COPAFS Quarterly Meeting

Ron Jarmin, Acting Director
U.S. Census Bureau

June 4, 2021
Synthetic Data for ACS Respondent Privacy

• In 2018, internal research revealed that traditional privacy protection techniques used for the 2010 Census were ineffective against today’s computational threats.
• That discovery prompted the Census Bureau to commit to modernizing protections of all its data releases.
• Synthetic data are a means of protecting respondent privacy while maintaining data utility.
• The Census Bureau currently uses synthetic data in several programs and products, such as the Longitudinal Employer-Household Dynamics (LEHD), ACS, SIPP, and OnTheMap web application.
Synthetic Data Continued…

- The Census Bureau is researching a new fully synthetic data product for the ACS to explore whether this method would allow us to produce more accurate data—correcting for known sources of error and potentially allowing for more tabulations at lower levels of geography—for our users while maintaining our respondents’ privacy.
- The Census Bureau hasn’t made any decisions yet about the use of fully synthetic data in the ACS.
- There are no plans to use fully synthetic data as a source for ACS table releases.
- Our goal has always been to produce reliable and high-quality demographic, socioeconomic, and housing data from the American Community Survey (ACS).
Disclosure Avoidance

Disclosure avoidance methods seek to make reconstruction and re-identification more difficult, by:

• Reducing precision
• Removing vulnerable records, or
• Adding uncertainty

Commonly used (legacy) methods include:

• Complementary suppression
• Rounding
• Top/Bottom coding of extreme values
• Sampling
• Record swapping
• Noise injection
2020 Census
A Success Story

• Determined a final resolution for over 99% of addresses.

• Increased use of technology at every level from offering an internet response option (without a single second of down time) to use of iPhones for field data collection.

• Exceeded goal for data collected through self-response – the highest quality response.

• Overcame the challenges of a global pandemic, wildfires, and a historic hurricane season.

• Achieved a similar quality profile to past censuses in terms of the metrics available at this point.
2020 Census
Resolution of Addresses

- Every address in the 2020 Census universe must be assigned a resolution.
- Resolution of an address involves a determination of status (occupied, vacant, or nonexistent) and when occupied, collection of census response information for all persons at an address.

- Over 97% of addresses were resolved through self-response (internet, phone or paper) or nonresponse followup (household interview, high-quality administrative records, or proxy interview).

- Almost 2% of addresses were resolved through operations targeted at unique geographies or populations such as addresses in remote areas of Alaska or addresses for group quarters, where people live and receive services.

- Less than 1% of addresses were unresolved and required imputation to assign a status, and when occupied, to assign a population count.
2020 Census
Operational Quality Metrics: April 2021

• No single number can definitively quantify the quality of the census.
• Operational metrics are data points related to progress and results of census operations.
• Insights gained through comparisons by state and comparisons to past census results.
• Differences are:
  – A result of changes within the Census Bureau’s control such as changes in the operational design since 2010.
  – A result of changes outside of the Census Bureau’s control such as normal changes in the population, respondent behavior, and COVID-19.
  – Expected across geographies and when making comparisons to the 2010 Census.
• Different doesn’t necessarily mean “better” or “worse.”
## U.S. Apportionment Population

### U.S. Apportionment Population
50 States, Excluding DC

<table>
<thead>
<tr>
<th></th>
<th>2020</th>
<th>2010</th>
<th>Numeric Change</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>331,108,434</td>
<td>309,183,463</td>
<td>21,924,971</td>
<td>7.1%</td>
</tr>
</tbody>
</table>

### Average Apportionment Population Per Representative

<table>
<thead>
<tr>
<th></th>
<th>2020</th>
<th>2010</th>
<th>Numeric Change</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>761,169</td>
<td>710,767</td>
<td>50,402</td>
<td>7.1%</td>
</tr>
</tbody>
</table>
Economic Program Update:
Experimental Data Products

- Guiding principles, processes, and external presence established 2019
- Innovative statistical products, use new data sources or methods
- Provide a benefit in the absence of other products
- Gave a framework for the products that came together quickly to measure COVID-19 impacts

2020
- Weekly Small Business Pulse Survey
- Weekly and Monthly Business Formation Statistics
- Monthly State Sales Tax
- Monthly State Retail Sales
What’s the SBPS?
A weekly survey that identifies challenges faced by small businesses during the pandemic. Each phase consisted of 20 questions or less, taking 5 - 6 minutes to complete. The survey reaches ~1M small businesses split across nine weeks.

What are we collecting?
Near real time data from small businesses on operational challenges, finances and outlook, vaccine requirements and expectations.

Methodology and Collection
- 1M sample representative of 5.6M small businesses
- Includes all single-location businesses with 1-499 employees and $1000+ revenue that reported an email address; excludes non-employers or multi-units.
- Initial email sent Monday asking for a response by Thursday, due date reminder sent Wednesday, and final notice sent Friday
- Businesses have until the end of the phase to respond

Collaboration Partners
SBPS content has been developed jointly between external stakeholders including: MBDA, FRB, SBA, ITA, NTIA, BTS, BEA and BLS

Phases
Phase 1: April 26, 2020 – June 25, 2020
Phase 2: August 9, 2020 – October 10, 2020
Phase 3: November 19, 2020 – January 10, 2021
Phase 4: February 15, 2021 – April 18, 2021

Key Facts
- In March-April 2020, Small Business Pulse Survey went from concept to data release in 39 days
- Utilizes Census Bureau data and IT infrastructure
- Is the first economic survey conducted solely by email
- Weekly data releases
- Provides national estimates as well as state, top 50 metro, and sector detail
- Response rate of ~25%
Growing need for future financial assistance

In the next 6 months, do you think this business will need to do any of the following?

- Obtain financial assistance or additional capital
- Increase marketing or sales
- Learn how to better provide for the safety of customers and employees
- Identify new supply chain options
- Identify and hire new employees
- Develop online sales or websites
- Remodel or renovate
- Permanently close this business
- None of the above

Responses Collected 01/04 to 01/10
- 42.8% Obtain financial assistance or additional capital (National: 33.5%)
- 16.9% Learn how to better provide for the safety of customers and employees (National: 13.3%)
- 22% Identify and hire new employees (National: 23.6%)
- 5% Permanently close this business (National: 4.5%)

Note: No data collected between 6/27 and 8/9. No data collected between 10/12 and 11/9.
SBPS Development Timeline

- Start of Pandemic
- Concept
- Pre Collection / Planning
- April 26th Data Collection Begins
- May 14th First Estimates Published

Approximately 2 months, or 39 business days
Business Formation Statistics (BFS)

What is the BFS?
An experimental data product that provides high frequency, timely data on early-stage new business activity in the U.S. including data on business applications and formations.

Coverage
Data cover Employer Identification Number (EIN) applications made in the U.S. including those associated with starting a new employer business.

External Collaboration
Began as a research collaboration with economists at the Board of Governors of the Federal Reserve System, Federal Reserve Bank of Atlanta, University of Maryland, and University of Notre Dame.

Releases
Weekly: released approximately 11-12 days after the end of the observed month
Monthly: released approximately 5 days after week’s end

Content
- Business applications as indicated by applications for an EIN
- Actual and projected business formations originating from business applications based on the record of first payroll tax liability for an EIN
- Delay in business formation as indicated by the average duration between business application and business formation

Key Facts
- First released in beta form in February 2018
- July 2019, moved to a quarterly publication
- April 2020, began releasing weekly business application data at the national, regional, and state levels
- February 2021, business formation data at the national, regional, and state with industry detail levels are released monthly

What's the SBPS?
A weekly survey that provides high frequency data about the challenges small businesses are facing due to COVID-19. Consisting of 16 questions, this 5 minute survey reaches ~1M businesses split across a 9 week rotation to reduce burden and lessen survey fatigue.

What are we collecting?
Survey captures information on small business operations, challenges, finances, and expectations.

- Business applications as indicated by applications for an EIN
- Actual and projected business formations originating from business applications based on the record of first payroll tax liability for an EIN
- Delay in business formation as indicated by the average duration between business application and business formation
Business Applications in 2020

Weekly (unadjusted)

Monthly (seasonally adj.)

Quarterly (seasonally adj.)

Source: Quarterly, Monthly, and Weekly Business Formation Statistics
BFS Production Timeline

**Publications by Year**
- **2019** – 2 Quarterly
- **2020** – 41 Weekly, 4 Quarterly, 1 Annual
- **2021** (estimated) – 49 Weekly, 11 Monthly, 1 Quarterly, 1 Annual

- **February 10**, 2021
  - First Monthly BFS Release, including monthly series by 2-digit NAICS

- **November 11**, 2020
  - Annual release of county level Business Application series through 2019

- **April 9**, 2020
  - First release of Weekly BFS (national, regional) Weekly state data added

- **April 16**, 2020
  - First release of Weekly BFS by NAICS for 2019-2020

- **January 14**, 2020
  - Final Quarterly BFS Release

- **October 22**, 2019
  - Special release of Weekly BFS by NAICS for 2019-2020

- **Late Spring**, 2019 (tentative)
  - Annual county BA series through 2020

- **Late 2021**, 2019 (tentative)
  - Implement new HBA NAICS Criteria
Blended Model for Data Products

• New monthly, state-level retail sales issued in September 2020
• Achieved by blending survey data, administrative tax data, and data obtained through a third-party data aggregator
• Harnessing the power of alternative data sources, coupled with our survey data, to produce more timely, more detailed, and less burdensome products
Total Retail Sales Excluding Nonstore Retailers by State

October 2020 Y/Y % Change

Source: October 2020 MSRS Report
$ = Estimate suppressed due to quality concerns
* The 90 percent confidence interval includes zero. There is insufficient statistical evidence to conclude that the actual change is different from zero.
Note: State retail sales data not adjusted for seasonal variation, trading-day differences, moving holidays or price changes.
High Frequency Data

• Proved in with the weekly Small Business Pulse Survey, Business Formation Statistics

• Demonstrated ability to rapidly develop and release new data products in near-real time

• Continuing to build on these capabilities, partnering with federal statistical counterparts and third-party data providers

• Developing generic clearance to streamline administrative processes in the future
FRAMES Program: The Challenge

- Various frame-like datasets are in use at the Census Bureau.
- Contain information about addresses, businesses, jobs, and people.
- Some information is unique within a particular frame.
- Other information is duplicative (e.g., group quarters exist in both the Master Address File and Business Register).
- Frames exist in an uncoordinated and unintegrated environment.
- No process exists that allows for the direct linkage of information contained in one frame with information in any other frame.
FRAMES Program: The Vision

To create Enterprise-wide frames linkable in nature, agile in structure, accessible for production or research on a need-to-know basis, and that adhere to best practices in terms of technology usage, data management, and methodology.

Linkable in Nature

• Each frame will include the necessary unique identifiers and keys for linkage to each other. For example:
  • Location information (i.e., address and/or longitude/latitude) will be available on each person, job, and business record.
  • A person record may have linkage information for a job.
  • A job record may have linkage information to a business.
Creating an Infrastructure to Modernize the Census Bureau’s Statistical Foundation

**Data Types**
- Admin Data
- Census / Surveys
- Public Records
- Commercial Data

**Data Ingest and Collection**
- Geospatial
- Job
- Business
- Demographic

**Enterprise Frames**
- Addresses, MAFIDs, geocodes, and attributes (i.e., address has a non-residential use)
- EINs and attributes (i.e., business has employees)
- PIKs and attributes (i.e., person is a worker)

**Cross-Cutting Capabilities**
- Data Curation and Quality Assurance
- Data Update
- Data Utility Assessment

**Assumptions**
- Dedicated research staff is available
- Data is available upon ingestion
- Ingest once, allow access across frames
- Built in "real-time-updates" triggers across the frames
- IT issues resolved for efficiency and security and usability

**Possible alternatives:**
1. Data Efficacy and Usage Innovation
2. Data Evolution

---

*This diagram provides a generalized view of the two-way flows of specified data elements (MAFIDs, PIKs, EINs, addresses, geocodes, etc.) between pairs of frames.
Evolving to Meet 21st Century Data Needs

JANUARY 11, 2021

At the Census Bureau, trail-blazing is part of our DNA. We strive to innovate and advance the science behind the nation’s statistical infrastructure.

Today, the nation demands data about our people, places and economy that is more timely, accurate, and granular than ever before. New technology and data science make that possible in ways barely conceivable just a generation ago.

Big Data. Limitless cloud storage. Artificial intelligence. Faster processing. Growing broadband access. We are working to safely and securely leverage these and other modern-day tools to produce the data the nation needs.
Thank you

ron.s.Jarmin@census.gov