The Future of Public Data: Innovation at the U.S. Census Bureau

Presentation to the Association of Public Data Users

John H. Thompson, Director
U.S. Census Bureau

September 14, 2016
The 2020 Census
A Changing Environment and a New Design

The 2020 Census requires a flexible design that takes advantages of new technologies and data sources, while minimizing risk to ensure a high quality population count.
The 2020 Census
Goals and Key Innovation Areas

**Goal:** To count everyone once, only once, and in the right place

**Challenge:** Conduct the 2020 Census at a lower cost per housing unit (adjusted for inflation) than the 2010 Census, while maintaining high quality results

Four Key Innovation Areas

- **Reengineering Address Canvassing**
- **Optimizing Self-Response**
- **Utilizing Administrative Records and Third-Party Data**
- **Reengineering Field Operations**
The 2020 Census

Estimated Lifecycle Costs

- $1.1 B in 1970
- $3.0 B in 1980
- $4.7 B in 1990
- $9.4 B in 2000
- $12.3 B in 2010
- $17.8 B (Traditional 2020 Census)
- $12.5 B (Innovative 2020 Census)

Less Staff
Fewer Offices
Less Burden

MORE THAN $5 BILLION IN SAVINGS
Adaptive Design
Making Data Collection Faster, Better and Cheaper

Information on costs
+ Fieldwork quality
+ Estimates of key variables

Modified and improved data collection efforts

National Survey of College Graduates
Adaptive treatments produced significantly better results in improving survey representativeness, without affecting response rates compared to experimental control groups

Survey of Income and Program Participation and National Health Interview Survey
Case prioritization through adaptive treatments helps survey directors make informed cost/quality trade-off decisions during data collection operations

Concurrent Analysis & Estimation System
Big Data distributed analytical processing using a Hadoop cluster, leading the way for high volume/high speed computer processing for real-time adaptive interventions
Administrative Records
Improving Data Collection and Dissemination

Before data collection
- Frame
- Predict (i.e. mode, best time to contact)
- Contact

During data collection
- Edit
- Impute
- Replace

After data release
- Evaluate
- Model
- Innovate
Disclosure Avoidance
Disseminating Data While Honoring Privacy & Confidentiality

U.S.C. Title 13 promises confidentiality to respondents; collected data is used for statistical purposes only

U.S.C. Title 26 mandates that statistical products are subject to disclosure avoidance procedures

Confidential Information Protection & Statistical Efficiency Act provides confidentiality protections for statistical information collections

Center for Disclosure Avoidance Research areas of study:

- Provably private disclosure limitation methods for the 2020 Census
- Synthetic data disclosure limitation methods for the American Community Survey
- Synthetic data with validation for the 2017 Economic Census
Improving Economic Statistics
Meeting User Demands for Timelier, More Granular, Linkable Data

Retail trade is our first focus:

- Improve quality of Monthly Retail indicator data
- Improve the timeliness of retail estimates
- Improve the granularity of retail estimates
- Reducing burden and improving value to retail data providers
- Improve and modernize Census Retail Trade Statistics
- Improve E-Commerce Measures
Center for Enterprise Dissemination Services and Consumer Innovation

Shared, Reusable Systems for Dissemination
CitySDK
Toolbox for Developers and Civic Innovators to Connect to Local & National Public Data
Opportunity Project
Development Initiative to Increase Access to Fair Housing

Collaboration between the U.S. Department of Housing and Urban Development, the White House, the Census Bureau, and cities and local governments

Helps cities and local governments use new, curated, open data to account for how they use federal housing dollars, and increase access to fair housing
Census Business Builder

Small Business and Regional Analyst Editions

Fairfax County, Virginia

- Total population: 1,117,072
- Median household income: $112,102
- Percent high school degree or higher: 91.9%
- Homeownership rate: 68.2%

Employer establishments (County - 2013, City - 2012):
- 9 - 11
- 7 - 8
- 5 - 6
- 3 - 4
- 1 - 2
- Not Available/Suppressed

Potential Customers?

Business Profile of Mobile Food Services (NAICS 72233) in Fairfax County, Virginia

- Total Population
- Median Household Income
- Median Household Unit Value

Source: 2009-2013 American Community Survey 5-Year Estimates
Center for Big Data Research and Applications

Innovation Measurement Initiative

Project from the Census Bureau, U. Michigan, and U. Chicago that links R&D and innovation by integrating data on federally-funded university grants with Census Bureau data assets.

*UMETRICS DATA*
University data on Federal awards:
- Employee, vendor, subaward transactions

*CENSUS DATA*
Secure data on people and businesses:
- Employment records, business dynamics & characteristics

*JOB PLACEMENTS*
Where research employees get their next jobs

*START-UP ACTIVITY*
What types of businesses research employees found

*VENDOR CHARACTERISTICS*
What types of businesses supply research

**Analyze by:** Occupational category | Funding agency | Research area | Years since leaving university
Future On
Embracing Transformation Internally
Thank you!