work, not only from a technical perspective, but from the perspective of those who will use the data?

The development of a **PUMS tabulation "validation server"** as a litmus test of this approach using real-life statistical data applications.

Photo by "My Life Through A Lens" on Unsplash

Age and Educational Attainment for Total Persons and Domestic In-Migrants				
New York City				
2009				
	Total, NYC		Domestic In-Migrants*	
Age				
Total, Persons	8,391,066	100.0	158,990	100.0
Under 20	2,098,722	25.0	33,244	20.9
20 - 24 Years	561,034	6.7	28,943	18.2
25 - 29 Years	718,696	8.6	34,133	21.5
30 - 34 Years	703,511	8.4	18,302	11.5
35 - 54 Years	2,389,111	28.5	30,818	19.4
55 and Over	1,919,992	22.9	13,550	8.5
Educational Attainment				
Total, 25 and Over	5,731,310	100.0	96,803	100.0
Less than High School	1,194,704	20.8	10,718	11.1
High School or GED	1,406,662	24.5	14,465	14.9
Some College	1,179,744	20.6	17,522	18.1
BS or higher	1,950,200	34.0	54,098	55.9
BS only	1,142,178	19.9	29,741	30.7
Masters	545,406	9.5	16,160	16.7
Doctorate or Professional Degree	262,616	4.6	8,197	8.5
*Domestic in-migrants are persons living in NYC in 2009 who lived in another state one year ago.				
U.S. Census Bureau				
2009 American Community Survey, Public Use Microdata Sample				
Population Division NYC Department of City Planning				

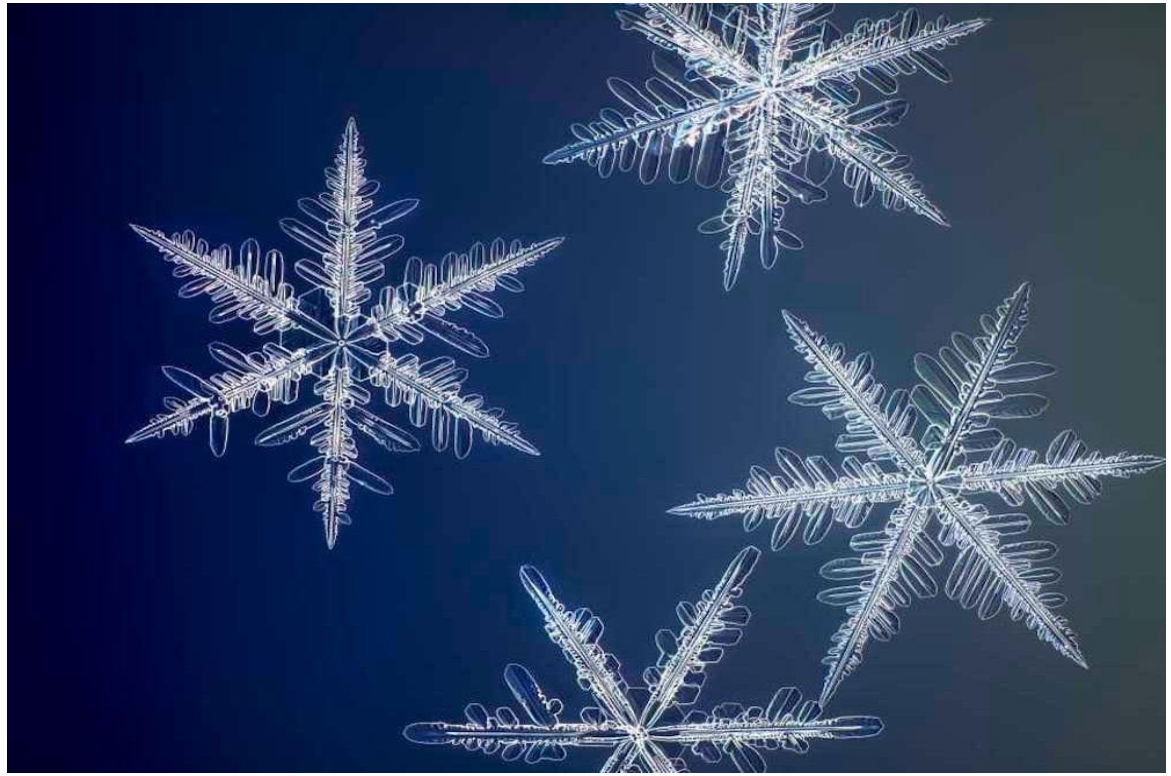
Request from City Hall at 10 AM on Monday that needs to be completed by 5:00 PM on Tuesday, with data and bullets for a press conference that includes information about the age and educational attainment of domestic in-migrants to New York City.

Collaboration and Disclosure Avoidance

There are no perfect data and there are ways of measuring the level of existing "noise," such as the level of imputation that is applied to the data.

Take existing "noise" into account when applying and evaluating disclosure avoidance algorithms as was done in the past (i.e., 2000 Census).

Noise associated with sampling, as per Spring 2023 CSAC Recommendation



"31. CSAC recommends the Bureau investigate recent research works on the implications of sampling on privacy, evaluate how that can influence the design of sampling methods, and determine how sampling could inform disclosure avoidance methodologies." March, 10, 2023

Key Questions

1. When does the application of a DAS algorithm compromise the quality of a statistical data product?
2. How can collaboration be achieved, given confidentiality issues? (Hint: FSRDCs are important but **not** the answer.)
3. Are there examples of successful collaborations with the data user community on product development?
4. How to achieve statistical product equity?

The Association of Public Data Users (APDU) ***Working Group on 2000 Census Products*****.**

"Prior to finalizing decisions about the number, content, and format of Census 2000 data products, **the Census Bureau sought advice from a wide variety of data users….** As part of this process, the agency **contracted with the Association of Public Data Users (APDU) to form a working group of data users who would review and provide advice on the details of individual data products as well as on the total integrated product proposal**."

U.S. Census Bureau, *History: 2000 Census of Population and Housing* (Volume 2) U.S. Government Printing Office, Washington, DC, 2009, P. 409.

A Hopeful Sign!

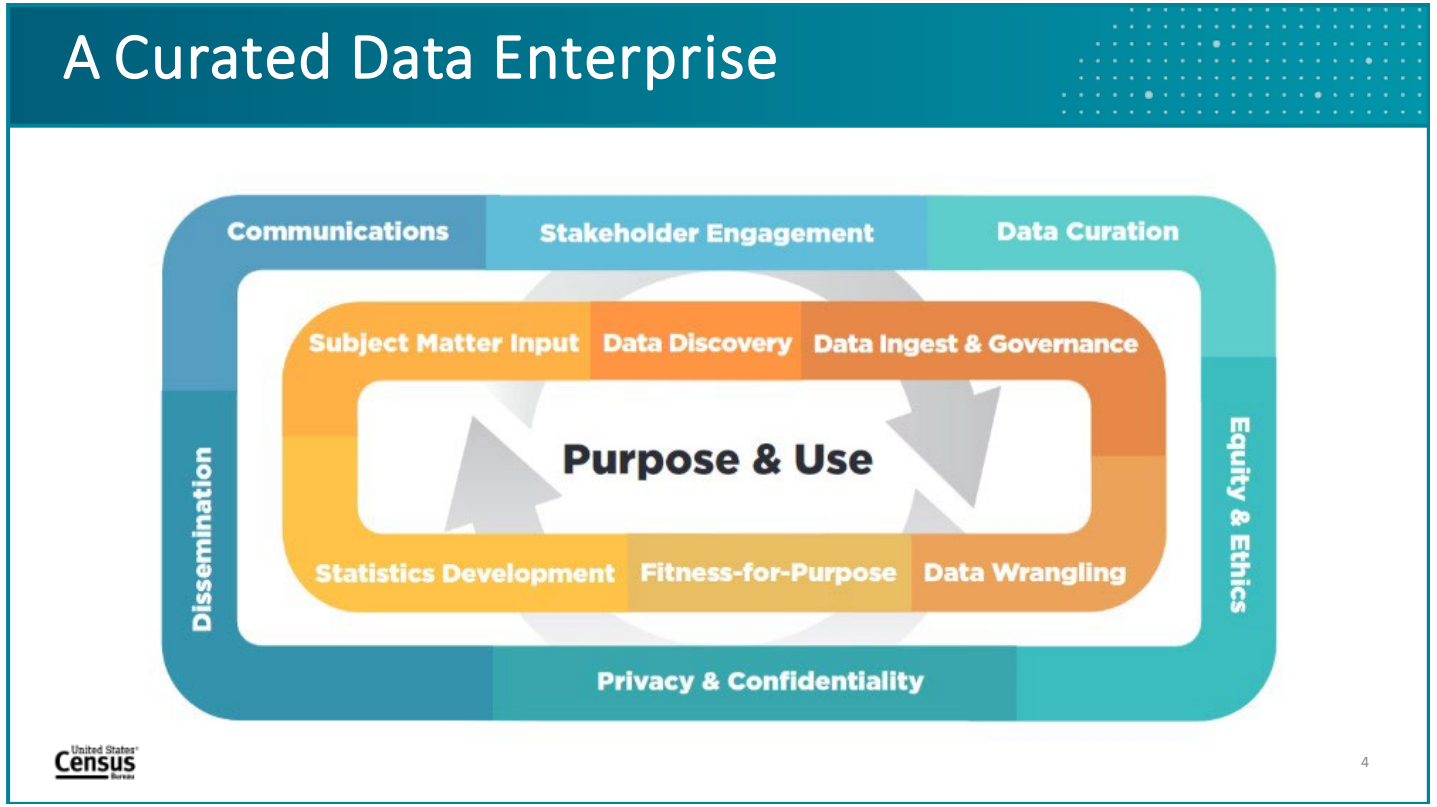
We're all in this *together!*

- We cannot escape the triple trade-off
- Successfully and efficiently managing the triple trade-off means embracing difficult conversations and accepting change
- Transparency and respectful, honest dialogue are key
- Stakeholders need to acknowledge there must be trade-offs and participate in the discussion

Presentation by Sallie Ann Keller at the 2023 Annual Meeting of the Population Association of America, April 13, New Orleans.

How to “Rise to the Occasion”

1. Use the **“statistical product first”** framework in the spirit of real collaboration.
2. Consider **alternative methods** that acknowledge the “noise” already inherent in the data.
3. Reality test any proposed DAS using **real-life applications** and **testing done in concert with data users to gauge utility, purpose and use**.



Thank you!

UVA | BIOCOMPLEXITY INSTITUTE