Census Bureau's Data Linkage Infrastructure for program evaluation

October 25, 2022

Federal Committee on Statistical Methodology

Carla Medalia, U.S. Census Bureau

carla.medalia@census.gov

DGWG report: leverage existing survey and administrative data

- Agencies should leverage existing survey and administrative data whenever possible when developing and executing program evaluation
    - Particularly useful to enable large scale observational studies
    - Enables consistent program evaluation across agencies and facilitates comparisons across variety of programs
    - Reduces burden of aid awardees when reporting on the work they did with the federal funds
    - Allows agencies to answer questions otherwise not possible to answer

Census Bureau survey and census data

People and households

Businesses,

Address level

Business/organization level

Person level

Administrative data

Federal

State/local

Commercial

data

governance

U.S. Code Title 13

Secure computing environments

Support high-quality research and evaluation, advancing the Census Bureau's mission of providing timely and unbiased data to support evidence-based decision making

United States® Census Bureau

Census data: people and households

Sources

- Decennial censuses
- American Community Survey
- Current Population Survey
- Survey of Income and Program Participation
- American Housing Survey

Uses

- Sociodemographic characteristics
- Income, poverty, health insurance
- Labor force, occupation, industries
- Data linked over time
- Individuals, families, households
- Geographic information down to Census blocks

Census data: businesses, governments, and economy

Sources

- Economic Censuses
- Census of Governments
- Firm Surveys
- Establishment Surveys
- Transaction or Trade data
- Longitudinal Employer-Household Dynamics (LEHD)

Uses

- Microdata about all U.S. businesses: precise geolocations
- Microdata about large samples of U.S. business: geolocations, payroll, tax records, foreign investments
- Detailed geographies and industries
- Data linked over time
- Employee and employer linked data

Administrative data

Federal data
Bureau of Labor Statistics
Bureau of Prisons
Corporation for National and Community Service
Department of Defense
Department of Veteran's Affairs
Health and Human Services
Housing and Urban Development
Indian Health Service
Internal Revenue Service
Office of Personnel Management
Selective Service System
Small Business Administration
Social Security Administration
U.S. Postal Service

State/local data
Unemployment Insurance
Supplemental Nutrition Assistance Program
Women, Infants, and Children
Temporary Assistance for Needy Families
Low Income Energy Assistance Program
Child Care Subsidy
Homeless Management Information System
Alaska Permanent Fund
Puerto Rico tax data
California tax data
University education data
Criminal Justice Administrative Records System

Third party/commercial data
Mortgage/address data
Contact frame
Property and tax foreclosure

Note: data access subject to approval

Data linkage quality depends on PII/BII

Linkage type	Linkage method	Linkage fields to collect
Address linkage	Master Address File Match (MAFMATCH)	Full address; coordinates for point of service delivery
Organization linkage	TF-IDF (term frequency-inverse document frequency), MAMBA	Businesses/non-profits: For both establishment and firm collect: EIN, business name, mailing address, physical location address, NAICS, company web address, UEI, SSN (owner of sole proprietorship) Governments: name, address, UEI, web address
Person linkage	Person Identification Validation System (PVS)	Full name (first, middle, last, suffix), complete date of birth (age is acceptable but less optimal), full address, sex, SSN/ITIN (for administrative records with authority to collect)

Legal framework and data governance

- Legal Framework: U.S.C. Title 13
    - **§ 6.** Acquire and utilize records to the greatest extent possible
    - **§ 8.** Reimbursable studies and joint statistical projects
    - **§ 9.** Protect confidential individual and establishment data, limit access, and statistical uses
    - **§ 23(c).** Swear in researchers to assist the Census Bureau
    - **§ 214.** Wrongful disclosure of information

- Data governance infrastructure
    - Data stewardship, disclosure review
    - Anonymized data
    - Secure computing environments

United States® Census Bureau

Census Bureau's Data Linkage Infrastructure enables measurement of program impacts

- Direct program impacts: defined in legislation

- Indirect program impacts: second-order effects

- Example: NTIA's Broadband Equity, Access, and Deployment (BEAD) Program
  - Direct: broadband built
  - Indirect: access to internet; educational attainment; commuting patterns

Prototype library of indirect program impact metrics

Category	Measure	Data source name
Demographic	Population characteristics	American Community Survey, decennial census
Demographic/ business	Business and owner characteristics	Annual Business Survey, Nonemployer Statistics by Demographics
Economic indicators	New business starts	Business Formation Statistics and Business Dynamics Statistics
Economic indicators	Manufacturing	Manufacturers' Shipments, Inventories, and Orders (M3)
Environment/ Climate	Coastal communities and businesses	American Community Survey; Annual Business Survey; Business Register
Jobs	Labor force participation rate	Current Population Survey, American Community Survey
Other economic	Tax revenues	Census of Governments
Socioeconomic	Poverty Rate	CPS ASEC, ACS, administrative data
Underserved communities index	At-risk neighborhoods	Community Resilience Estimates