

Conducting Economic Research at the University of Missouri Research Data Center

Jacob Cronin

University of Missouri Research Data Center Administrator

United States Census Bureau

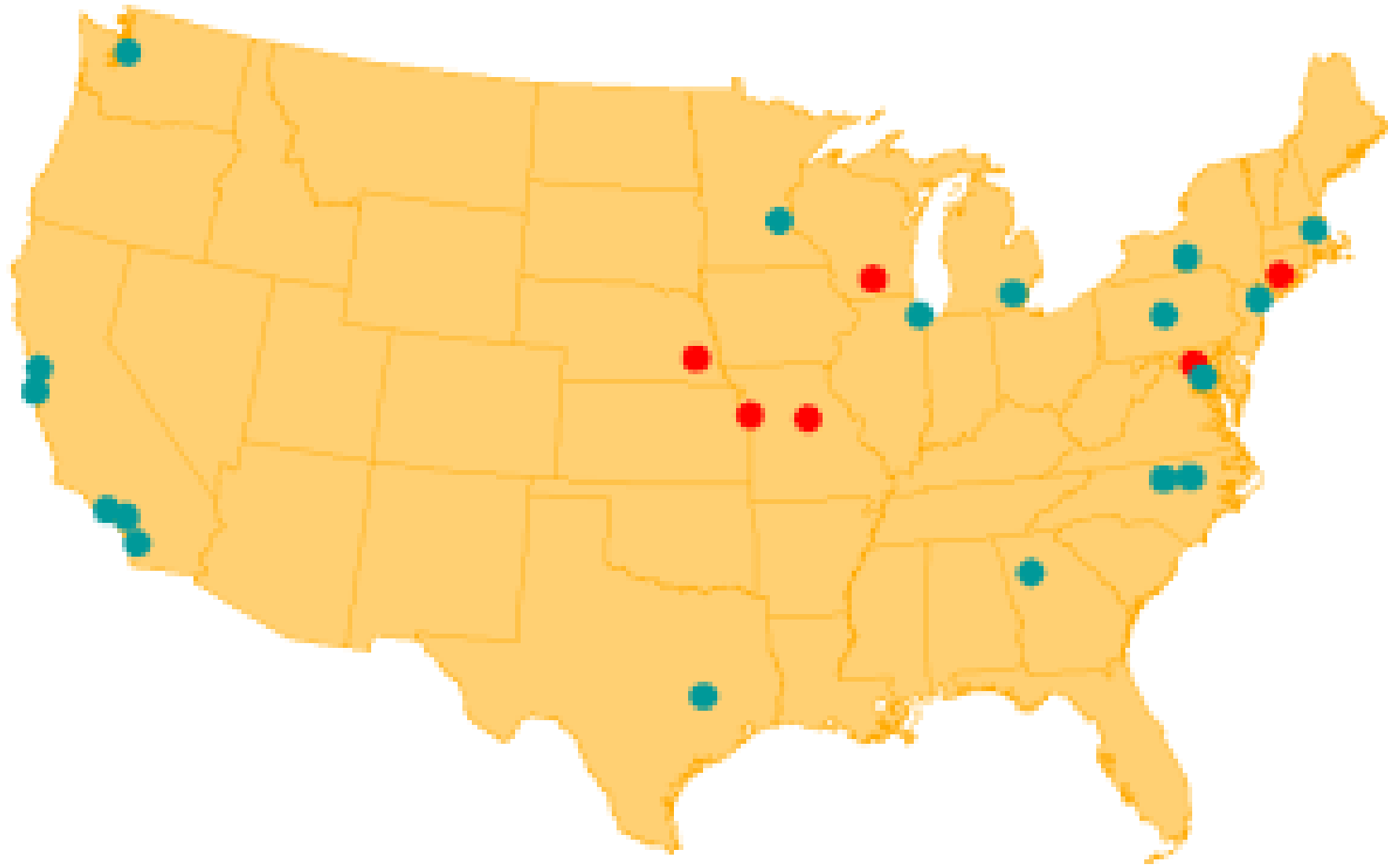
October 8, 2016

Overview

- Background and purpose of Federal Statistical Research Data Centers
- Overview of economic data available through the RDC
 - Business Register and LBD, economic censuses, surveys of firms and establishments, LEHD, and transactions data
 - Examples of research using these different datasets
- Process for accessing RDC data
- Contact information and sources for additional information
- Questions

What is a Federal Statistical Research Data Center?

- A secure environment where researchers can undertake analysis of non-public data from federal agencies
- MU joined approximately 25 RDC sites nationwide.
- The RDC allows for research that would be difficult (if not impossible) otherwise



Types of Data Available

- **Demographic Data**

- Census data (Decennial Census, CPS, ACS, SIPP, etc.)
- Am. Housing Survey, Natl. Crime Victimization Survey

- **Health Data**

- National Center for Health Statistics
- Agency for Healthcare Research & Quality

- **Economic Data (today's focus)**

- Business register data
- Transactions data (trade flows)
- LEHD – Longitudinal Employer/Household Data

Economic Data

- The Census RDC website lists over 40 different dataset as “Economic Data”
- Economic data refer to the Economic Census of establishments and various surveys and data for establishments and firms to measure economic activity.
- With very few exceptions, the public use versions for these files are limited to data presented in aggregate form.

Business Register (BR) or the Standard Statistical Establishment List (SSEL)

- Comprehensive list of business establishments with payroll at any time during the past three years.
 - 1974-2013
- Uses administrative records from the IRS, SSA, and BLS as well as data from the economic censuses
 - Has data on payroll, employment, sales, and industry classification
 - Has physical and mailing address and business name
- Considered backbone of economic research because it serves as the sampling frame for censuses and surveys and provides definitive list of all establishments and firms

Data on Establishments in Single Merged Longitudinal Database

- Longitudinal Business Database (LBD)
 - Annual information on all US establishments with paid employees
 - Over 8.5 million observations in 2013
 - Available 1976-2013
 - Can observe mergers, acquisitions, deaths, births of establishments and firms
 - Can be merged with economic censuses and surveys

Recent publications using restricted-use LBD

- Emek Basker and Javier Miranda. (2014). “Taken by Storm: Business Financing, Survival, and Contagion in the Aftermath of Hurricane Katrina.” *University of Missouri Department of Economics Working Paper 14-06 and Center for Economic Studies Working Paper 14-20*

Economic Censuses

- Based on survey of businesses
 - Since it is a census, it attempts to cover all units in universe
 - Response is required by law.
 - Most industries covered.
- Occurs every 5 years (years ending in “2” and “7”)
- Unit is the establishment
 - Defined as a site of work
 - Multi-establishment firms can be identified with their establishments

Economic Censuses

Data Set	Frequency	Available
Census of Auxiliary Establishments	5 years	1977-2007
Census of Construction Industries	5 years	1972-2012
Census of Finance, Insurance, Real Estate	5 years	1992-2012
Census of Manufactures	5 years	1963, 1967-2012
Census of Mining	5 years	1987-2012
Census of Retail Trade	5 years	1997-2012
Census of Services	5 years	1977-2012
Census of Transportation, Communications, Utilities	5 years	1987-2012
Census of Wholesale Trade	5 years	1977-2012

Economic Censuses

- Large number of variables available, depending on industry, including
 - Employment and labor costs
 - Number of employees
 - Employment by type (e.g. construction vs. others)
 - Payroll and wages
 - Sales, receipts, revenue, value of shipments/construction work done
 - Cost of materials, supplies and fuels
 - Often includes purchased services (e.g. advertising, legal services, etc.)
 - Capital expenditures, inventories, and assets

Recent publications using restricted-use economic censuses

- Andrew Bernard and Teresa Fort. (Forthcoming). “Factoryless Goods Producing Firms.” *American Economic Review*.
- Hector Perez-Saiz. (Forthcoming). “Building New Plants or Entering by Acquisition? Estimation of an Entry Model for the U.S. Cement Industry.” *The RAND Journal of Economics*.

Samples of Establishments

Dataset	Frequency	Availability
Annual Survey of Manufactures	Annual	1973-2013
Medical Expenditure Panel Survey – Insurance Component	Annual	1996–2006, 2008–2014
National Employer Survey	Various	1994, 1996, 1997, 1998
Quarterly Survey of Plant Capacity Utilization	Quarterly	1974–2013
Manufacturing Energy Consumption Survey	Various	1985, 1988, 1991, 1994, 1998, 2002, 2006, 2010

Surveys of Establishments

- Annual Survey of Manufacturers
 - Sample drawn from Census of Manufacturers to provide annual estimates
 - Approximately 50,000 establishments surveyed annually
 - Contains many of the same variables as the Census of Manufacturing
 - Employment, payroll, worker hours, cost of materials, capital expenditures, inventories, and energy consumption
- Can be linked with Census of Manufacturing

Survey of Establishments

- Medical Expenditure Panel Survey – Insurance Component
 - Survey of establishments to collect data on the number and types of private health insurance plans offered, benefits associated with these plans, annual premiums, annual contributions by employers and employees, eligibility requirements, and employer characteristics.
 - Valuable for studying the Affordable Care Act (ACA)
- National Employer Survey
 - Covers all establishments with 20 or more employees
 - Focuses on worker education, employer training and employer business characteristics, including business productivity.

Survey of Establishments

- Quarterly Survey of Plant Capacity Utilization
 - Collects data quarterly from about 7,500 manufacturing and publishing establishments selected from the Business Register
 - Collects statistics on establishment operational status, value of actual production, estimated production attainable at full and emergency conditions, and reasons for operating at less than full production capacity.
- Manufacturing Energy Consumption Survey
 - National sample survey that collects information on the stock of U.S. manufacturing establishment, their energy-related building characteristics, and their energy consumption and expenditures.

Samples of Firms

Dataset	Frequency	Availability
Annual Capital Expenditures Survey	Annually	1993–2012
Annual Retail Trade Survey	Annually	1999–2012
Annual Wholesale Trade Survey	Annually	1995–2012
Survey of Business Owners	Every 5 years	1982–2007
Business Research & Development and Innovation Survey (formerly SIRD)	Annually	2008–2012
Survey of Industrial Research and Development	Annually (discontinued)	1972–2007
Service Annual Survey	Annually	1991–2013

Samples of Firms

- Annual Capital Expenditures Survey
 - Annual sample of firms to provide broad-based statistics on business spending for new and used structures and equipment
- Annual Retail Trade Survey
 - Sample covers employer businesses classified in retail trade sector and accommodation and food services sector
 - Collects data on total annual sales, e-commerce sales, end-of-year inventories, purchases, total operating expenses, inventories held outside the United States, gross margins, and end-of-year accounts receivable.

Sample of Firms

- Annual Wholesale Trade Survey
 - Sample covers companies primarily engaged in wholesale trade
 - Sales, e-commerce, end-of-year inventories, inventories held outside the U.S., purchases, and total operating expenses
- Survey of Business Owners
 - Part of the Economic Censuses (sent to firms)
 - 1.75 million observations in 2012
 - Collects detailed information from business owners such as gender, race, ethnicity, and veteran status.

Sample of Firms

- Business Research & Development and Innovation Survey
 - Covers all firms with five or more employees
 - Collects R&D expenses of US based companies as well as statistics on the R&D workforce, intellectual property, technology transfer activities and innovation
- Survey of Services
 - Covers companies whose primary business or operation is to provide services to individuals, businesses, and governments
 - Include operating revenue, sources of revenue and expenses by type for selected industries, operating expenses for tax-exempt firms, and selected industry-specific items. Also has e-commerce data after 1999.

Transactions Microdata

Dataset	Frequency	Availability
Commodity Flow Survey (CFS)	Various	1993, 1997, 2002, 2007, 2012
Foreign Trade Data – Exports (EXP)	Monthly	1992–2011
Foreign Trade Data – Imports (IMP)	Monthly	1992–2011
Longitudinal Firm Trade Transactions Database (LFTTD)	Monthly	1992–2011

Transactions Microdata

- Commodity Flow Survey
 - Provides data on the movement of goods within the United States
 - Unit is establishment
 - Surveyed about 100,000 establishments from manufacturing, mining, wholesale, and selected retail and services establishments.
 - Commodities shipped, their value, weight, and mode of transportation, as well as the origin and destination of shipments of commodities Conducted in partnership with the Bureau of Transportation Statistics

Transactions Microdata

- Foreign Trade Data: Imports
 - Covers the universe of firms operating in the U.S. that engage in merchandise import from a foreign destination
 - Transaction level datasets: over 59 million observations in 2011
- Foreign Trade Data: Exports
 - Covers the universe of trade transactions data collected by the Customs Bureau during the exportation process
 - Transaction level datasets: over 35 million observations in 2011
- Longitudinal Firm Trade Transactions Database
 - Links individual trade transactions to the U.S. firms that make them
 - Allows for linkage to other census data products

Combining Workers and Firms

- LEHD: Longitudinal Employer-Household Dynamics Database:
 - Employment History Files (EHF): Quarterly UI Wage Records
 - Employer Characteristic Files (ECF): Employer data
 - Individual Characteristics Files (ICF): Person level demographic data
- Firm and worker information combined to create *job level* quarterly earnings history data, data on where workers live and work, data on firm characteristics, such as industry

Recent publications using restricted-use LEHD

- Sari Pekkala Kerr, William Kerr, and William Lincoln. (Forthcoming). “Skilled Immigration and the Employment Structures of US Firms.” *Journal of Labor Economics*
- Paige Quimet and Rebecca Zarutskie. (2014). “Who Works for Startups? The Relation Between Firm Age, Employee Age, and Growth.” *Journal of Financial Economics*.

Data Linkages

- Most census data products can be linked with each other via employer identification number (EIN) or other methods (such as name and address)
- Can link with outside data if the data has a suitable identifier (e.g. name and address or EIN)
 - Can link with ACS data or other aggregated data using geographic identifiers (e.g. state, county, census tract) to get contextual data
- Compustat-SSEL bridge
 - Compustat is a commercially available database providing stock prices and related financial information
 - Bridge file available in RDC to link at the establishment level

RDC Proposal Process

Getting authorization to use the MU RDC to
access federal data sources

Overview

- Researchers interested in using restricted-use Census datasets will need to go through a proposal process and obtain Special Sworn Status (SSS)
- Jake Cronin is the RDC Administrator who works full-time at MU to help researchers through the proposal process.

Proposal Process

- The goal is to draft proposals for feasible projects that provide academic merit as well as benefits to Census, with minimal disclosure risk
 - Feasible—Is it possible to answer this research question with the RDC data?
 - Academic merit—staff at Census with expertise in topic will review. Generally very lenient.
- You work with the RDC Administrator to write the proposal. The RDC Administrator then forwards the final proposal to Census or other agency for approval.

Benefits to Census

- By law, researchers must document how their research benefits the mission of the Census Bureau
 - MU's RDC administrator will work with you to identify the appropriate benefit
 - There are thirteen different benefit criterion. Most research meets multiple criterion.

Special Sworn Status

- After the proposal has been approved, researchers need to obtain Special Sworn Status (SSS) to physically access the RDC
 - SSS entails background check and oath to protect the confidentiality of the data
 - Requires completing paperwork, getting fingerprints, online trainings, identification verifications
 - Can take about 3 months to issue SSS
 - Process is tedious, but denial is uncommon

Working in an RDC

- Only researchers with approved projects and Special Sworn Status are allowed in the RDC.
- Researchers will work on “thin client” computers that tunnel into secure Census servers.
 - No data is stored on local machines
- All analysis must be done in the RDC.
- Output cannot be removed from the RDC prior to a formal disclosure avoidance review

Disclosure Avoidance Review

- There is a formal process to remove aggregated output out of the RDC.
- This is the only approved process for moving results from inside the confidential RDC to the outside public domain.
- The general idea is to ensure that no identifiable information on establishments, people, or households is disclosed.
 - In the proposal phase, it is important to think ahead about the output of the research and if it poses any disclosure risk

More information

- For information on restricted-use data available through the RDC:
- Economic Data:
 - <https://www.census.gov/ces/dataproducts/economicdata.html>
- LEHD Data
 - <https://www.census.gov/ces/dataproducts/lehddata.html>
- MU RDC website (google 'PEHC Missouri'):
 - <https://truman.missouri.edu/population-education-and-health-center/mu-research-data-center/>

MU RDC Contact Information

Jacob Cronin

MU RDC Administrator

Jacob.Cronin@census.gov

573-884-9122

Colleen Heflin

MU RDC Co-Director

heflincm@Missouri.edu

Peter Mueser

MU RDC Co-Director

mueserp@Missouri.edu