Research Using Restricted Data

Nichole Szembrot
Cornell Federal Statistical Research Data Center
Census Bureau, Center for Enterprise Dissemination
Outline

• FSRDC overview
• Available data
• Proposal process
• Questions
Overview
What are Research Data Centers (RDCs)?

• Secure Federal facilities that provide approved researchers access to restricted microdata

• A partnership between local institutions, the US Census Bureau, and other Federal statistical agencies
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
Restricted Data Advantages

• No publicly-available microdata
  • Internal data at the establishment and firm level
  • Universal scope
  • Detailed industry and geography

• Linking data
  • Consistent identifiers
  • Business register
  • External data

• Less top-coding and more detail for demographic data
Data
## Restricted-Use Economic Census Microdata

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Enumeration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census of Auxiliary Establishments</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Construction Industries</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Finance, Insurance, and Real Estate</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Manufactures</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Mining</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Retail Trade</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Services</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Transportation, Communications, and Utilities</td>
<td>Establishment</td>
</tr>
<tr>
<td>Census of Wholesale Trade</td>
<td>Establishment</td>
</tr>
</tbody>
</table>
## Restricted-Use Establishment Microdata

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Enumeration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Survey of Manufacturers</td>
<td>Establishment</td>
</tr>
<tr>
<td>Manufacturing Energy Consumption Survey</td>
<td>Establishment</td>
</tr>
<tr>
<td>Medical Expenditure Panel Survey – Insurance Components (MEPS-IC)</td>
<td>Establishment</td>
</tr>
<tr>
<td>Quarterly Survey of Plant Capacity Utilization / Survey of Plant Capacity Utilization</td>
<td>Establishment</td>
</tr>
<tr>
<td>Survey of Manufacturing Technology</td>
<td>Establishment</td>
</tr>
<tr>
<td>Survey of Pollution Abatement Costs and Expenditures</td>
<td>Establishment</td>
</tr>
<tr>
<td>Management and Organizational Practices Survey (MOPS)</td>
<td>Establishment</td>
</tr>
</tbody>
</table>
# Restricted-Use Firm Microdata

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Enumeration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Capital Expenditure Survey</td>
<td>Firm</td>
</tr>
<tr>
<td>Annual Retail Trade Survey</td>
<td>Firm</td>
</tr>
<tr>
<td>Business Research &amp; Development and Innovation Survey (BRDIS)</td>
<td>Firm</td>
</tr>
<tr>
<td>Survey of Business Owners</td>
<td>Firm</td>
</tr>
<tr>
<td>Annual Survey of Entrepreneurs</td>
<td>Firm</td>
</tr>
<tr>
<td>Kauffman Firm Survey</td>
<td>Firm</td>
</tr>
<tr>
<td>Quarterly Financial Report</td>
<td>Firm</td>
</tr>
</tbody>
</table>
# Restricted-Use Transactions Microdata

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Enumeration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity Flow Survey</td>
<td>Establishment</td>
</tr>
<tr>
<td>Foreign Trade Data – Exports</td>
<td>Transaction</td>
</tr>
<tr>
<td>Foreign Trade Data – Imports</td>
<td>Transaction</td>
</tr>
<tr>
<td>Longitudinal Firm Trade Transactions Database</td>
<td>Transaction</td>
</tr>
</tbody>
</table>
# Restricted-Use Business Register Microdata

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Enumeration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Statistical Establishment List / Business Register</td>
<td>Establishment</td>
</tr>
<tr>
<td>Longitudinal Business Database (LBD)</td>
<td>Establishment</td>
</tr>
<tr>
<td>Integrated Longitudinal Business Database (ILBD)</td>
<td>Establishment</td>
</tr>
</tbody>
</table>
Demographic Data

<table>
<thead>
<tr>
<th>Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decennial Census</td>
</tr>
<tr>
<td>American Community Survey</td>
</tr>
<tr>
<td>American Housing Survey</td>
</tr>
<tr>
<td>Current Population Survey (and Supplements)</td>
</tr>
<tr>
<td>National Crime Victimization Survey (and Supplements)</td>
</tr>
<tr>
<td>National Survey of College Graduates</td>
</tr>
<tr>
<td>Rental Housing Finance Survey</td>
</tr>
<tr>
<td>Survey of Income and Program Participation</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study</td>
</tr>
</tbody>
</table>
Demographic Administrative Datasets

• Numident (Social Security Administration)
  • Birth place, Birth date, Death Date (Social Security records)

• 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)
  • Data collected on participants multiple times from 1994 through 2010
  • Individuals in RDC files have been PIKed
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-Linked Data

• **Master Beneficiary Record:** Entitlement and payment data for retirement and disability
• **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-Linked Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Linkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed Earnings Record</td>
<td>SIPP, CPS</td>
</tr>
<tr>
<td>Summary Earnings Record</td>
<td>SIPP, CPS</td>
</tr>
<tr>
<td>Master Beneficiary Record</td>
<td>SIPP, CPS</td>
</tr>
<tr>
<td>Payment History Update System</td>
<td>SIPP, CPS</td>
</tr>
<tr>
<td>Supplemental Security Record</td>
<td>SIPP, CPS</td>
</tr>
<tr>
<td>Master Earnings File</td>
<td>SIPP</td>
</tr>
</tbody>
</table>
Linked Data - LEHD

- Link 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 Census business data via firm identifier
- Can link LEHD to Census household data via individual identifier
- Available on a state-by-state basis
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
## NCHS Data

https://www.cdc.gov/rdc/b2accessmod/acs220.htm

rdca@cdc.gov

### 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
AHRQ Data

https://meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp#
CFACTDC@AHRQ.HHS.GOV

<table>
<thead>
<tr>
<th>Restricted MEPS Data Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Component-Insurance Component linked file</td>
</tr>
<tr>
<td>Nursing Home Component</td>
</tr>
<tr>
<td>Medical Provider Component (except directly identifiable data)</td>
</tr>
<tr>
<td>Two-Year, Two-Panel Files</td>
</tr>
<tr>
<td>Area Health Resources Files</td>
</tr>
<tr>
<td>MEPS Link Files to NHIS</td>
</tr>
</tbody>
</table>
BLS Data

https://www.bls.gov/rda/?#
RDA_Admin@bls.gov

<table>
<thead>
<tr>
<th>Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Longitudinal Surveys</td>
</tr>
<tr>
<td>Survey of Occupational Injuries and Illnesses</td>
</tr>
<tr>
<td>Census of Fatal Occupational Injuries</td>
</tr>
<tr>
<td>National Compensation Survey</td>
</tr>
</tbody>
</table>
## BEA Data

**https://www.bea.gov/research/special-sworn-researcher-program**  
specialswornresearch@bea.gov

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. MNEs worldwide (BE-10/11)</td>
<td>In FSRDC</td>
</tr>
<tr>
<td>Foreign MNEs in the U.S. (BE-12/15)</td>
<td>In FSRDC</td>
</tr>
<tr>
<td>Trade in selected services and intellectual property (BE-120/125)</td>
<td>In FSRDC</td>
</tr>
<tr>
<td>Trade in financial services (BE-180/BE-185)</td>
<td>Expected in 2019</td>
</tr>
<tr>
<td>Trade in insurance transactions by U.S. insurance companies (BE-45/140)</td>
<td>Expected in 2020</td>
</tr>
<tr>
<td>New foreign direct investment in the U.S. (BE-13)</td>
<td>Expected in 2020</td>
</tr>
<tr>
<td>U.S. direct investment abroad (BE-577)</td>
<td>Expected in 2020</td>
</tr>
<tr>
<td>Foreign direct investment in the U.S. (BE-605)</td>
<td>Expected in 2020</td>
</tr>
</tbody>
</table>
Proposal Process
Approval Process

• Step 1: Approval from RDC
• Step 2: Census approval
• Step 3: Sponsoring agency approval
• Step 4: Background check
Timeframe

• Census data
  • Plan on 2-3 months for Census review and 6 months for IRS (if necessary)
  • Title 13 (Census only) vs. Title 26 (Census and IRS)

• NCHS/AHRQ/BLS/BEA Data
  • Timeline depends on agency
  • Census approval not required

• Special Sworn Status
  • 2-3+ months for security clearance (concurrent with IRS review, if necessary)
Background Check

• Offline paperwork and documentation
• Online trainings and certifications
• Background check
  • Submitted online and followed with interview
    • Residential history
    • Foreign travel
    • Education and employment history
    • References

• Fingerprinting
Proposal Structure

• Introduction / Background / Literature Review
• **Census Benefits**
• Methodology
  • Estimation equations and constructed measures
• Data requested
• Output / Disclosure Risk
• Timeline / Project Duration / Funding
Proposal Evaluation

• Proposal readers are mostly data experts
• Key points to emphasize in proposal:
  • **Benefit to Census**
  • Feasibility and scientific merit
  • Justified need for restricted data
  • Avoids disclosure risk
  • Emphasizes statistical models over tabular output
Key Proposal Questions

- Is this proposal feasible?
  - What data will be used and how will it be linked?
  - Is restricted data necessary? Can the question be answered with public data?
  - What methods will be used?

- What is the disclosure risk?
  - What are the sample sizes?
  - Are implicit samples being created?

- What is the Census Bureau benefit?
Benefits

1. Evaluating practices, continued relevance
2. Demographic and economic processes that affect Census programs
3. Increasing utility of data for analyzing public programs, public policy, and/or demographic, economic, or social conditions
4. Conducting data collection, processing, or dissemination
5. Understanding/improving data quality
6. Improved methodology to collect, measure, or tabulate
7. Enhancing the data collected (e.g. imputation, links)
Benefits

8. Limitations/improving the Business Register, Master Address File, industrial/geo classification
9. New data collection needs
10. Constructing, verifying, improving sampling frame
11. Estimates/characteristics of population
12. Methodology for estimating non-response
13. Developing statistical weights for survey
Output / Disclosure Avoidance Review

• No output can leave the RDC without review
• Clearly identify samples
  • No individual person or business can be identifiable in the release
• Performed by Administrator and a disclosure officer
• 2-4 weeks (generally)
  • Intermediate output discouraged
  • Descriptive results may be problematic
  • Focus on statistical estimates for release
Timeline

• List of major milestones
  • When will you complete the data matching, construction of extracts, etc.?  
  • How do you expect the project to unfold?  
  • When you will request disclosure?

• Extensions are not granted
Conclusion

• “Upon completion of the project...we will include a report describing how the project met Title 13, Chapter 5 requirement...

• ...We will also provide all programs, outputs, and findings to the Census Bureau and submit a technical paper to the Working Paper Series.”
Nuts and Bolts of RDC Research

• Research conducted on-site
• Computing environment (SAS, Stata, R, Matlab, etc.)
• Physically secure location, restricted area with badge access
• No internet, phones, or laptops allowed in lab
• No paper or output allowed outside of lab
• Disclosure Avoidance review required to present results
  • Discussion of specific results allowed only inside RDC (even among co-authors)
Contacts

• Nichole Szembrot, Cornell RDC Administrator
  • nichole.e.szembrot@census.gov
  • 607-255-8603
• Warren Brown, Cornell RDC Research Director
  • wab4@cornell.edu
• William Block, Cornell RDC Executive Director
  • block@cornell.edu