Big Data Projects at the Census Bureau
presented to the
Council of Professional Associations on Federal Statistics

William G. Bostic, Jr.
Associate Director, Economic Programs
U.S. Census Bureau

March 1, 2013

Is the Census Bureau working on Big Data projects? If so, how are these different from other projects?

- Yes, incorporating Big Data solutions to improve the efficiency of operations throughout the survey lifecycle

- Projects will result in enterprise-wide solutions that support **all** surveys and census operations; this is different from our current processes and technology that are developed to support individual surveys and census operations

United States™
Census
Bureau

Big Data Projects at the Census Bureau

**The Census Bureau
Survey Life Cycle**

Data Collection

- Multi-Mode Data Collection model
- Systems architecture

Data Analysis

- Methodological techniques
- Exploring technology solutions

Data Dissemination

- New platforms for integrating census and survey data with other Big Data

Potential Uses of Big Data at the Census Bureau

How can the official statistical community integrate Big Data with census data, survey data or other administrative data?

- Integrate novel sources of data to:
  - increase timeliness,
  - increase explanatory power of the data (i.e., more granular levels of geography or industry classifications,) and
  - reduce operational costs of data collection.

Potential Uses of Big Data at the Census Bureau

- Electronic transactions and administrative data to supplement or improve statistics

- Data analysis that takes into account larger than ever collections of data

- Using Big Data techniques, such a modeling, looking for patterns, or mashing up data, to improve logistical operations

United States™
Census
Bureau

- 2020 Census

- Construction Statistics

- Retail and Service Statistics

Opportunities Using Big Data: 2020 Census Opportunities with Big Data

- More effective use of existing data such as administrative records
- Incorporating new data into decennial operations
  - Paradata from Internet Data Capture
  - Information from Social Media Feeds
- Aligning organization assets (systems, personnel) to maximize the use of big data
- Building flexibility into decennial census operations to accommodate unknown sources of big data

Opportunities Using Big Data: Construction Statistics

- Foreclosures
  - Use vendor data on new residential properties in foreclosure to aid analysis of data on new construction and sales
- Public Construction
  - Partner with government agencies to obtain data on construction spending
- Building Permits
  - Best opportunity for using Big Data
  - Local jurisdictions and state agencies posting public records online

United States Census Bureau

Opportunities Using Big Data: Retail and Service Statistics

- Evaluating electronic payment processing to fill data gaps such as geographical detail and revenue measures by firm size
  - New data products
  - Improvements to current data quality
- Considerations/Challenges/Risks
  - Confidentiality restrictions
  - Cost
  - Differences in survey concepts and definitions; differences in classification
  - Data quality and completeness
  - Where's the money?
- Other potential opportunities

Integrating other Big Data with Census Bureau Data

- Embracing the business intent of the Federal Digital Government Strategy of releasing public use data to drive digital services that promote innovation

Deliver better services to customers at a lower cost.

Increased speed to market.

Provide meaningful insights that can drive business decisions.

Proliferate innovation opportunities.

Provide accessibility to products and services anywhere and any time.

Integrating other Big Data with Census Bureau Data

- Recently released an API (http://www.census.gov/developers/) to make data from the American Community Survey and the 2010 Census available for easy access integration with other community Big Data

- Making data available through a diverse array of digital products and services (http://www.census.gov/main/www/access.html) including:
  - Custom applications such as Data Ferrett, Hot Reports, American Fact Finder, and Census Indicators Data Repository
  - FTP and hard copy DVDs for large data download
  - Mobile Applications such as America's Economy and OnTheMap
  - Remote Data Centers to facilitate research

United States™
Census
Bureau

Providing Big Data: American Community Survey

- Five-year estimates, released yearly
  - Contain almost 11 billion estimates
  - Deliver on the Big Data promise
  - Represent a major statistical innovation, providing detailed economic, social, demographic, and housing information for over 700,000 communities
- December 2012: Census released the Equal Employment Opportunity Tabulation, relying on the ACS estimates
  - More than 19 billion statistics, requiring more than 1 trillion calculations
  - Provides a comprehensive profile of the American workforce

Considerations

- Statistics provide a benchmark

- Statistics often inform public policy
  - Billions in federal funds are allocated to communities based on official statistics

- Data quality is critical
  - Potential for false discovery
  - Explore new statistical methods to ensure solutions

What's on the Horizon

- Developing a process to coordinate and facilitate Big Data activities across the Census Bureau

- Looking to implement a process to systematically explore data products and processes that might be improved by Big Data sources and techniques

- Thinking carefully about how we tackle Big Data economically

United States™
Census
Bureau