Innovation at the Census Bureau

How Big Data, New Technologies and Advanced Analytics Are Revolutionizing How We Measure America

John Thompson
Director, U.S. Census Bureau

United States™ Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

THE BACKGROUND

The **U.S. Census Bureau** is the leading source of quality data about the America's people, places and economy.

We collect and disseminate a wide array of data:

- Decennial census
- Economic census
- American Community Survey
- 13 principle economic indicators
- Demographic surveys – income, poverty & health insurance data
- Economic surveys
- Reimbursable surveys for other federal agencies

Hansen, Morris H., William. N. Hurwitz, and William G. Madow. *Survey Sample Methods and Theory.* Wiley, 1953.

Cochran, William G. *Sampling Techniques.* Wiley, 1953.

Waksberg, Joseph. "Sampling methods for random digit dialing." Journal of the American Statistical Association, 73 (1978): 40-46.

Fay, R.E. and Herriot, R.A. "Estimates of income for small places: An application of James-Stein procedures to census data." Journal of the American Statistical Association, 74 (1979): 269-277.

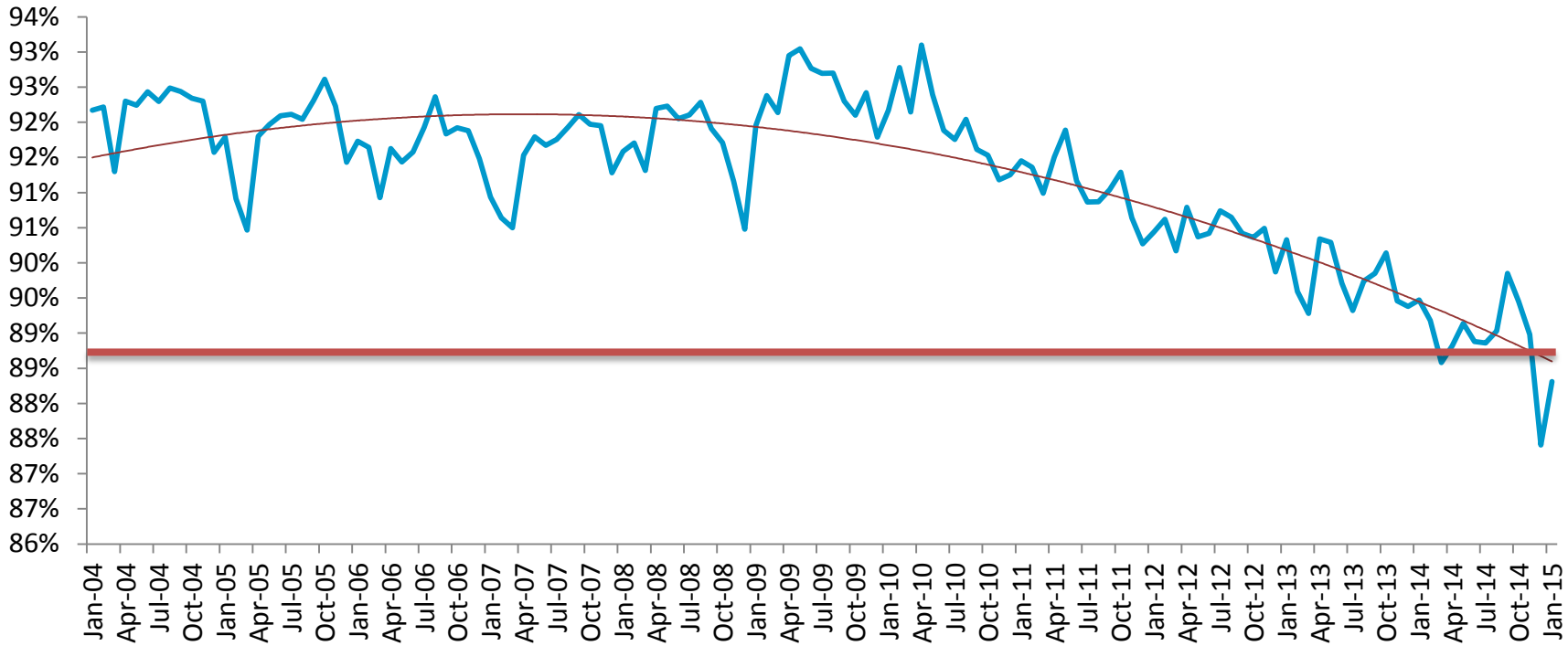
Fellegi, I.P. and Sunter, A.B. "A theory for record linkage." Journal of the American Statistical Association, 64 (1969): 1183-1210.

Dalenius, Tore. "Towards a methodology for statistical disclosure control." Statistisk tidskrift/ Statistical Review, 5 (1977): 429-444.

21st century challenges

Declining response rates

Current Population Survey response rates, 2004-2015



21st century challenges

Doing more with fewer resources

In an era of declining resources, data users want high-quality data.

- Federal budgets are shrinking
- Users' demands for more timely, granular data are increasing

21st century challenges

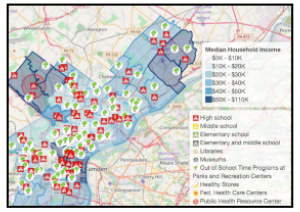
Emerging data needs

We need to meet our data users' growing demand for data that can be easily combined with other data sources

- Opportunities to use new technology and methods to combine data from multiple sources – federal, state, local, and private sector
- Exposing our data to meet users' needs in efficient ways

Examples of data user-generated tools from the Opportunity Project

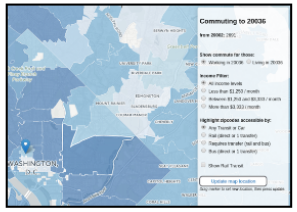
School ResourceMapper



Brings together economic, health center, and community institution data to help school leaders find and develop community partnerships.

Learn More

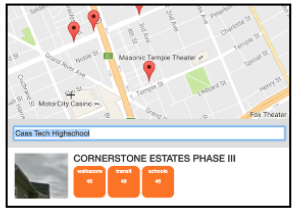
Commuter Map



Uses commuting data from the LEHD Origin-Destination Employment Statistics, bus & rail transit routes to show commuter flows and gaps in access to transit.

Learn More

FindHome



Helping low income families overcome the stress of finding affordable housing this tool combines transit access, walkability and school information into one unique feed of available properties.

Learn More

21st century challenges

Protecting privacy

We want to disseminate the most accurate, granular data possible. At the same time, we have the responsibility to protect the confidentiality of our respondents' data.

- Uncertain implications of the **database reconstruction theorem** for a data product as large-scale as the census

- Actively researching ways to deliver statistics that meet users' needs and limits reconstruction

Database reconstruction theorem
Too many statistics published too accurately from a confidential database exposes the entire database with certainty

Challenge: How many statistics are too many, and too accurate?

21st century challenges

Concerns about federal data collection

Respondent, public and oversight concerns about the American Community Survey, economic surveys & other data collection efforts

- Mandatory nature of surveys
- Intrusive questions
- Respondent burden

AMERICAN COMMUNITY SURVEY
U.S. CENSUS BUREAU

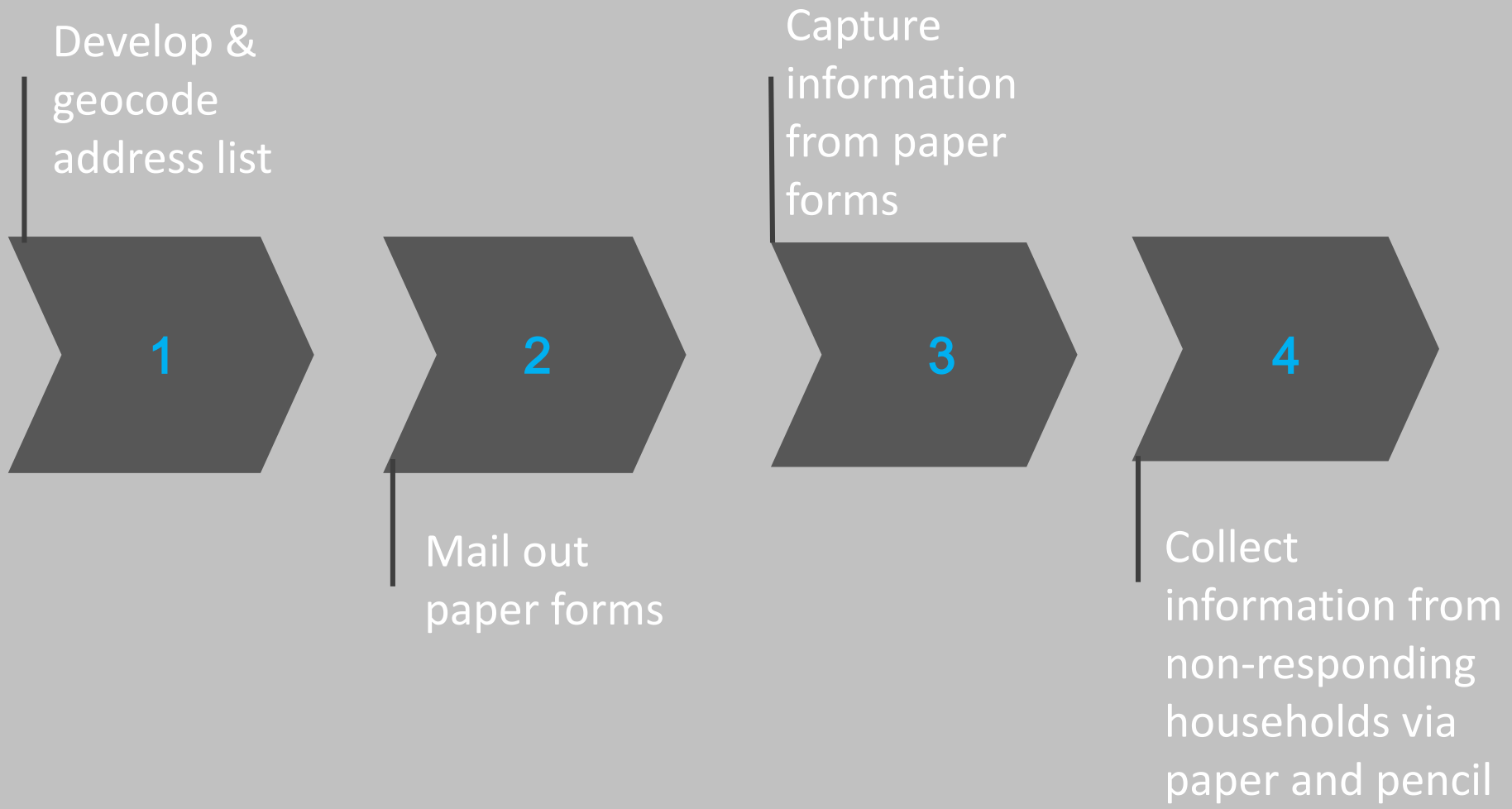
THE 2020 DECENNIAL CENSUS

Representatives and direct Taxes shall be apportioned among the several States which may be included within this union, according to their respective Numbers… The actual enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten years, in such Manner as they shall by Law direct.

U.S. Constitution, Article 1, Section 2

- Apportion representation among states
- Draw congressional and state legislative districts, school districts & voting precincts
- Enforce voting rights and civil rights legislation
- Distribute federal dollars to states
- Inform federal, tribal, state and local government planning decisions
- Inform business & nonprofit organization decisions (where to locate, size of market)
- Population benchmark for nearly every other U.S. survey

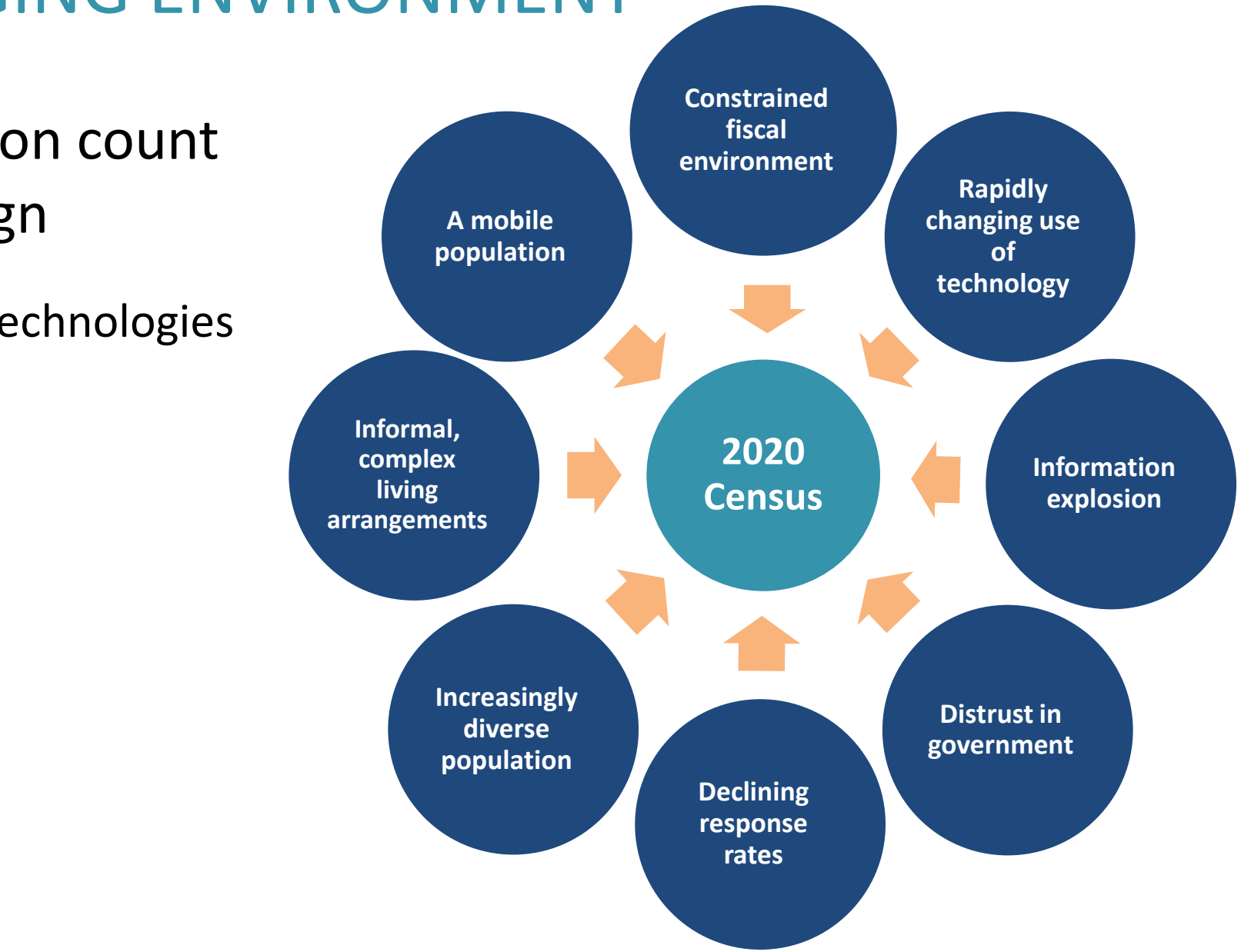
CENSUS METHODOLOGY, 1970-2010



A RAPIDLY-CHANGING ENVIRONMENT

A high-quality population count requires a flexible design

- Take advantages of new technologies and data sources
- Minimize risk



OVERARCHING GOAL:

Count every person once, only once, and in the right place

CHALLENGE GOAL:

Use modern technology to reduce the cost of the census, while maintaining the accuracy of previous censuses

- Reengineer address canvassing
- Optimize self response
- Use administrative records and third party data
- Reengineer field operations

Check out the 2020 Census Operational Plan at www.census.gov/2020census

2020 CENSUS INNOVATION

Reengineering address canvassing

Use new geospatial tools to eliminate the requirement to walk every block in the United States

- Conduct in-office review using imagery
- Procure private sector services
- Only walk 25% of blocks in the U.S.

Check out TIGER at www.census.gov/geo/maps-data/data/tiger.html

2020 CENSUS INNOVATION

Optimizing self response

Make the internet the primary response option

- Only use mail response on paper where internet is not a good option (20% of housing units)

Predict response rates and enumeration challenges

Use micro-targeted advertising

DON'T require a pre-assigned ID code to respond

Bates, Nancy and Chandra Erdman.
"The Low Response Score:
A metric to locate, predict and manage hard-to-survey populations."
Public Opinion Quarterly (2016).

✸ *Coming soon* ✸
Low Response Score app

2020 CENSUS INNOVATION

Using administrative and third party data

Remove vacant housing units from the workload for nonresponse follow-up (NRFU)

After one in-person visit, remove 6 million occupied units from the NRFU workload

Mule, Thomas, Andrew Keller, Scott Konicki, Ingrid Kjeldgaard and Darcy Steeg Morris.

"Algorithms for Including Administrative Data to Address NRFU Efforts."

Census Scientific Advisory Committee Meetings, March 2017. (*available on census.gov*)

2020 CENSUS INNOVATION

Reengineering field operations

Take advantage of technology to eliminate paper-and-pencil processes

- Smartphone for data collectors
- Tablet for supervisors

Optimize assignments based on spatial location and likelihood of contact

- Provide real-time information for supervisors
- Cloud computing

Adams, Tamara S. **"Field Reengineering for the 2020 Census."** Joint Statistical Meetings (2016).

2020 CENSUS INNOVATION

A new design for the 21st Century

Motivate People to Respond

Conduct a nation-wide communications and partnership campaign

Maximize outreach using traditional and new media

Target ads to specific audiences

Work with trusted sources to inspire participation

TELEPHONE AND PAPER SELF-RESPONSE

NONRESPONSE FOLLOWUP

INTERNET SELF-RESPONSE

Count the Population

Collect data from all households, including group and unique living arrangements

Make it easy for people to respond anytime, anywhere

Encourage people to use the new online response option

Use the most cost-effective strategy to contact and count nonrespondents

Knock on doors only when necessary

Streamline in-field census-taking

Establish Where to Count

Identify all addresses where people could live

IN-FIELD

IN-OFFICE

Conduct a 100% review and update of the nation's address list

Minimize field work with in-office updating

Use multiple data sources to identify areas with address changes

Get local government input

Release Census Results

Process and Provide Census Data

Deliver apportionment counts to the President by December 31, 2020

Release counts for redistricting by April 1, 2021

Make it easier for the public to get data

2020 CENSUS INNOVATION

COST OF REPEATING 2010 CENSUS METHODS IN 2020: $17 billion

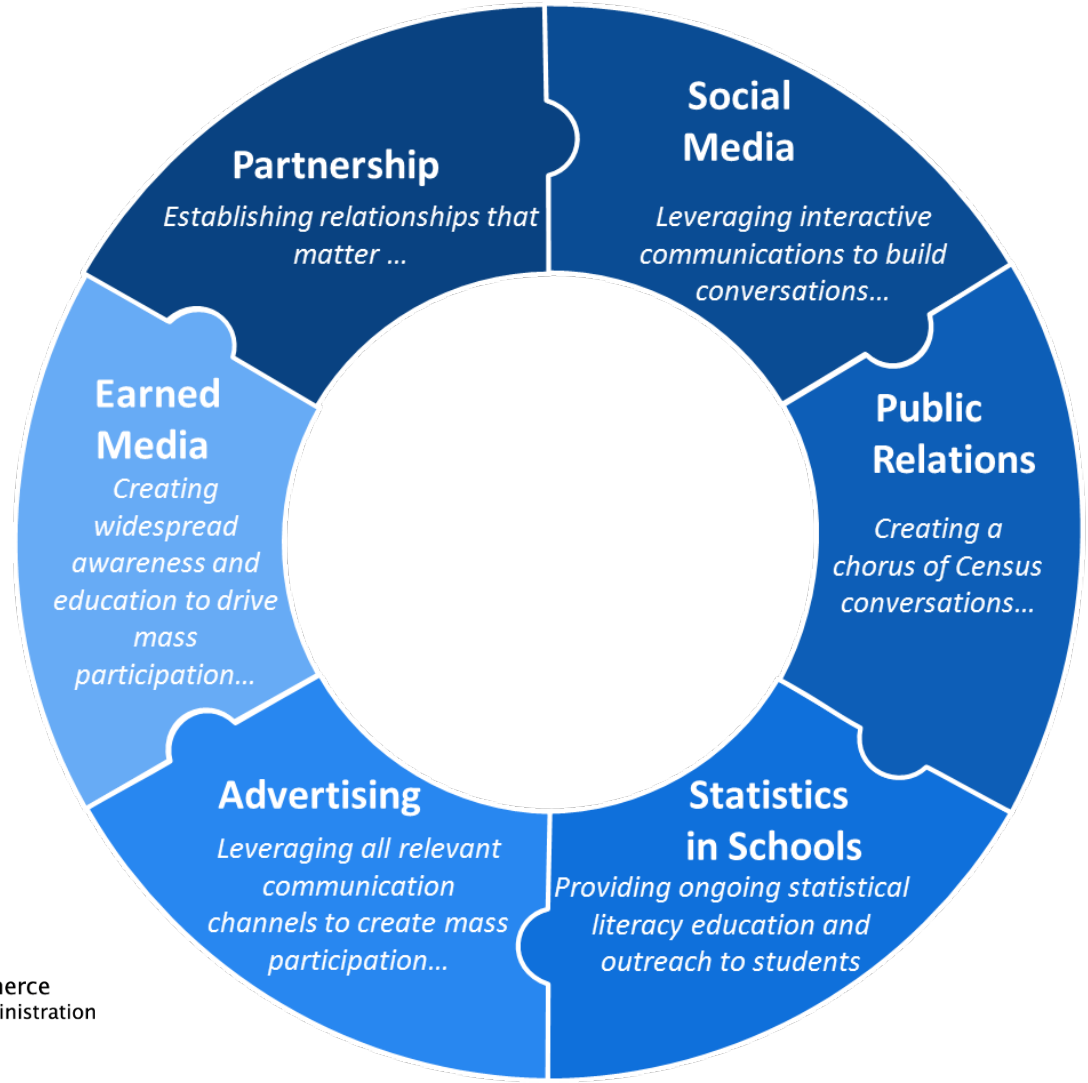
COST OF REDESIGNED 2020 CENSUS: $12 billion

POTENTIAL COST AVOIDANCE: $5 billion

* These are approximate cost estimates based on the 2020 Census Operational Plan as of September 30, 2015. Any changes to the Operational Plan will result in corresponding changes to the cost estimate. This cost estimation approach follows the GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs.

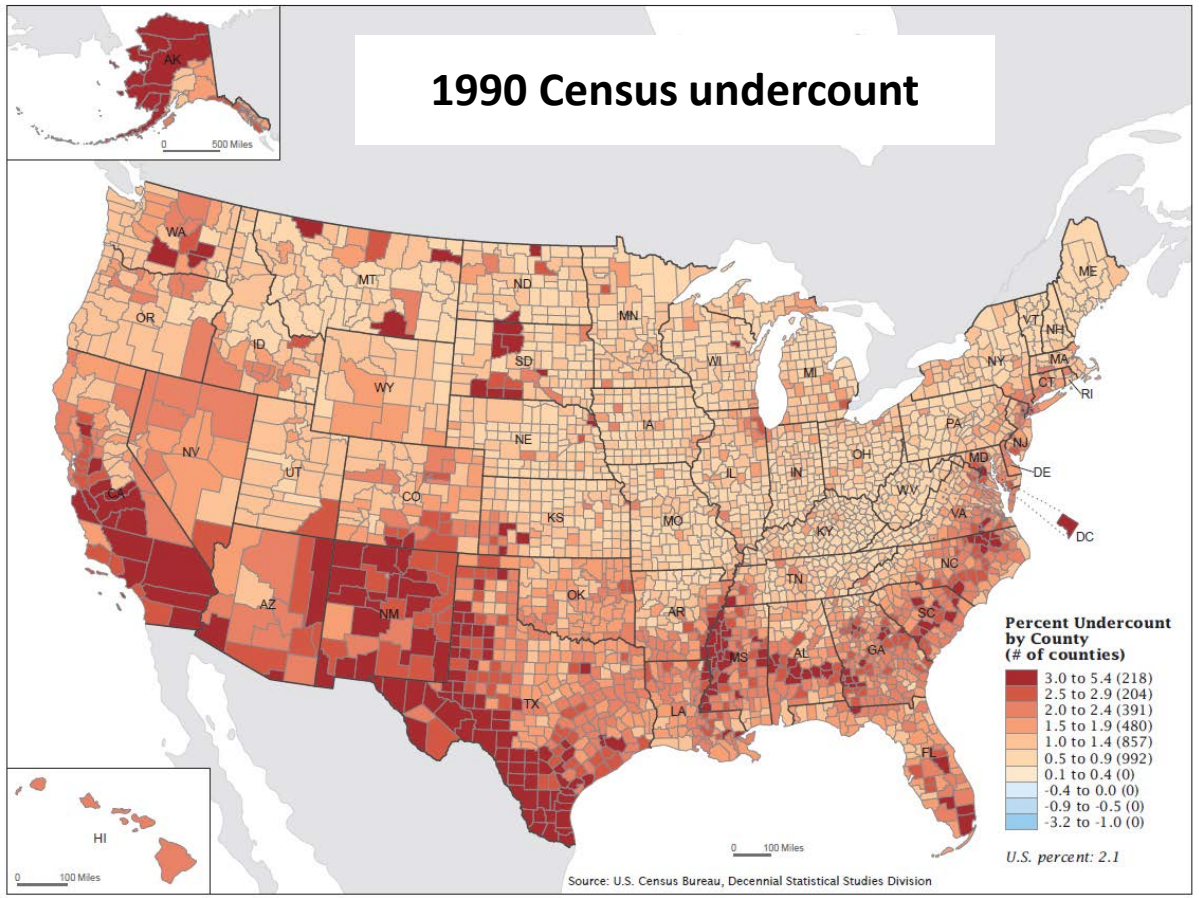
2020 CENSUS INNOVATION

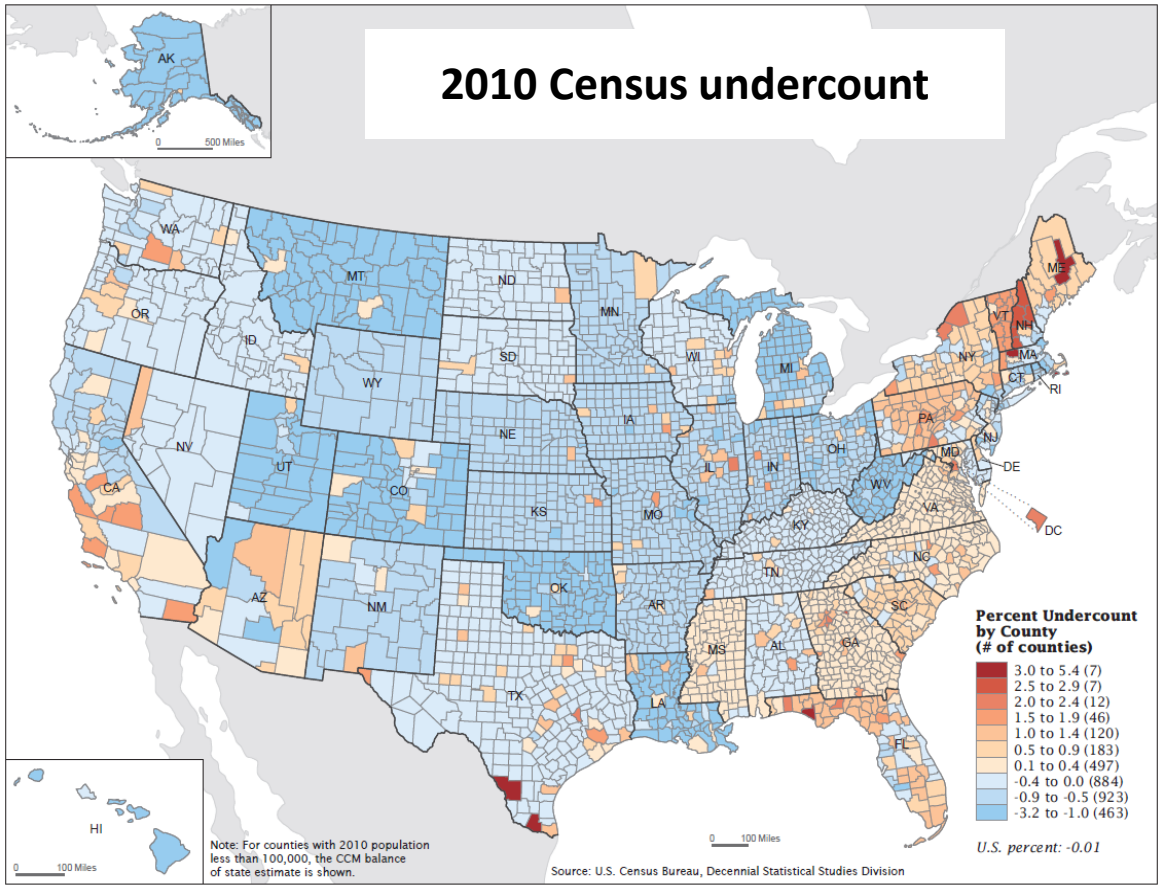
Integrated communications & partnerships program



2020 CENSUS INNOVATION

Integrated communications & partnerships program

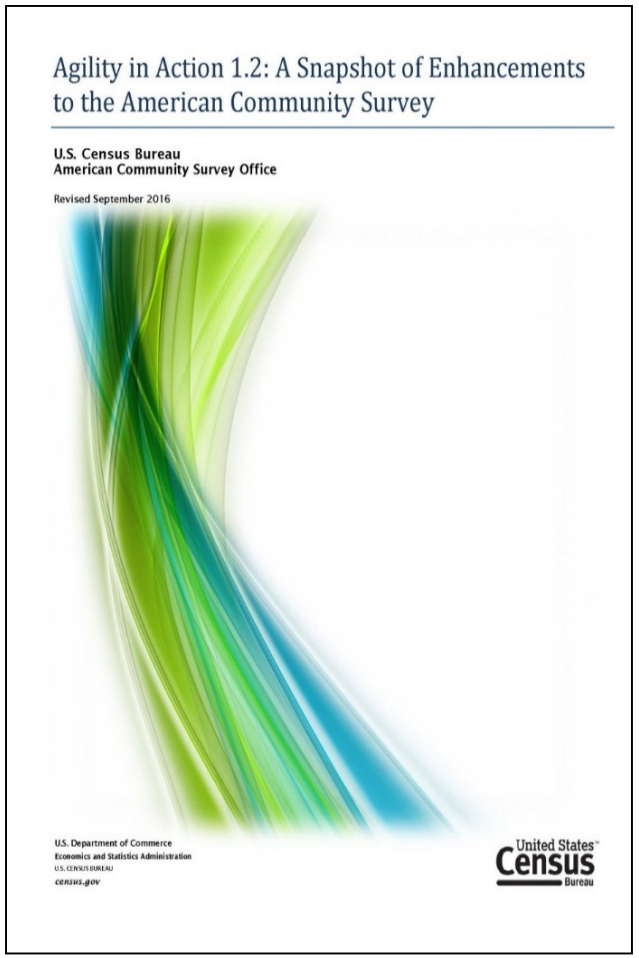




United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

REDUCING RESPONDENT BURDEN IN THE AMERICAN COMMUNITY SURVEY



The National Academies of Sciences, Engineering, and Medicine's Committee on National Statistics conducted a public workshop to shed light on ACS burdens – and opportunities for addressing them – through public discussion among stakeholders.

Key strategies for improvement:

- Reduce follow-up contacts
- Improve survey materials and the way we ask questions
- Obtain data from other sources
- Remove questions or ask questions less frequently

THE ECONOMIC CENSUS

The Economic Census is the U.S. government's five-year benchmark of the American economy.

It's conducted by the Census Bureau in years ending in 2 and 7.

- Provides a benchmark of the economy for other economic surveys and measures
- Supports decisions and planning for businesses
- Informs trade associations and chambers of commerce
- Government agencies, analysts and business organizations nationwide rely on it for planning and key economic reports

THE ECONOMIC CENSUS, 2017

100% internet response

Metadata

Questionnaires

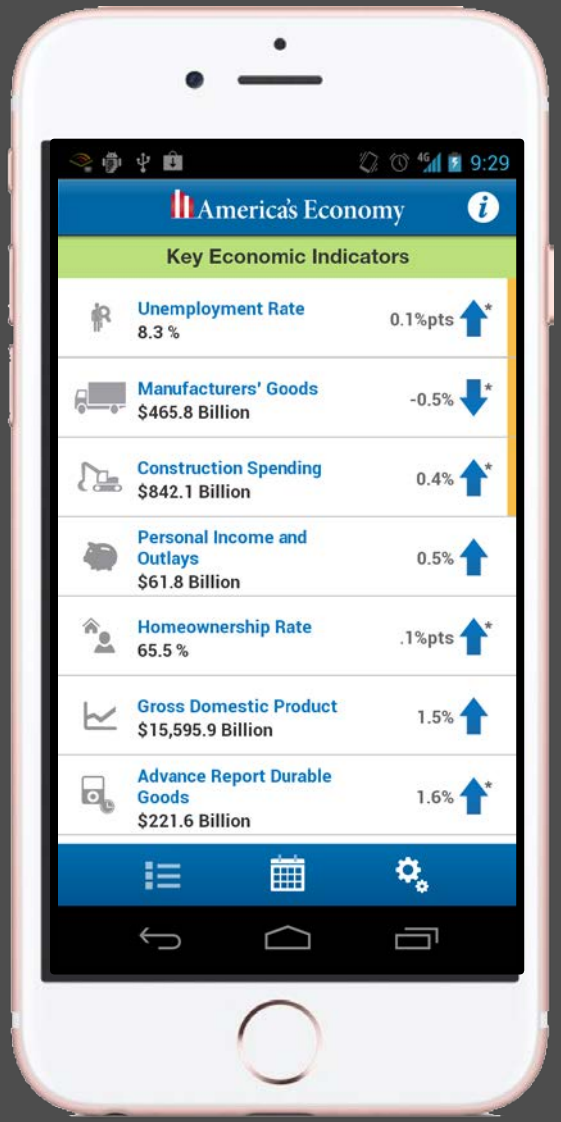
Responses

IMPROVING ECONOMIC STATISTICS

We produce:

➔ 13 key economic indicators

➔ Gross Domestic Product in conjunction with BEA



America's Economy app

2017 ECONOMIC CENSUS

The impetus for innovation

Increasing costs

Steady or declining budgets

Declining response rates

Demands for timely, granular data that's linkable with data from multiple sources

Econ Hub's guiding principles can help with this effort:

Content Harmonization ■ Data Coherence ■ Process Alignment ■ Innovative Methods

COMPONENTS OF SUCCESSFUL MODERNIZATION

Methodology

produce scientifically valid estimates and uncertainty measures of economic and social statistics from data collected from a wide variety of sources – most of which were not designed to produce inputs to the production of official statistics

Computational

develop the hardware and software infrastructure to compute and disseminate statistics constructed from a variety of sources – surveys, administrative sources, transaction data, social media, sensors, and so on.

Policy

secure legal permissions and stakeholder buy-in to utilize non-traditional sources of data for the production of official statistics, so that everyone understands the cost, benefits and risks of expanding the capabilities of this next-generation federal statistical system

Outreach & Marketing

satisfy users that the data products produced employing new data sources and techniques actually accurately measure the phenomena we intend them to /// educate users to properly draw inferences from estimates constructed in novel ways

WORKING WITH BEA TO IMPROVE GDP ESTIMATES

Advance reports provide broader range of accelerated statistics to BEA

More complete info when putting out early GDP estimates >>> fewer revisions

➔ **Advance Monthly Retail Trade Report**

➔ **Advance Economic Indicators Report**

➔ **Advance Quarterly Services Report**

IMPROVING ECONOMIC STATISTICS

Retail trade is our first focus.

➔ Changing sector requires new ways to measure and collect data

➔ Retail sector surveys are critical inputs:

- Monthly indicators for economic activity
- Personal consumption expenditures (GDP, benchmarked industry and I/O accounts, industry productivity statistics)

Improve quality of Monthly Retail indicators data

Improve timeliness of retail estimates

Improve granularity of retail estimates

Reduce burden & improve value to retail value estimates

Improve e-commerce measures

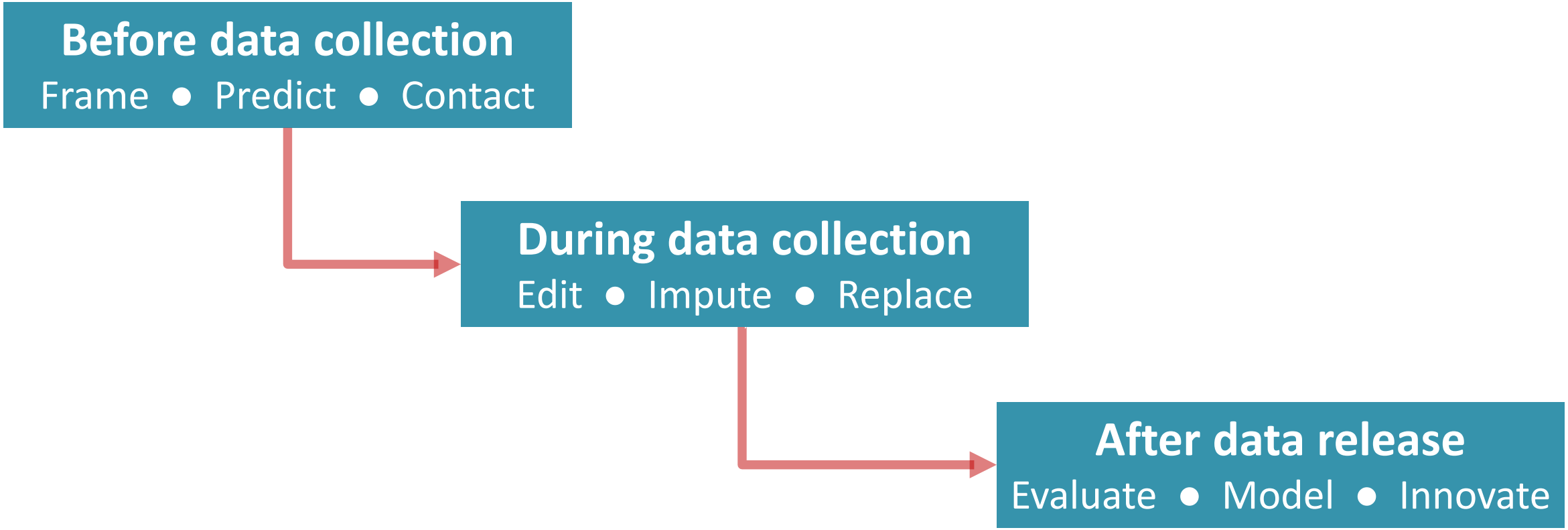
IMPROVING ECONOMIC STATISTICS

We're using non-traditional methods to meet those challenges and improve economic statistics:

- Credit card data
- Big Data
- Transaction data
- Passive data collection

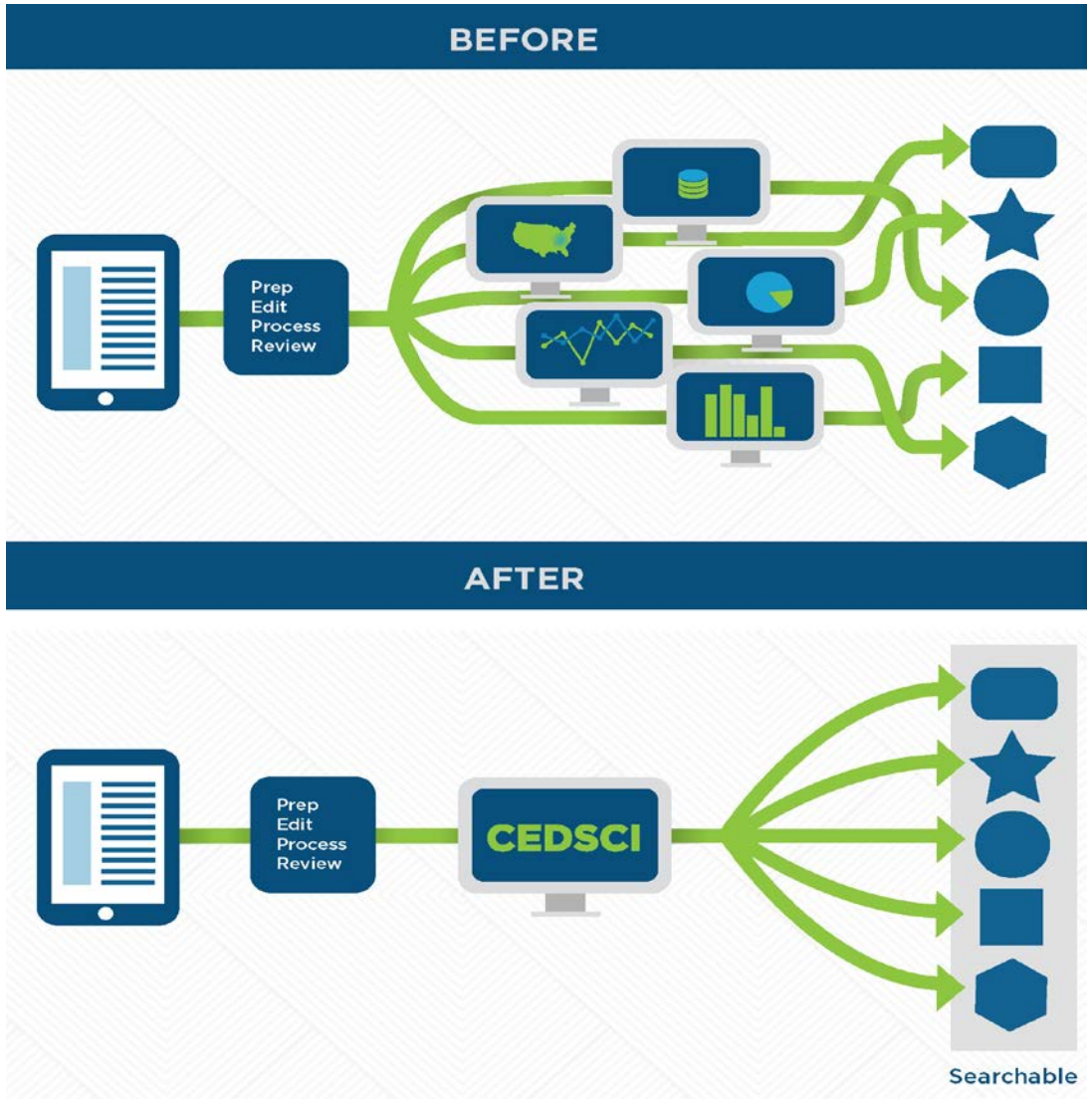
USING ADMINISTRATIVE RECORDS

Improving data collection & dissemination for the American Community Survey and demographic surveys



DISSEMINTING DATA IN NEW, EFFICIENT WAYS

Making Our Data Available for the 21st Century



Center for Enterprise Dissemination Services & Consumer Innovation

Suite of services and supporting infrastructure to handle data dissemination for the 130+ censuses and surveys within the Census Bureau

- Improved customer satisfaction
- Personalized customer experience
- Efficient operations

Making Our Data Available for the 21st Century

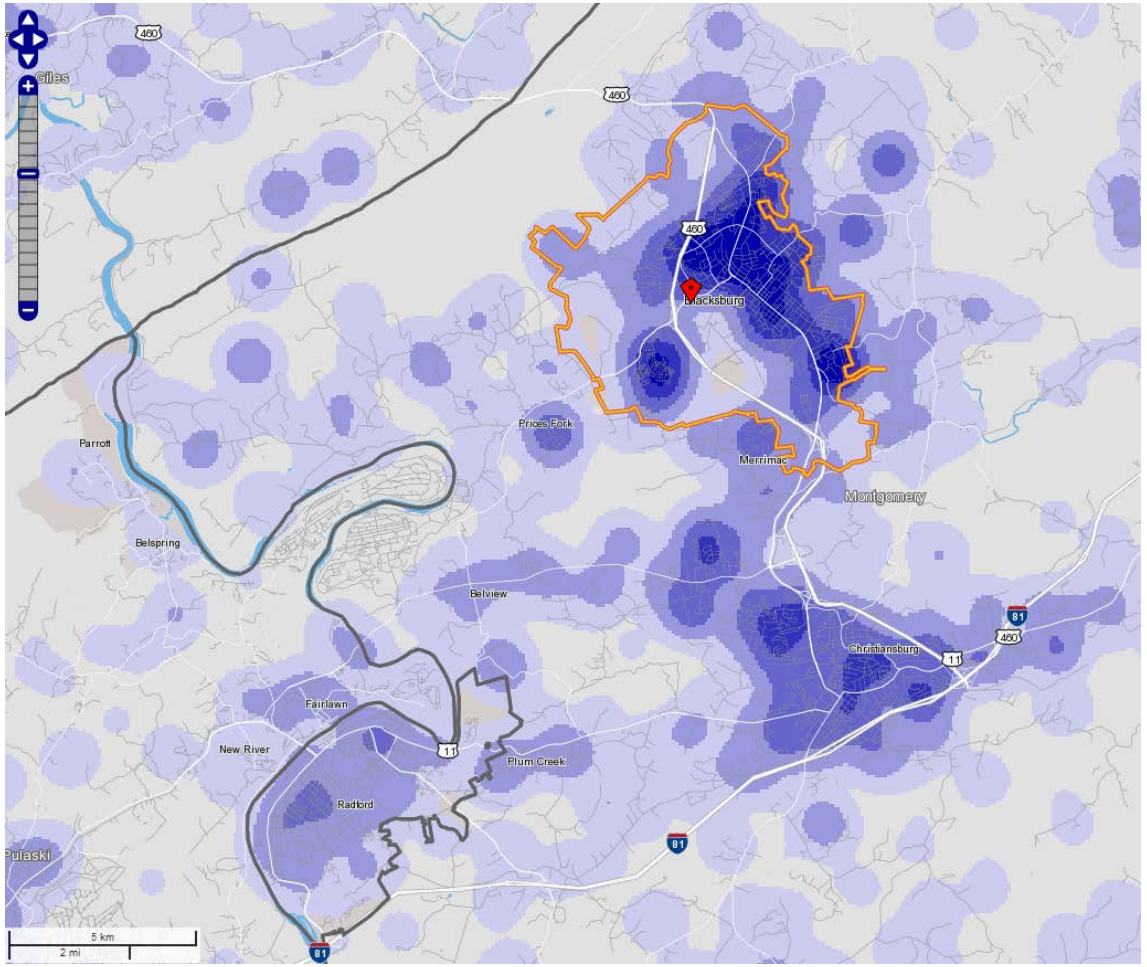
LONGITUDINAL EMPLOYER-HOUSEHOLD DYNAMICS

Machanavaijjhala, Ashwin and Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber.

"Privacy: Theory Meets Practice on the Map."

International Conference on Data Engineering (2008).

Heat map of residences for jobs in Blacksburg, VA



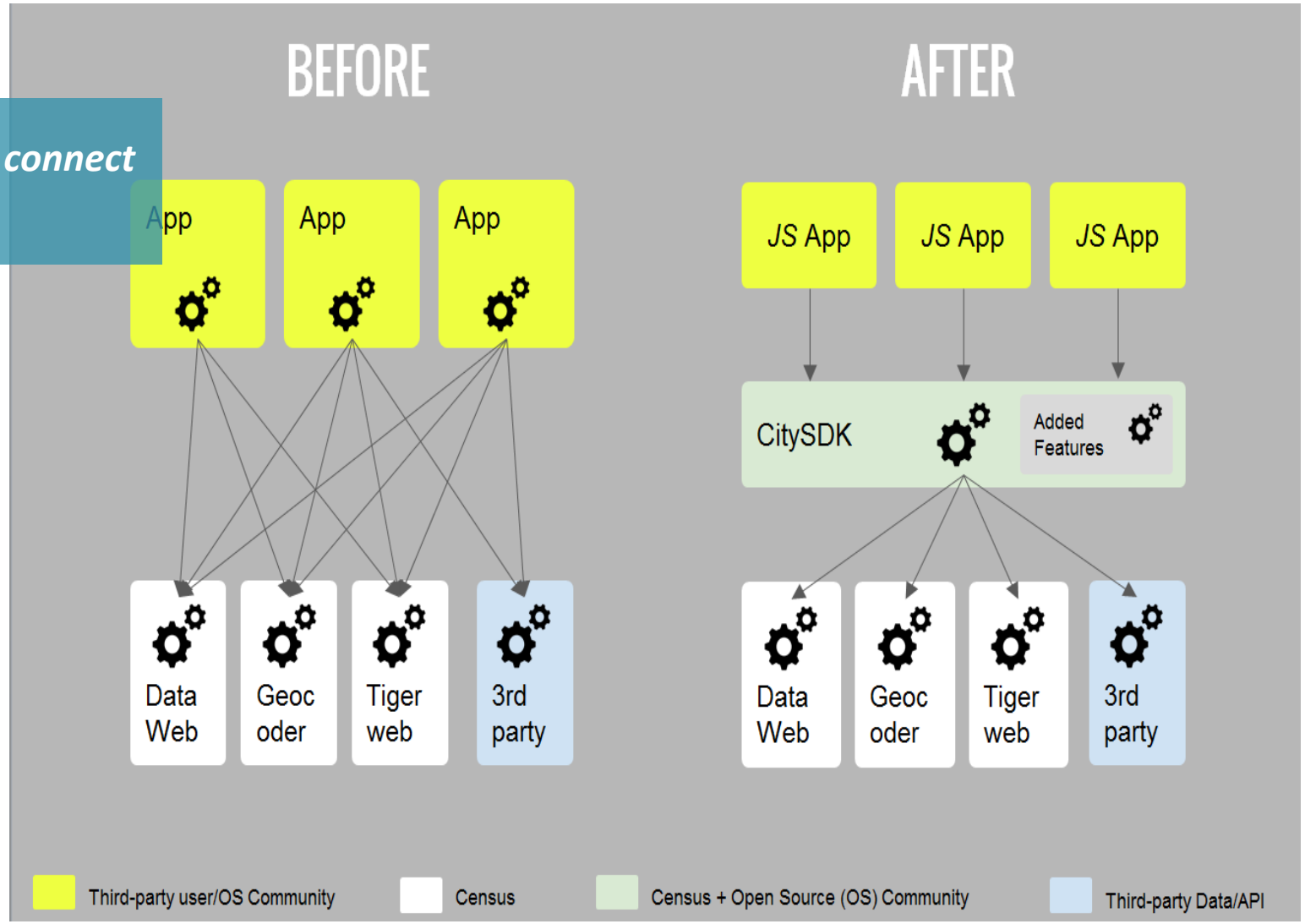
Produced using OnTheMap web tool, based on LEHD Origin-Destination Employment Statistics

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Making Our Data Available for the 21st Century

CITY SDK

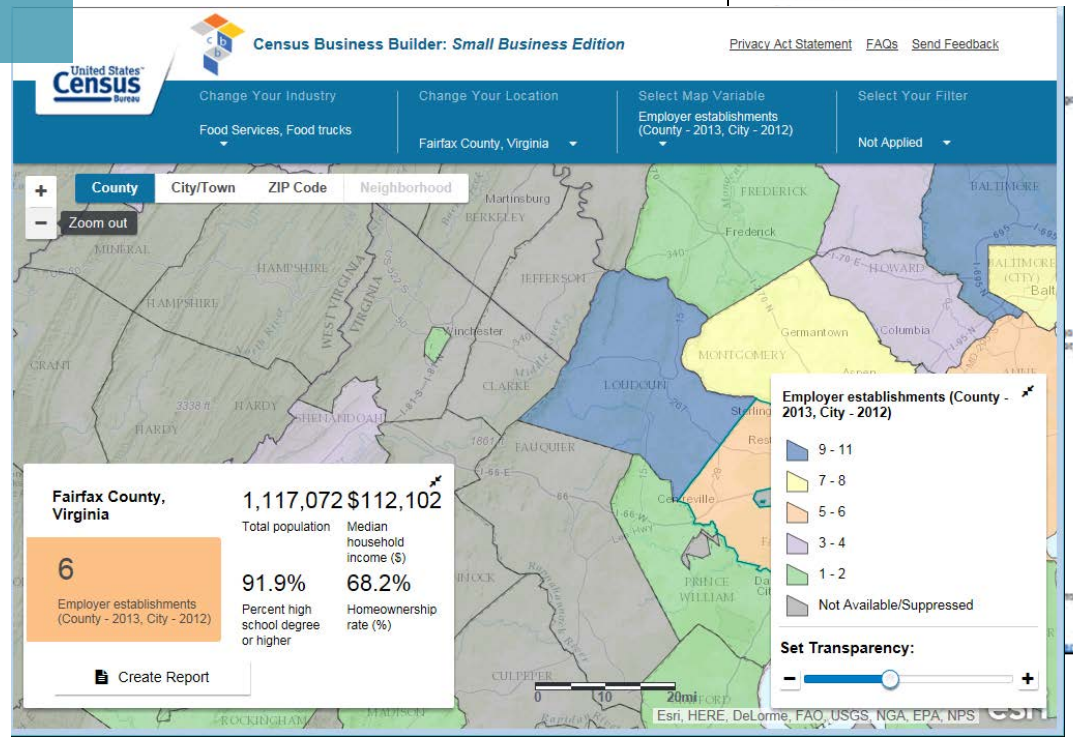
Toolbox for civic innovators to connect local and national public data

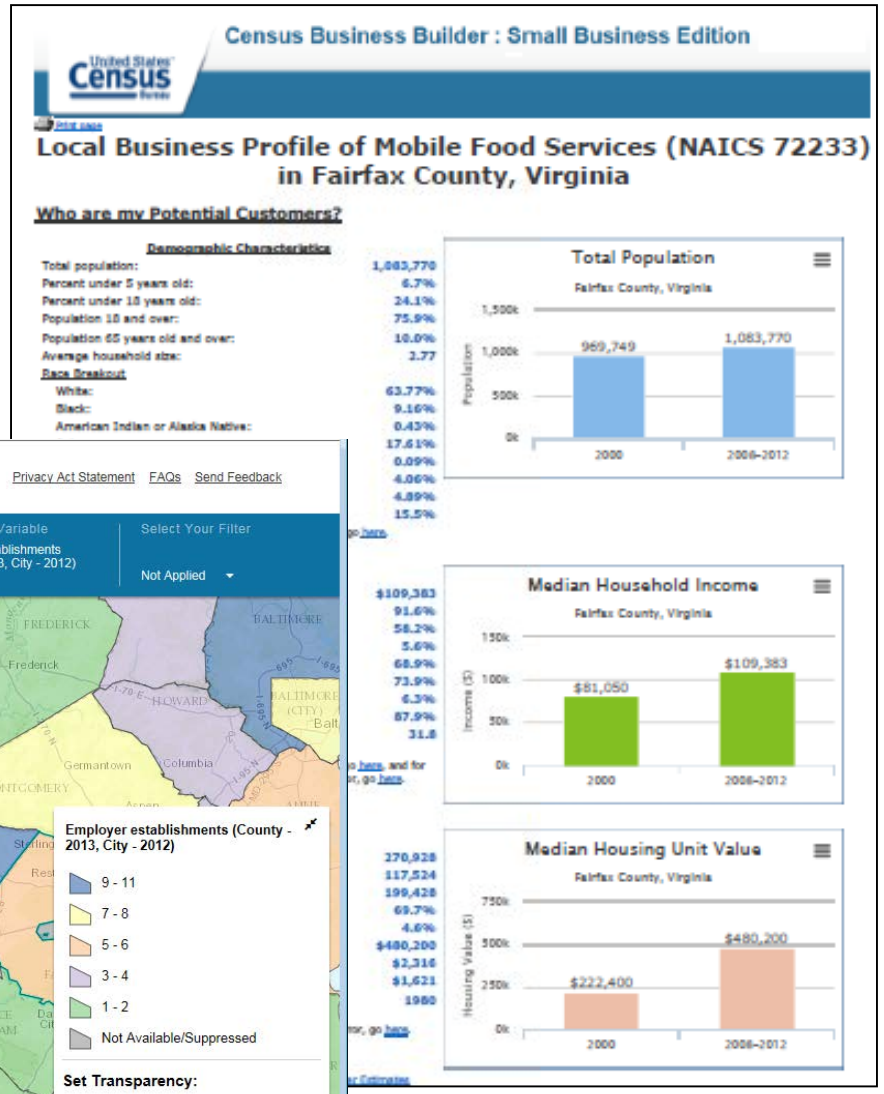


Making Our Data Available for the 21st Century

CENSUS BUSINESS BUILDER

Two editions **– Small Business** ***and*** **Regional Analyst**





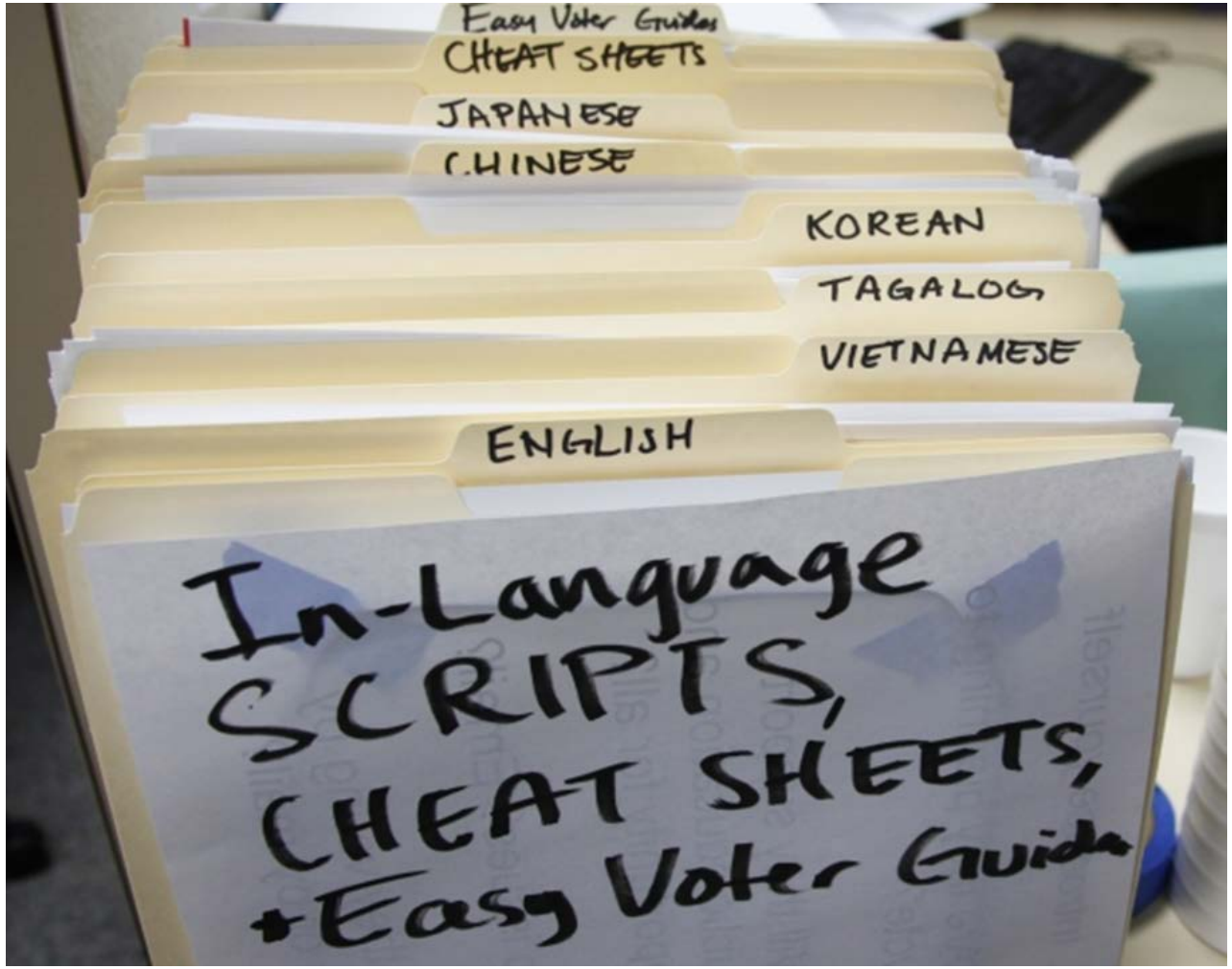
Making Our Data Available for the 21st Century

VOTING RIGHTS ACT SECTION 203

Joyce, Patrick M., Donald Malec, Roderick J.A. Little, Aaron Gilary, Alfredo Navarro and Mark E. Asiala.

"Statistical Modeling Methodology for the Voting Rights Act Section 203 Language Assistance Determinations."

Journal of the American Statistical Association (2014).



Making Our Data Available for the 21st Century

OPPORTUNITY PROJECT

CENSUS BUREAU – DEPARTMENT OF HOUSING & URBAN DEVELOPMENT – WHITE HOUSE – PRIVATE SECTOR – CITY GOVERNMENTS

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

United States Census Bureau | THE WHITE HOUSE WASHINGTON | U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

FORD FOUNDATION | Socrata The Data Experience Company | GREAT!SCHOOLS | Zillow | MEASURE OF AMERICA of the Social Science Research Council | COMMUNITY COMMONS together for the common good | Lifting Up What Works | Streetwyze | PolicyLink | REDFIN | esri | URBAN INSTITUTE | diversitydatakids.org data for a diverse and equitable future | azavea | National Equity Atlas

MODERNIZING HOW WE COLLECT AND PRODUCE DATA

Interagency Council on Statistical Policy strategic priorities

- **Quality:** Develop standards and methods for combined statistical data
- **Research access:** Continue to grow and improve Federal Statistical Research Data Centers
- **Public access:** Increase & improve access and creating value-added products
- **Human capital:** Invest in employees and competencies
- **State and federal program data:** Improve intergovernmental and interagency relationships, in order to acquire & use data
- **Respondent burden:** Develop strategies to counter falling response rates

The Census Bureau works with other federal statistical agencies and OMB on the ***Federal Commission on Statistical Methodology*** to develop recommendations on modernization

Thank you!