Georgetown Federal Statistical Research Data Center
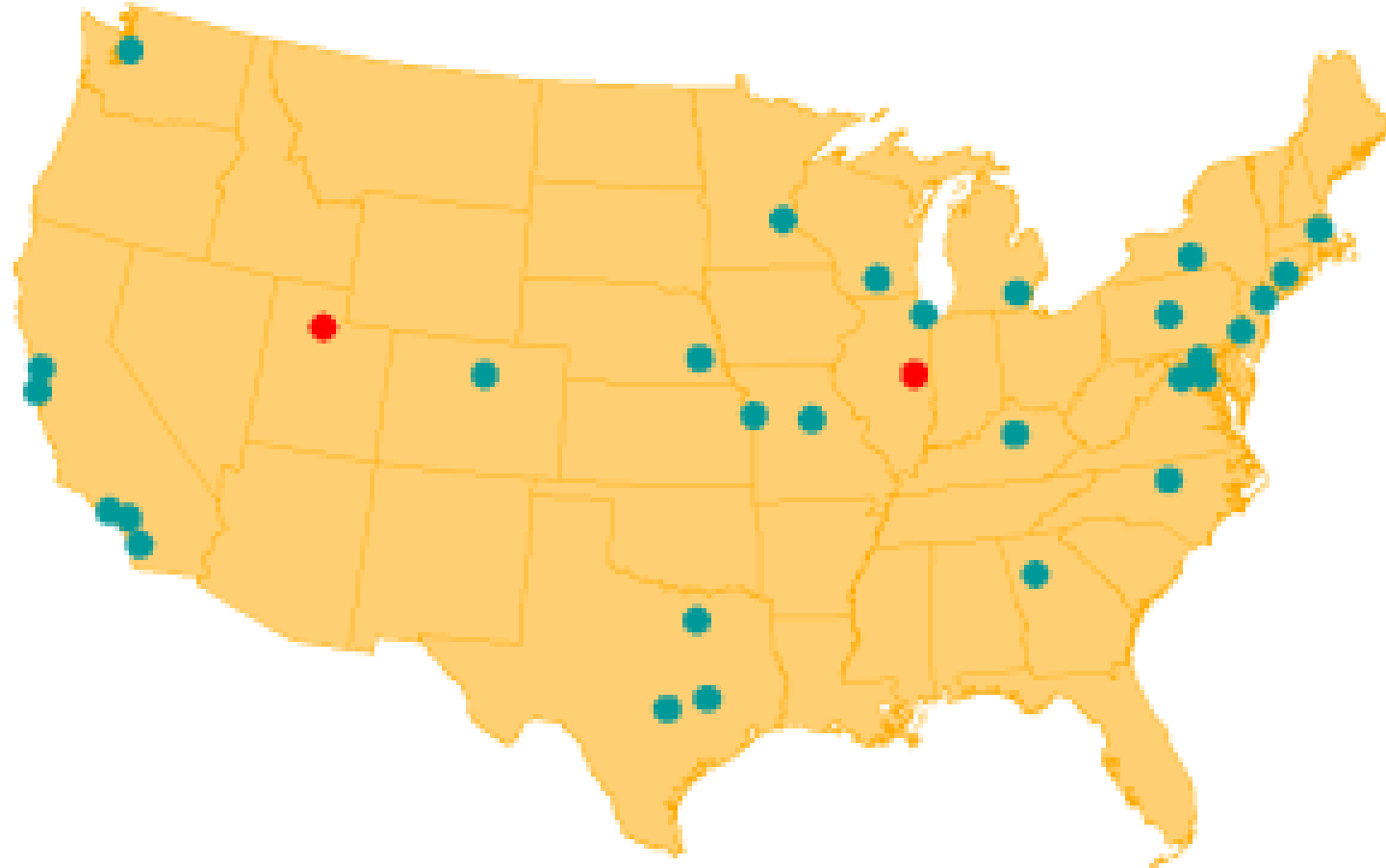
Amy O'Hara, Executive Director

Jelena Leathard, Administrator

United States Census Bureau

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

What is a Federal Statistical Research Data Center (FSRDC)?

- A secure Federal facility that provides approved researchers access to restricted microdata from multiple federal agencies

- Made possible by a contractual agreement between local institutions, the U.S. Census Bureau, and other Federal agencies

- FSRDCs are managed by an on-site Census employee—the administrator—who guides researchers on proposal development, enforces security guidelines, and serves as liaison with the research community

United States™
Census
Bureau

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

FSRDC Locations

Why is Census required to restrict access?

- Titles 13 (Census); 26 (IRS) U.S.C.; CIPSEA protect confidentiality:
  - The respondent cannot be identified
  - Only Census employees and temporary staff can access microdata
  - **Access must help the Census Bureau fulfill its statutory mission**

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

Federal Statistical Agencies

- Bureau of the Census
- National Center for Education Statistics
- National Agricultural Statistics Service
- Bureau of Labor Statistics
- Bureau of Economic Analysis
- National Center for Science and Engineering Statistics
- Statistics of Income Division (Internal Revenue Service)
- Bureau of Justice Statistics
- Energy Information Administration
- National Center for Health Statistics
- Bureau of Transportation Statistics
- Office of Research, Evaluation and Statistics (Social Security Administration)

United States™
Census
Bureau

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

5

Advantages to Researchers

- Greatly expands the policy and basic questions that can be addressed
- Builds on past research findings with richer data
- Improves competitive edge for grants and publications (60% of pubs in top journals)
- Improves graduate education (big data/statistical techniques) and placement

Opportunities for Graduate Students

- Cornell University Course by Videoconference, Spring Semester, held in IBS on Mondays

- Census Dissertation Mentorship Program

- Census Dissertation Fellowship Program (up to $50,000)

FSRDCs Address Many Research Topics

- Business, Trade, Finance, and Management
- Crime and Crime Victimization
- Demography, Population Distributions and Trends, Migration, and Immigration
- Economics, Labor Markets, Entrepreneurship, Employment and Industry
- Education and Education Policy
- Hazard Mitigation, Environmental Impact Assessment, Pollution Abatement
- Health and Well-Being, Health Insurance, Health Policy
- Housing, Housing Markets, and Residential Patterns
- Poverty, Social Welfare Policy, and Social Mobility
- Transportation Analysis and Planning
- Urban and Regional Economics and Planning
- Energy Efficiency and Greenhouse Gas Emissions in Manufacturing

Advantages of Using FSRDC for Census data

- Microdata not available publicly
  - firms and establishments
  - individuals and households
- Variables not available in public versions of data sets
  - (e.g., low level geography)

- Full population counts or larger samples (Decennial Census, ACS, CPS)

- Full range of response items (e.g., industry codes, occupational codes, detailed race answers, income is not top-coded, etc.)

- Ability to make linkages

  - with external data (e.g., via geocodes, establishment ID, etc.)

  - between multiple internal data sets via non-public link keys

Partners in the FSRDC System

- U.S. Census Bureau

  - Economic Data (microdata on firms and establishments)

  - Demographic Data (survey data on individuals and households; administrative data on individuals)

  - Employer-Employee Linked Data (Data on employees linked with data on employers)

- Bureau of Labor Statistics (BLS)

- National Center for Health Statistics (NCHS)

- Agency for Healthcare Research and Quality (AHRQ)

- Other Federal Partners

United States™ Census Bureau

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

GURDC

- **Location:** Healy Hall G05

- 12 thin client workstations to access FSRDC servers

- Secure communications that tunnel over campus internet

- Contains the Administrator's office

- Badge Reader at Entrance

- 24/7 Security System with camera

- No electronic devices allowed

- NOTHING leaves the secure lab without approval

United States™
Census
Bureau

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

Requirements for using FSRDC

- Research projects must undergo a formal approval process with the agency that owns the data, e.g., Census, NCHS, BLS

- Researchers must go through a background investigation that qualifies them for "Special Sworn Status (SSS)" which makes them an unpaid Census Bureau employee

- Results must be formally reviewed for disclosure violation before they leave the secure facility

Guidelines for Census Proposals

- 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

Proposal Differences by Agency

	Census	NCHS	AHRQ	BLS
Approval rate	70%	Nearly 100%	90-100%	TBD
Time to Approval	4-6 months on average	8-10 weeks on average	3-6 months	TBD
Benefit to Agency	Required and weighted heavily	Not required	Not required	Benefit statement required
Fee	None	One time extraction fee ($3000)	None	$4000 annual
Scope	Broad (max of 30 pages)	Precise	5-10 pages	TBD

Timeframe – "Patience is a Virtue"

- Census Data
  - Plan on 6 to 12 months
  - Title 13 (Census approval only) vs. Title 26 (Census & IRS approval)

- NCHS/AHRQ/BLS Data
  - Timeline dependent on agency approval process
  - Census approval NOT required

- Special Sworn Status
  - 3 additional months for your "security clearance"

United States **Census** Bureau

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

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, etc.

- Agency reviews output before releasing
  - Penalty for disclosure is $250,000 and/or 5 yrs in prison (inadvertent or otherwise)

United States™
Census
Bureau

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

How and When Do I Get Started?

- See materials at the Census RDC website, the CES website, and other relevant sites.

- Contact the GURDC Director and Administrator for
  - data availability
  - project budget and timeline
  - contact information

- The Administrator will give invaluable guidance on Census projects:
  - the proposal development process
  - the benefits to Census
  - help navigate the project approval process

United States **Census** Bureau

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

Contact Information

- Resources:
  - FSRDC website:
    https://www.census.gov/fsrdc
- People:
  - Amy O'Hara – Amy.OHara@georgetown.edu
  - Jelena Leathard – Jelena.Leathard@census.gov

United States Census Bureau

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

Restricted-Use Economic Census Microdata

Data Set	Unit of Enumeration
Census of Auxiliary Establishments	Establishment
Census of Construction Industries	Establishment
Census of Finance, Insurance, and Real Estate	Establishment
Census of Manufactures	Establishment
Census of Mining	Establishment
Census of Retail Trade	Establishment
Census of Services	Establishment
Census of Transportation, Communications, and Utilities	Establishment
Census of Wholesale Trade	Establishment

Restricted-Use Establishment Microdata

Data Set	Unit of Enumeration
Annual Survey of Manufacturers	Establishment
Manufacturing Energy Consumption Survey	Establishment
Medical Expenditure Panel Survey – Insurance Components (MEPS-IC)	Establishment
Quarterly Survey of Plant Capacity Utilization / Survey of Plant Capacity Utilization	Establishment
Survey of Manufacturing Technology	Establishment
Survey of Pollution Abatement Costs and Expenditures	Establishment
Management and Organizational Practices Survey (MOPS)	Establishment

Restricted-Use Firm Microdata

Data Set	Unit of Enumeration
Annual Capital Expenditure Survey	Firm
Annual Retail Trade Survey	Firm
Business Research & Development and Innovation Survey (BRDIS)	Firm
Survey of Business Owners	Firm
Annual Survey of Entrepreneurs	Firm
Kauffman Firm Survey	Firm
Quarterly Financial Report	Firm

United States Census Bureau™

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

Restricted-Use Transactions Microdata

Data Set	Unit of Enumeration
Commodity Flow Survey	Establishment
Foreign Trade Data – Exports	Transaction
Foreign Trade Data – Imports	Transaction
Longitudinal Firm Trade Transactions Database	Transaction

United States™ Census Bureau

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

Restricted-Use Business Register Microdata

Data Set	Unit of Enumeration
Standard Statistical Establishment List / Business Register	Establishment
Longitudinal Business Database (LBD)	Establishment
Integrated Longitudinal Business Database (ILBD)	Establishment

Economic Example #1

"Raising the Barcode Scanner: Technology and Productivity in the Retail Sector" Basker *American Economic Journal: Applied Economics* (2012)

- The author examines the effect of technology on the retail sector by investigating the impact of bar code scanners on productivity.

- She matches data from the Food Marketing Institute on store scanner installations in the 1970s and 80s to the **Census of Retail Trade** for 1972, 1977, 1982.

- She finds productivity increased by an average of 4.5% in stores that installed scanners, reducing labor costs by reducing demand for both skilled and unskilled workers.

Demographic Survey Datasets

- Decennial Surveys 1950-2010
- 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

Demographic Administrative Datasets

- Numident (Social Security Administration)
  - Birth place, Birth date, Death Date (Social Security records)
  - Has PIK

- Moving to Opportunity Survey (MTO)
  - Designed to help low-income families in public housing move to "opportunity" neighborhoods with vouchers and counseling.
  - Ran in five large cities -- Baltimore, Boston, Chicago, Los Angeles, and New York -- between September 1994 and August 1998 (4,608 families randomly assigned)

United States **Census** Bureau

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

Demographic Example #1

"Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes" Bayer, Ross, and Topa *Journal of Political Economy* (2008)

- The authors find a significant effect of social networks on hiring, especially among those with similar socio-demographic characteristics.

- Use **Decennial** survey with census block of residence and census block of work to look for social hiring networks.

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 #1

"Human Capital Loss in Corporate Bankruptcy" Graham et al., CES Working Paper (2013)

- This paper quantifies the "human costs of bankruptcy" by estimating employee wage losses induced by bankruptcy filings and finds annual wages decline by 30% one year after bankruptcy.

- 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)
  - People paid on the grants
  - Businesses paid by the grants

- Can link grant recipients to internal Census Bureau data on individuals and firms

UMETRICS Example #1

"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 #1, 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.

United States Census Bureau

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

Health Data

- These data are collected by:
  - National Center for Health Statistics (NCHS)
  - Agency for Healthcare Research and Quality (AHRQ)

- FSRDC hosts the data

- Proposal review and approval is handled by these agencies

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

Restrictions for NCHS/AHRQ

- Geographic variables
- Most dates (date of interview, date of death, date of birth)
- Income and employment data
- Specific diagnoses
- Details about facilities (accreditation, payments, number of employees)
- Some information about children and adolescents
- Some information about race, ethnicity, and country of origin
- Contextual data (nearest hospital, % of population with diploma)
- Sample design variables (necessary for estimating variances)

NCHS Example

"Tuskegee and the Health Of Black Men" Aslan and Wanamaker *Quarterly Journal of Economic* (Forthcoming)

- Use **NHIS** data in addition to other data sources to examine the effect of the public disclosure of the Tuskegee Study in 1972 on behavior and health of older black men.

- They find decreases in both outpatient and inpatient physician interactions for older black men, among other results.

AHRQ Example

"The Effects of Medical Marijuana Laws on Utilization of Prescribed Opioids and Other Prescription Drugs" Ozluk, Working Paper (2017)

- Uses 1996 and 2014 of **MEPS** to estimate the effects of Medical Marijuana Laws (MMLs) on prescription drug utilization, with a focus on opioids.

- She finds find MMLs lead to a $2.47 decrease in per person prescribed opioid spending among young adults (ages 18-39) over a year.

United States **Census** Bureau™

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

Census Health Data

- Medical Expenditure Panel Survey-Insurance Component (MEPS-IC)
  - Surveys employers and collects data on health insurance options
  - Collects data on characteristics of firms
  - Can link to other Census business microdata (e.g., ASM, LBD)

- Economic Census and annual economic surveys (e.g., ASM)
  - Collects data on employers' annual health insurance expenditures
  - Can link to other Census business microdata

BLS Datasets

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

SNAP / TANF / WIC

- Can be linked with Census data (e.g. ACS, CPS, SIPP, etc.)
- Data availability varies by state
- Must provide benefits to the Food and Nutrition Service
  - Evaluate and improve surveys
  - Provide estimates to improve understanding of poverty and food insecurity
  - Contribute to evidence-based policy and program evaluation

SSA-Survey Linked Data

- **Master Earnings File:** Individual earnings records since 1951

- **Payment History Update System:** Monthly payment records

- **Summary Earnings Record:** Lifetime earnings. Total earnings from 1937-1950, annual earnings from 1951-present

- **Supplemental Security Record:** Data for SSI program

SSA-Survey Linked Data

Data Set	Linkage
Detailed Earnings Record	SIPP, CPS
Summary Earnings Record	SIPP, CPS
Payment History Update System	SIPP, CPS
Supplemental Security Record	SIPP, CPS
Master Earnings File	SIPP

United States Census Bureau

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