

Restricted Data in the FSRDCs – New Developments

Melissa Ruby Banzhaf, PhD
Administrator, Atlanta FSRDC
U.S. Census Bureau

October 4, 2019

Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

What are Federal Statistical Research Data Centers (RDCs)?

- Secure computing labs where qualified researchers conduct approved statistical analysis on non-public data.
- These data are collected by various government agencies (Census Bureau, NCHS, BEA, BLS, SSA, etc.).
- Established through an agreement between federal statistical agencies and a local research community.
- Managed by the Census Bureau.

The Atlanta Research Data Center

- Located in the Federal Reserve Bank of Atlanta
 - corner of 10th & Peachtree
- Consortium Members
 - Emory University
 - University of Georgia
 - Georgia State University
 - University of Alabama
 - Georgia Institute of Technology
 - Federal Reserve Bank of Atlanta
 - University of Tennessee – Knoxville
 - Florida State University
 - Tulane University
 - Vanderbilt University



FSRDC Locations

Atlanta

Boston

California - Berkeley

California - Irvine

California - Stanford

California - UCLA

California - USC

Census HQ (Suitland, MD)

Central Plains (Lincoln, NE)

Colorado (Boulder)

Dallas-Forth Worth

Georgetown University

Illinois (Chicago, UIUC)

Kansas City

Kentucky (Lexington)

Maryland (College Park)

Michigan (Ann Arbor)

Minnesota (Minneapolis)

Missouri (Columbia)

New York (Baruch, Cornell)



Philadelphia

Penn State

Seattle

Texas (Austin, College Station)

Triangle (Duke)

Wisconsin (Madison)

Yale

Restricted Data Available from Multiple Agencies

- Census
- Bureau of Economic Analysis (BEA)
- National Center for Health Statistics (NCHS)
- Agency for Healthcare Research and Quality (AHRQ)
- Bureau of Labor Statistics (BLS)
- National Center for Science and Engineering Statistics (NCSES)
- In Future (?):
 - Statistics of Income (IRS)
 - Bureau of Transportation Statistics

Types of Restricted Census Data Available

- Economic Data
 - Microdata on firms and establishments
- Demographic Data
 - Microdata on individuals and households
- Employer-Employee Linked Data (LEHD)
 - Data on employees linked with data on employers
- University Grant Data (UMETRICS)
 - Data on university grant awardees linked with data on employees and vendor firms

Most Popular Economic Microdata

Data Set	Unit of Enumeration
Economic Censuses (Manufactures, Retail Trade, Services, etc.)	Establishment
Annual Surveys (Manufactures, Services, etc.)	Establishment/Firm
Standard Statistical Establishment List (SSEL)/Business Register	Establishment
Compustat - SSEL Bridge	Firm
Longitudinal Business Database	Establishment
Longitudinal Firm Trade Transactions Database (LFTTD)	Firm
Survey of Business Owners/Annual Survey of Businesses	Firm
Many more economic datasets!	Establishment/Firm

Economic Example

"Firms' Internal Networks and Local Economic Shocks" Giroud and Mueller
American Economic Review (2019)

- The authors examines how local shocks spread across US regions through firms' internal networks of establishments.
- They construct a spatial network of an entire firm's establishments using the **LBD** and categorize establishments using zip codes and 4-digit NAICS codes. Shocks to consumer demand are measured using zip code level changes in the housing prices from 2006-2009.
- They find that establishment-level employment is sensitive to shocks in distant regions in which the parent firm is operating, and that aggregate county level employment is sensitive to shocks in distant counties linked by firms' internal networks.

Demographic Survey Datasets

- Decennial Surveys 1940-2010
- American Community Survey (annual microdata, 1996-2015)
- Current Population Survey (selected supplements)
- Survey of Income and Program Participation
- National Crime Victimization Survey
- National Survey of College Graduates
- National Longitudinal Study of Mortality
- American Housing Survey/ Rental Housing Finance Survey (HUD)

Linking - the Power of the PIK

- PIK – a unique Census identifier assigned to individuals using name, birthdate, address information, and social security number.
- Many restricted Census demographic datasets are PIKed
- PIK crosswalks permit linking restricted demographic datasets together:
 - 1940 Decennial Census to 2000 Decennial Census
 - 2000 Decennial Census to 2010 Decennial Census
 - American Community Survey to Decennial Census

Linking to the Census Numident

- Information from Social Security Number registrations
 - Place of birth (Census place/county)
 - Date of Birth
 - Date of Death (if deceased)
- Includes PIK
- Includes Census Household Composition Key (formerly known as Census Kidlink) links kids' PIKs to PIKs of parents (from SSN registrations)

Linking to the Master Address File (MAF)

- Master Address File (MAF)
 - Created for 2000 Census
 - Complete list of addresses with lat/long and geography variables
- MAF Extract
 - Yearly snapshot of MAF can be used for research
- MAF Auxiliary Reference File (ARF)
 - Links address key (MAFID) to individuals' PIKs

Demographic Example

“The Long-Run Effects of Recessions on Education and Income” Stuart CES Working Paper (2017)

- The author obtains adult education and income outcomes from the 2000 **Decennial** and the **ACS**. He links those adults to their county of birth found in the Numident File.
- He examines the effect of the 1980-82 recession on the education and income outcomes of individuals who were children, adolescents, and young adults during the recession. He finds that for children age 0-10 in 1979, a 10% decrease in earnings per capita in their county reduces college graduation by 10% and adult income by 3%. He does not find significant effects for adolescents or young adults.

Linked Data - LEHD

- Links individuals to place of employment
- Based on unemployment insurance administrative records
- Contains quarterly earnings for every company employee
- “Tracks” a person based on their place of employment
- Contains some demographic information on individuals
- Can link LEHD to other Census business data via firm identifier
- Can link LEHD to other Census household data via individual identifier
- Available on a state-by-state basis

LEHD Example

“Employee Costs of Corporate Bankruptcy” Graham et al., NBER Working Paper (2019)

- This paper quantifies the earnings losses to employees induced by bankruptcy filings. The researchers find wages decline by 10% in the year of the bankruptcy filing and decline by a cumulative present value of 67% over 7 years.
- Authors use an external database on bankruptcy filings and merge it to the **Business Register**. They then link these firms to their respective workers in the **LEHD** and examine how their earnings change over time.

Linked Data - UMETRICS

- **Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science**
- Database containing:
 - Information on university grants (32 universities currently) – (Awards files)
 - People paid on the grants (Personnel File)
 - Businesses paid by the grants (Vendor File)
- Can link grant recipients to internal Census Bureau data on individuals and firms

UMETRICS Example

“STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census” Buffington et al. *AER: Papers and Proceedings* (2016)

- The authors examine gender differences in graduate training and early career of men and women in STEM-related fields, possibly as an explanation as to why women are underrepresented in STEM fields, especially at later career stages.

UMETRICS Example, Cont.

- Authors use the following datasets:
 - UMETRICS Personnel file linked to ProQuest Dissertation database (identify graduate students and field of study)
 - UMETRICS linked to Decennial Census (gender, marital status of student, children)
 - UMETRICS linked to W-2 earnings
 - UMETRICS linked to LEHD-ECF, LBD, ILBD (sector of employment)
 - UMETRICS grant information (identify research team)

UMETRICS Example, Cont.

- The authors find:
 - Gender “separation” among students – women work on teams with larger shares of women than men.
 - Gender differences in field of study:
 - Biology, chemistry, health – 59% of females, 27% of males
 - Engineering – 21% of females, 45% of males
 - Women earn 31% less than men overall; 11% less controlling for field of study and funding source.
 - Gap disappears once gender interacted with marital status and children.

Bureau of Economic Analysis Datasets

- Allow researchers to analyze economic behavior of
 - Multinational Enterprises (MNEs)
 - Firms that trade in services
- Can be linked to Census establishment level data

Bureau of Economic Analysis Datasets

Currently Available	Expected in 2020
U.S. MNEs worldwide (BE-10/11)	Trade in insurance transactions by U.S. insurance companies (BE-45/140)
Foreign MNEs in the U.S. (BE-12/15)	New foreign direct investment in the U.S. (BE-13)
Trade in selected services and intellectual property (BE-120/125)	U.S. direct investment abroad (BE-577)
Trade in financial services (BE-180/BE-185)	Foreign direct investment in the U.S. (BE-605)

Health Data

- These data are collected by:
 - National Center for Health Statistics (NCHS) -
 - Recently includes the National Survey on Drug Use and Health (NSDUH)
 - Agency for Healthcare Research and Quality (AHRQ)

What types of NCHS data?

National Health Status Surveys

- National Health and Nutrition Examination Survey (NHANES) I, II, and III
- National Health Interview Survey (NHIS)
- Longitudinal Study on Aging I and II (LSOA)
- National Survey of Family Growth
- National Survey of Children's Health
- National Survey of Early Childhood Health
- National Survey of Children with Special Health Care Needs
- National Asthma Survey

National Health Care Surveys

- National Ambulatory Medical Care Survey

- National Hospital Ambulatory Medical Care Survey
- National Survey of Ambulatory Surgery
- National Hospital Discharge Survey
- National Nursing Home Survey (NNHS)
- National Home and Hospice Care Survey
- National Employer Health Insurance Survey
- National Health Provider Inventory
- National Immunization Survey

Vital Statistics

- Mortality and Multiple Mortality
- Birth
- Fetal Death
- National Death Index
- Marriage and Divorce

What types of AHRQ Data?

- Medical Expenditure Panel Survey (MEPS) files include:
 - Household Component
 - Provider Component
 - Insurance/Employer Component
 - Nursing Home Component (1996 only)
 - Area Resource File
 - Two-year two panel file
 - MEPS-NHIS linked data
- Only Household Component and portions of Provider Component are publicly available

Bureau of Labor Statistics Datasets

- Currently available in the FSRDCs:
 - National Longitudinal Surveys (individual)
 - Survey of Occupational Injuries and Illnesses (establishment)
 - Census of Fatal Occupational Injuries (CFOI) (individual)
- More coming soon:
 - National Compensation Survey (establishment)

National Center for Science and Engineering Statistics

- Survey of Earned Doctorates
- Survey of Doctoral Recipients
- Data will be PIKed so can links with other Census data

How to Access the FSRDC

- Develop proposal
 - Different guidelines for the various agencies
 - Submit proposal for agency review
 - Review time varies by agency
- Obtain Special Sworn Status (SSS)
 - Residency Requirement (3 years)
 - Only U.S. citizens can access BEA datasets
- Pay fees for NCHS, BEA, BLS data (may soon include fees for Special Sworn Status)

Timeframe – “Patience is a Virtue”

- Census Data
 - Plan on 3 to 7 months for review process
 - Economic data requires IRS approval in addition to Census
- Other agency data
 - Timeline dependent on agency approval process
 - Census approval NOT required
- Special Sworn Status
 - 3 - 4 additional months for your “security clearance”

Guidelines for Census proposal

- Proposal must meet Basic Requirements
 - Need for *Non-Public* data
 - Maintains Confidentiality
 - Must emphasize statistical models vs. tabular output
 - Feasibility
 - Describes Census Benefits (LEGAL REQUIREMENT)
 - Scientific Merit
- Work with Census Administrator to Craft Final Proposal

Working in the FSRDC lab

- All analysis conducted in the RDC lab
 - Data located on server in Maryland
 - Access data via thin client terminals located in cubicles
- No internet access or personal computers allowed in lab
- Statistical software available: SAS, Stata, R, Matlab, Python etc.
- Agency reviews output before releasing
 - Penalty for disclosure is \$250,000 and/or 5 yrs in prison (inadvertent or otherwise)
- On-site RA free of charge
- Some grants available (\$1,000 travel grants, \$15,000 proposal)

Contact Information

- People:

- Melissa R. Banzhaf, ARDC Administrator
melissa.r.banzhaf@census.gov, 404-498-7538
- Julie L. Hotchkiss, ARDC Executive Director
Julie.L.hotchkiss@atl.frb.org, 404-498-8198

- Resources:

- ARDC website:
www.atlantardc.org
- CES Working Paper Series:
<https://ideas.repec.org/s/cen/wpaper.html>