

Big Data at Census

Cavan Capps

Big Data Lead

U.S. Census Bureau



U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
[census.gov](https://www.census.gov)

Census Mission and Purpose

The **Census Bureau's mission** is to serve as the leading source of quality data about the nation's people and economy. We honor privacy, protect confidentiality, share our expertise globally, and conduct our work openly.



Census developed national surveys in 1942

Opportunities to improve available data via Big Data technologies

- During the Great Recession
 - No Supply-Chain data
 - Job loss could be linked to Supply-Chain business failure
 - No Housing Fore-closure data
 - Could improve low level geography data for
 - Retail Sales
 - Housing Starts
 - Faster Local Gov't Finance data

Big Data and Related Technologies

- Machine learning (supervised, unsupervised)
- Recommendation systems/ Propensity modeling
- Text Mining
- Deep learning (neural networks)
- Web crawling
- Streaming real time data
- Data Visualizations for Data Review etc
- Graphs

Opportunities for Census



- Improve efficiencies for older survey processing
- Opportunities to get new types of data cheaper and faster

Improving current Census Processes

- Adaptive Design : Propensity modeling
- Industry/Occupation/Product classification (ML)
- Data Imputation & edits (ML classification)
- Data Review (ML classification)
- Data Linkage (cluster processing)
- Data Lake (data enclave)
- New Formal Privacy Preserving techniques
- Secure multi-party computing
- Real time electronic transactions

Potential “big data” measures of economic activity

- Consumer behavior
 - Retail Trade by MSA
 - By demographic group
 - By income groups
 - Consumer confidence
- Housing activity
 - building permits
 - starts
 - sales
 - value
 - foreclosures
- Local Economic Development data
 - Local incomes
 - Demographic data
- Transportation data
 - Supply Chains
 - By actual roads
 - By season
 - On/off rush hour
 - By demographic
 - Different trends by age for transportation
 - Walking, biking, metro, driving
 - Seasonal differences..

Privacy becomes more critical



No longer small samples
So much data to match to...

- New Techniques: Formal Privacy & Differential Privacy
 - Apple and Google using Differential Privacy to collect data
- Secure Multi-party Computing

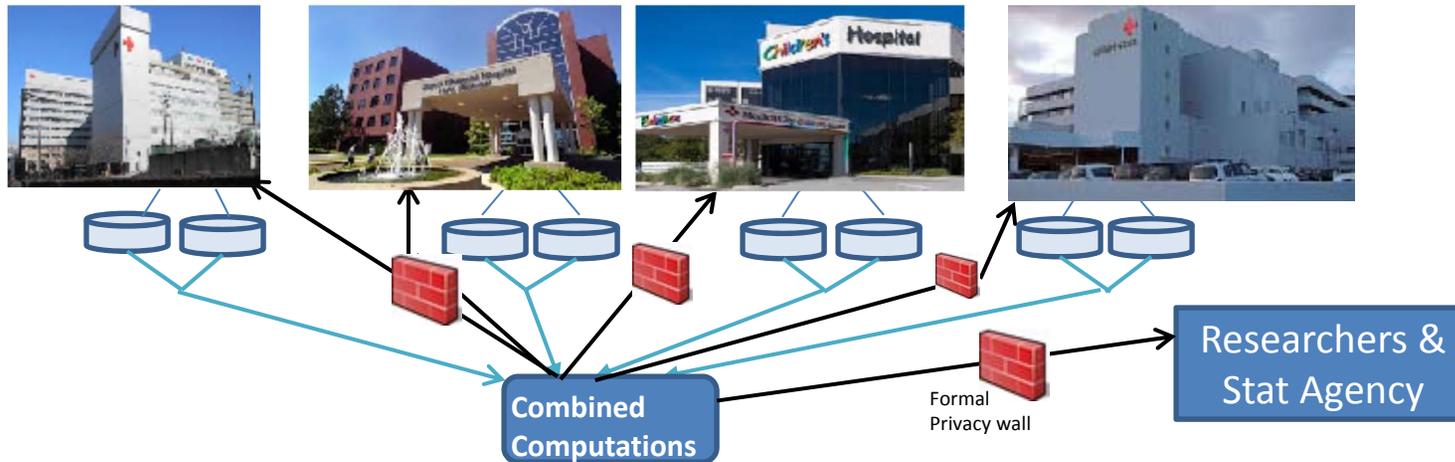
Collecting new Data

- Web scraping public data
- Electronic Transactions
 - Businesses
 - People



- Perhaps using Differential Privacy like Apple or Google
 - <https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/>
 - <http://research.google.com/pubs/pub42852.html>
- Or using Secure Multi-party computing for business near real time data

Secure Multi-party Computing



Statisticians or researchers might be able to do analysis across transactions or records located across a set of different businesses or health providers while keeping data strictly private.

This would by pass the need to move all records to a complex central constantly updated database and simplify research/privacy issues.

Current or Potential Projects

- Getting data more effectively
 - Response Propensity Modeling
 - Text mining to improve survey response rates
 - Aggregate Electronic Retail Transactions
 - Web Scraping for Governments, Housing Data, International data
 - Pilot for “Secure Multi-party computing”
 - Machine Learning to estimate foreign populations from satellite imagery
- Processing and Cleaning Data
 - Machine Learning for Imputations for Retail Trade
 - Machine Learning code classifications for NAICS and Product Codes

Current or Potential Projects

- Improved Research Capabilities
 - Parallel processing to speed Record Linkage
 - Administrative Data Enclave “data lake”
 - Improving Medicare operational effectiveness
 - Measuring and tracking the "gig" economy
- Improve Data Dissemination Capabilities
 - Privacy Protection Innovation
 - Improving searching using Elastic Search for data Dissemination

Questions

Cavan.Paul.Capps@Census.gov



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