Federal Statistical Research Data Centers (FSRDC)

Samuel R. Bondurant
Dallas-Fort Worth Federal Statistical Research Data Center
Center for Enterprise Dissemination
US Bureau of the Census

Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau.
Goals of this Presentation

• Overview of FSRDC program
  • Description of working environment
• Phases of an RDC Project
• Available data products and advantages
• How to get started
Federal Statistical Research Data Center (FSRDC)

• Federally secure research lab where authorized researchers are granted secure access to restricted-use microdata
  • Census Bureau and other statistical agencies
    • NCHS, AHRQ, BLS, BEA, CJARS, etc.
  • Fees to maintain facility typically paid by a local consortium of research institutions
    • Additional fees for non-Census agency projects
Research Data Centers

1982
Census Bureau
Research Data Centers

1994
Boston/NBER
Research Data Centers

1995 - 2004

UC Berkeley
UCLA
Duke/RTI
Chicago Fed
U Michigan
Cornell
Research Data Centers

2005 - 2013
CUNY Baruch
Stanford
U Minnesota
Atlanta Fed
U Washington
Texas A&M
Research Data Centers

2014 - Present

- USC
- UC Irvine
- Penn State
- Yale
- U Missouri
- U Wisconsin
- Kansas City Fed
- U Maryland
- U Nebraska
- Georgetown
- U Kentucky
- Philadelphia Fed
- Colorado, Boulder
- UT Austin
- Dallas Fed
- Federal Reserve
- U Utah
- UIUC
- U Florida
What is an RDC?

*Not the actual lab
Census Servers
- Project space
- Statistical packages
- Data
Why Is Census Required to Restrict Microdata Access?

• Titles 13 (Census) & 26 (IRS) U.S.C. and CIPSEA protect confidentiality
  • respondents cannot be identified
  • only Census employees and temp staff can access microdata
  • use limited to statistical purposes
  • access must benefit Census Bureau programs
RDC Resources

• www.dfwrdc.org
  • Data
  • Proposal Process
  • Presentations

• https://ideas.repec.org/s/cen/wpaper.html
  • Working paper series
  • Google “census ces working paper series”
Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census
Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census
Proposal Process

• Email me and Wenhua
  • Include a quick description of your research interest
  • Schedule a video call

• Preliminary proposal submission portal
  • *portal currently being updated*
  • ResearchDataGov.org
Proposal Process

• Census proposals
  • Involve RDC Admin (potentially other Census staff)
  • Research_Proposal_Guidelines.pdf

• NCHS, AHRQ, BLS, BEA, etc. proposals
  • Separate administrators from respective agencies
  • Additional fees associated

• Note: on the dfwr dc.org website
  • Census proposal guidelines (pdf)
  • Links to NCHS, AHRQ, BLS, BEA proposal processes
*Census* Proposal

1. Proposal (Project Description)
   - Overview
   - Requested Datasets and years
   - Why you need internal data
   - Methodology
     - Linking process
     - Estimating equations
     - Measures constructed
   - Expected Output
   - Duration and Funding

2. Benefits to the Census Bureau (Predominant Purpose Statement)
   - Minimum of two (one being the working paper)

3. Abstract
13 Criteria for Benefits

1. Evaluating practices, continued relevance
2. Demog., econ. 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 the quality of data
6. Improved methodology to collect, measure, or tabulate
7. Enhancing the data collected (e.g. imputation, links)
8. Limitations/improving the Business Register
9. New data collection needs
10. Constructing, verifying, improving sampling frames
11. Estimates CHARACTERISTICS OF POPULATION
12. Methodology for estimating non-response
13. Developing statistical weights for a survey
Distribution of Benefit Criterion

Percent

Benefit Criterion

Data Quality

Increasing Data Utility

Data Enhancement

Population Estimates

United States Census Bureau
Phases of an RDC Project

• Proposal drafting
• Submit proposal for review
• Security clearance (SSS)
• RDC analysis
• Statistical output for Disclosure Avoidance Review
• Provide working paper and benefit to Census
Phases of an RDC Project

- Proposal drafting
- **Submit proposal for review**
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census
Proposal Review

• Census Review
  • Scientific merit
  • Need for non-public data
  • Risk of disclosure
  • Feasibility
  • Benefits to Census strength

• Additional Agency Review, if needed
  • Economic data (IRS), Numident (SSA), LEHD (states), CPS (BLS)
Phases of an RDC Project

• Proposal drafting
• Submit proposal for review
• Security clearance (SSS)
• RDC analysis
• Statistical output for Disclosure Avoidance Review
• Provide working paper and benefit to Census
Phases of an RDC Project

• Proposal drafting
• Submit proposal for review
• Security clearance (SSS)
• RDC analysis
• Statistical output for Disclosure Avoidance Review
• Provide working paper and benefit to Census
Special Sworn Status (SSS)

• All persons working in the RDC require SSS
  • Census, NCHS, AHRQ, BLS, BEA, etc.
  • Census absorbs SSS cost, other agencies might not

• Couple months to process (longer for non-citizens)
  • Application process can start after Census project approval
  • Non-citizens must reside in the U.S. for 3 of last 5 years to qualify
  • All SSS researchers must reside in U.S. for majority of the year
Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census
Phases of an RDC Project

• Proposal drafting
• Submit proposal for review
• Security clearance (SSS)
• RDC analysis
• Statistical output for Disclosure Avoidance Review
• Provide working paper and benefit to Census
RDC Analysis

- All work involving restricted-use data will occur in the RDC
- Free parking available at Federal Reserve Bank of Dallas
  - RDC accessible 24/7
- Statistical software packages available on server
- External user data/code can be uploaded into the project space
- Large data files, takes time to organize
- Projects last from 3-5 years

*Not the actual lab*
Virtual RDC Analysis

• Title 13 data only (non-FTI data, person/household)
  • ACS, CPS, SIPP, Decennial, AHS, etc.
  • No SSA data available (Numident, LEHD ICF, etc.)

• Access the RDC environment from home
  • Start project in the physical RDC
  • Additional VRDC orientation
  • Work the remainder of project from home

• Simulate the security requirements of the RDC
Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census
Phases of an RDC Project

• Proposal drafting
• Submit proposal for review
• Security clearance (SSS)
• RDC analysis
• Statistical output for Disclosure Avoidance Review
• Provide working paper and benefit to Census
Disclosure Avoidance Review

• All output from an RDC projects goes through disclosure avoidance review
  • Involved process assisted by RDC Admin
• Disclosure requirements different across data
• HQ Disclosure Office will email you approved output
Phases of an RDC Project

• Proposal drafting
• Submit proposal for review
• Security clearance (SSS)
• RDC analysis
• Statistical output for Disclosure Avoidance Review
• Provide working paper and benefit to Census
Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census
13 Criteria for Benefits

1. Evaluating practices, continued relevance
2. Demog., econ. 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 the quality of data
6. Improved methodology to collect, measure, or tabulate
7. Enhancing the data collected (e.g. imputation, links)
8. Limitations/improving the Business Register
9. New data collection needs
10. Constructing, verifying, improving sampling frames
11. Estimates/characteristics of population
12. Methodology for estimating non-response
13. Developing statistical weights for a survey
Timeline

• Census review, (1-3 months)
  • Assign to reviewers
  • Reviewing committee
  • (if needed) IRS review (3-7 additional months)

• After approval, SSS Packet

• Security clearance (2-3 months)
  • Additional time for non-citizens

• Project work from 3-5 years
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
Data Availability

• Census Bureau Projects:
  • Economic Data
  • Demographic Data
  • Longitudinal Employer-Household Dynamics (LEHD)
  • UMETRICS Data
  • Criminal Justice Administrative Records System (CJARS)

• Agency for Healthcare Research and Quality (AHRQ)
• National Center for Health Statistics (NCHS)
• Bureau of Labor Statistics
  • NLS, NCS, SOII, CFOI
• Bureau of Economic Analysis
## Economic Data

### Economic Censuses

<table>
<thead>
<tr>
<th>Establishment</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auxiliaries (AUX)</td>
<td>1977-2012</td>
</tr>
<tr>
<td>Construction Industries (CCN)</td>
<td>1972-2017</td>
</tr>
<tr>
<td>Finance, Insurance, Real Estate (CFI)</td>
<td>1992-2017</td>
</tr>
<tr>
<td>Manufacturers (CMF)</td>
<td>1963-2017</td>
</tr>
<tr>
<td>Mining (CMI)</td>
<td>1987-2017</td>
</tr>
<tr>
<td>Retail Trade (CRT)</td>
<td>1977-2017</td>
</tr>
<tr>
<td>Services (CSR)</td>
<td>1977-2017</td>
</tr>
<tr>
<td>Trans., Comms., Utilities (CUT)</td>
<td>1987-2017</td>
</tr>
<tr>
<td>Wholesale Trade (CWH)</td>
<td>1977-2017</td>
</tr>
</tbody>
</table>

### Business Register Data

<table>
<thead>
<tr>
<th>Establishment</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compustat-SSEL Bridge (CSB)</td>
<td>1976-2016</td>
</tr>
<tr>
<td>Form 5500 Bridge File</td>
<td>1992-2018</td>
</tr>
<tr>
<td>Integrated Longitudinal Business Database (ILBD)</td>
<td>1977-2018</td>
</tr>
<tr>
<td>Longitudinal Business Database (LBD)</td>
<td>1976-2019</td>
</tr>
<tr>
<td>Ownership Change Database (OCD)</td>
<td>1963-2002</td>
</tr>
<tr>
<td>Standard Statistical Establishment List/ Business Register (SSEL/BR)</td>
<td>1974-2019</td>
</tr>
</tbody>
</table>

### Transactions Data

<table>
<thead>
<tr>
<th>Establishment</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity Flow Survey (CFS)</td>
<td>1993-2017</td>
</tr>
<tr>
<td>Foreign Trade Data - Export (EXP)</td>
<td>1992-2019</td>
</tr>
<tr>
<td>Foreign Trade Data - Import (IMP)</td>
<td>1992-2019</td>
</tr>
<tr>
<td>Longitudinal Foreign Trade Transactions Data (LFTTD)</td>
<td>1992-2019</td>
</tr>
</tbody>
</table>

### Firm Surveys

<table>
<thead>
<tr>
<th>Establishment</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Business Survey</td>
<td>2017</td>
</tr>
<tr>
<td>Annual Capital Expenditures Survey (ACES)</td>
<td>1993-2019</td>
</tr>
<tr>
<td>Annual Retail Trade Survey (ARTS)</td>
<td>1999-2018</td>
</tr>
<tr>
<td>Business R&amp;D and Innovation Survey (BRDIS)</td>
<td>2008-2018</td>
</tr>
<tr>
<td>Exporter Database (EDB)</td>
<td>1992-2010</td>
</tr>
<tr>
<td>Service Annual Survey (SAS)</td>
<td>1991-2019</td>
</tr>
<tr>
<td>Survey of Business Owners (SBO)</td>
<td>1982-2012</td>
</tr>
<tr>
<td>Survey of Industrial R&amp;D (SIRD)</td>
<td>1972-2007</td>
</tr>
</tbody>
</table>
Economic Censuses

• Quinquennial, years ending in 2 or 7
• Covers ~4 million business locations
  • Covers most industries and all geographic areas of the US
  • 9 different versions tailored to industry type
• Establishment level
• Variables included:
  • Name/address/ownership information/industry codes
  • Revenues, employment, payroll
  • Customer types (households, contractors, gov’t, etc)
  • Sales by item code
  • Special inquiries depending on year and industry
Economic Censuses

• Intercensal years covered by other firm/establishment level datasets
  • CMF: Annual Survey of Manufactures (ASM)
  • CRT: Annual Retail Trade Survey (ARTS)
  • CSR: Services Annual Survey (SAS)
  • CWT: Annual Wholesale Trade Survey (AWTS)
Business Register (CBPBR) and Longitudinal Business Database (LBD)

• County Business Patterns Business Register
  • Covers all firms in the United States
  • From IRS tax records
  • # of employees, revenue, address, etc.

• Longitudinal Business Database
  • Universe of all US businesses w/ paid employees
  • Links establishments across years/surveys
  • Payroll, size, birth/death, ownership, location

• Integrated Longitudinal Business Database (ILBD)
  • Non-employer businesses
Additional Firm/Establishment Surveys

- Longitudinal Foreign Trade Transactions Data (LFTTD)
  - All transactions of $2,500+

- Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)

- Annual Business Survey (ABS)
  - Info on selected economic and demographic characteristics for businesses and business owners
    - Survey of Business Owners
    - Annual Survey of Entrepreneurs
    - Business R&D and Innovation for Microbusinesses Survey
# Demographic Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey (ACS)</td>
<td>1996-2020</td>
</tr>
<tr>
<td>American Housing Survey (AHS)</td>
<td>1984-2019</td>
</tr>
<tr>
<td>CPS Supplements</td>
<td>(1997-2017)</td>
</tr>
<tr>
<td>Decennial Census</td>
<td>1950-2010</td>
</tr>
<tr>
<td>National Crime Victimization Survey (NCVS)</td>
<td>2006-2017</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study (NLMS)</td>
<td>1973-2011</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>1984-2020</td>
</tr>
<tr>
<td>Census Numident</td>
<td>Present</td>
</tr>
</tbody>
</table>
RDC Demographic Data Advantages

- Individual and household data
- Less top-coding/censoring
- Additional variables
  - DOB
- Larger sample
- More detailed geography
- Individual level linkages across time/surveys
## More Detailed Geography

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Geography</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Housing Survey (AHS)</td>
<td>Tract</td>
</tr>
<tr>
<td>American Community Survey (ACS)</td>
<td>Block</td>
</tr>
<tr>
<td>Current Population Survey (CPS) – March Supplement</td>
<td>Tract</td>
</tr>
<tr>
<td>Decennial Census</td>
<td>Block</td>
</tr>
<tr>
<td>National Crime Victimization Survey (NCVS)</td>
<td>Block</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Mature Men/Women</td>
<td>Tract</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study (NLMS)</td>
<td>County</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>Tract/County</td>
</tr>
<tr>
<td>Master Address File Extract/Auxiliary Reference File</td>
<td>Address</td>
</tr>
</tbody>
</table>
Protected Identification Key (PIK)

• Unique person level identifier used to link data
• Link longitudinally and across surveys

Available in:

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS</td>
<td>2000 Decennial SF/LF</td>
</tr>
<tr>
<td>AHS</td>
<td>2010 Decennial</td>
</tr>
<tr>
<td>CJARS</td>
<td>NCVS</td>
</tr>
<tr>
<td>CPS-ASEC</td>
<td>Numident</td>
</tr>
<tr>
<td>CPS-Supplements</td>
<td>SIPP</td>
</tr>
<tr>
<td>LEHD data</td>
<td>And more</td>
</tr>
</tbody>
</table>

• ~80-90% of the samples are PIK’d
**LEHD System**

Covers 96% of all private-sector jobs

*Majority are matched using just the SEIN and not the SEIN+SEINUNIT pairing*
Longitudinal Employer-Household Dynamics (LEHD)

Economic Data
- Economic Census
  - Manufacturing
  - Wholesale Trade
  - Retail Trade
  - Etc.
- Annual Survey Manu.
- Kauffman Firm Survey
- Etc.

ECF/BRB

LEHD

Demographic Data
- ACS
- CPS
- SIPP
- Decennial
  - Short form
  - Long form
  - Etc.
Criminal Justice Administrative Records System (CJARS)

- Harmonized criminal justice records that can be linked anonymously at the person level
  - [https://cjars.isr.umich.edu/](https://cjars.isr.umich.edu/)

![Map of CJARS geographic and criminal justice domain coverage](image-url)
Data from Partnering Agencies Accessible in the RDC

- Agency for Healthcare Research and Quality
- National Center for Health Statistics
- Bureau of Labor Statistics
- Bureau of Economic Analysis

- Proposal process is with individual agencies and not Census
  - Security background check still with Census
Agency for Healthcare Research and Quality (AHRQ)

- Medical Expenditure Panel Survey – Household Component (MEPS-HC)
- Nursing Home Component (1996)
- Medical Provider Component (except directly identifiable data)
- Two-Year, Two-Panel Files
- Area Resource File (county-level data that can be linked to MEPS-HC)
Agency for Healthcare Research and Quality (AHRQ)

- Fully specified ICD-9 medical condition codes
- Fully specified industry and occupation codes
- Lower levels of Geography
  - State and county FIPS codes
  - Census tract and block-group codes
- Non-public use data elements
  - Asset information
  - Imputed NDC codes
- Federal and state marginal tax rates
NCHS Data Sets

• National Health and Nutrition Examination Survey (NHANES)
• National Health Care Surveys
  • National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS)
  • National Hospital Discharge Survey (NHDS)
  • National Nursing Home Survey (NNHS) and National Nursing Assistant Survey (NNAS)
  • National Home and Hospice Care Survey (NHHCS) and National Home Health Aide Survey (NHHAS)
• National Health Interview Survey (NHIS)
• National Survey of Family Growth (NSFG)
• State and Local Area Integrated Telephone Survey (SLAITS)
• NCHS Data Linkage Activities
  • Linked Mortality Data Products
  • Linked Medicare Enrollment and Claims Files Data
  • Linked Social Security Benefit History Data
  • Linked Air Quality Data
• National Vital Statistics System (NVSS)
National Center for Health Statistics (NCHS)

Types of Restricted Variables

- Geographic Variables
- Linked Data Products
- Genetic Variables (NHANES phenotype data)
- Temporal Variables
  - e.g. dates of birth, death, exams
- Detailed Race/Ethnicity Variables
- Sensitive Variables
  - e.g. youth sexual behavior and mental health
- Long Term Health Care Survey Merging Variables
# Publications by RDC Researchers and CES Staff: 2018 and Forthcoming

<table>
<thead>
<tr>
<th>Econ Journals (by rank)</th>
<th>RDC Researchers</th>
<th>CES staff</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA (1-5)</td>
<td>12</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>AA (6-20)</td>
<td>14</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>A (21-102)</td>
<td>22</td>
<td>15</td>
<td>37</td>
</tr>
<tr>
<td>B (103-258)</td>
<td>12</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>C (259-562)</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>D (562-1,202)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Journals outside of Econ</th>
<th>14</th>
<th>11</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book Chapters</td>
<td>3</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>82</td>
<td>63</td>
<td>145</td>
</tr>
</tbody>
</table>
How to get started?

[Emails]
samuel.r.bondurant@census.gov
wenhua.di@dal.frb.gov

• Send us a quick proposal pitch
  • Send me simple data questions

• Request a meeting in person
  • Include an itinerary if possible

• [Website 1]
  • [Website 2]