The California Census
Research Data Center

UC Davis  Vocational Education Cluster
April 25, 2014  Jon Stiles
Census RDCs

- What is an RDC?
- What data are available in the RDC?
- What kinds of research can be done with RDC resources?
- What is the process for getting access to RDC data?
What is a Census RDC?

- A partnership
- A set of services, tools and data
- A secure & vetted environment
The CCRDC is a joint project of the U.S. Bureau of the Census, UC Berkeley, Stanford and UCLA to enable qualified researchers with approved projects to access confidential, unpublished Census Bureau data.

CES on the web: http://www.census.gov/ces/
CCRDC on the web: http://www.ccrdc.ucla.edu/
Stanford RDC: https://iriss.stanford.edu/Securedata
RDCs as partnerships

- For researchers:
  Access to huge corpus of non-public use data

- For universities:
  Support for cutting-edge research
  Attract and keep data-intensive faculty

- For Census Bureau:
  Extends pool of expertise on substantive, methodological, and statistical issues
RDCs as partnerships

For researchers:
Access to huge corpus of non-public use data
Must address topics of interest to the Census Bureau in developed proposal
Must provide working papers and written annual updates
Must attempt to provide the benefits promised in proposal
Must financially support project in most cases
Must adhere to security requirements
RDCs as partnerships

- For universities:
  - Support for cutting-edge research
  - Attract and keep data-intensive faculty
    - Finances, provides and maintains secure facility
    - Funds Census Bureau administrators
    - Enters into legal contract delineating responsibilities
RDCs as partnerships

For Census Bureau:
Extends pool of expertise on substantive, methodological, and statistical issues

- Provides and supports administrator
- Provides feedback on proposals
- Provides security infrastructure, oversight
- Provides data access, software, disclosure avoidance review
- Network with other federal agencies to extend corpus of research data
Current RDCs

- Washington, DC (1983)
- Boston, Mass. (1994)
- Ann Arbor, Michigan (2002)
- Baruch (NYC, 2006) and Ithaca (Cornell, 2004)
- Minnesota (2010)
- Atlanta (2011)
- Texas – College Station (2012)
- Seattle (2012)
- Penn State (2014)
Why RDCs?
(Rationale for partnering)

- Perceptions of improper use could
  - Reduce response rates
  - Induce Congress to cut funding/programs

- Title 13 U.S.C protects confidentiality
  - Identifying microdata cannot be released
  - Only Census Employees/temporary staff can look at individually identifiable data
  - Projects must provide legitimate benefits to Census Bureau programs
Why use CCRDC data?

- Not available elsewhere
  - Establishment level business data
  - Linked data (e.g. worker-firm)

- More detail than anywhere else
  - Detailed geo-spatial variables
  - Virtually no top or bottom coding
  - Possible to link to other non-Census data

- Bigger Samples

- High Quality Sampling Frames

- Extensibility
Access and Disclosure Issues

- All researchers must be Census Bureau employees or have Special Sworn Status
  - Fingerprints, security forms, penalties
- Projects must show
  - Benefits to Bureau
  - Scientific Merit
  - Feasibility
  - Need for non-Public use Data
  - Minimal Risk of Disclosure
- All output goes through disclosure avoidance review (Interim and Final Outputs)
  - Statistical output: Yes
  - Tabular Output: No
Data in the RDC’s

- Demographic Data
- Economic Data
- Trade Data

Partner-Data
- Health Data (Hosted)
- Crime Victimization (Sponsored)
- Other
Go to web...... and skip next few slides
Key Demographic Surveys & Censuses

- Decennial Census of Population and Housing (1970-2010)
- American Community Survey (1996-2012)
- Current Population Survey (1967-2013) *
- American Housing Survey (1984-2012)
Decennial Census of Population and Housing

- Flagship Data Collection of Census Bureau
- Includes both universe and sample data

Public Use products include
- Summary Files
  - Pre-tabulated counts, multiple geographic summary levels
- Public Use Microdata
  - Individual/Household level data, PUMAs
Decennial Census

vs. Public Use Microdata

- Lowest level of geography available in the PUMS is an area that contains 100,000 people (PUMA)
- RDC version includes more detailed geographic information
  - current residence
  - place of work
  - prior place of residence
Decennial Census

- vs. Public Use Microdata
  - Larger sample size
    - 100% of short form respondents
    - One in six answered long form
      - PUMS has 5% of population
      - Improves analysis of small populations/sample sizes
  - Less top-coding
    - Continuous variables, such as income, are top-coded at a higher level
  - More detailed codes (race, education, multi-race, e.g. type of native American)
What can you do with it?

- Analyses of Segregation
- School Choice Preferences
- Impacts of Indian Casinos
- Patterns of Migration
- Impacts of Subsidized Childcare
- Residential and Work Enclaves
- Spatial Mismatch
- Impacts of Vietnam Draft

Look for yourself (CES Discussion Paper Series)
Center for Economic Studies (CES)

Research

CES conducts research in economics and other social sciences, and creates new public-use data from existing data.

- Researchers
- Research Topics
- CES Discussion Paper Series
- CES Seminar Series
- Restricted-Use Research Data
  - Economic data on business establishments and firms
  - Demographic data on individuals and households
  - Longitudinal Employee-Household Dynamics (LEHD) data

Data Products

CES creates innovative public-use data products from existing census, survey, and administrative data.

- Business Dynamics Statistics (BDS)
  Tabulations on establishments, firms, and employment with unique information on firm age and firm size
- LED Data Extraction Tool
  Beta tool providing easy access to the raw data products produced through the Local Employment Dynamics (LED) Partnership
- OnTheMap
  Online mapping and reporting application showing where the U.S. population and workforce live and work
- OnTheMap for Emergency Management
  Intuitive web-based interface for accessing U.S. population and workforce statistics, in real time, for areas being affected by natural disasters
- Quarterly Workforce Indicators (QWI)
  Workforce statistics by demography, geography, and industry for each state
- Synthetic Longitudinal Business Database (SynLED)
  Beta version of synthetic microdata on all U.S. establishments

Research Data Centers (RDCs)

RDCs are secure Census Bureau facilities. RDCs provide secure access to restricted-use microdata for statistical purposes only. Qualified researchers with approved projects can conduct research that benefits Census Bureau programs.

- Partner with leading research organizations
- Currently operate in 16 locations (green dots) with 2 new locations coming soon (yellow dots)
American Community Survey

- All surveys with all information collected on survey
  - Household or person-level data
  - Detailed geography (census block)
  - No top or bottom coding
- 1996 through 2012 currently available
- Can be linked to other data sources, where feasible and permissible
Confidential Versions of Your Favorite Public Use Datasets

- Survey of Income and Program Participation (SIPP)
- National Longitudinal Survey
- Current Population Survey (March)
- American Housing Survey
## Economic datasets: Economic Census

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census of Auxiliary Establishments (AUX)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2007</td>
</tr>
<tr>
<td>Census of Construction Industries (CCN)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1972–2007</td>
</tr>
<tr>
<td>Census of Finance, Insurance, and Real Estate (CFI)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1992–2007</td>
</tr>
<tr>
<td>Census of Island Areas – Puerto Rico (CIAPR)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1992–2007</td>
</tr>
<tr>
<td>Census of Mining (CMI)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1987–2007</td>
</tr>
<tr>
<td>Census of Retail Trade (CRT)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2007</td>
</tr>
<tr>
<td>Census of Services (CSR)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2007</td>
</tr>
<tr>
<td>Census of Transportation, Communications, and Utilities (CUT)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1967–2007</td>
</tr>
<tr>
<td>Census of Wholesale Trade (CWH)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2007</td>
</tr>
</tbody>
</table>
# Economic datasets: Firms

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
<th>Sponsor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Capital Expenditures Survey (ACES)</td>
<td>Annually</td>
<td>Firm</td>
<td>1993–2010</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Includes Information and Communication Technology (ICT) Supplement starting in 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual Retail Trade Survey (ARTS)</td>
<td>Annually</td>
<td>Firm</td>
<td>1999–2010</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Annual Wholesale Trade Survey (AWTS)</td>
<td>Annually</td>
<td>Firm</td>
<td>1995–2010</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Business Expenditures Survey (BES)</td>
<td>Every 5 Years</td>
<td>Firm</td>
<td>1992–2007</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Also known as Business Expenses Supplement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>See also <a href="#">SIRD</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaufman Firm Survey (KFS) [△]</td>
<td>Annual</td>
<td>Firm</td>
<td>2004–2010</td>
<td><a href="#">Ewing Marion Foundation</a></td>
</tr>
<tr>
<td>Must be used in conjunction with Census Bureau restricted-use microdata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturers' Shipments, Inventories, and Orders (M3)</td>
<td>Monthly</td>
<td>Firm</td>
<td>1992–2011</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Service Annual Survey (SAS)</td>
<td>Annually</td>
<td>Firm</td>
<td>1991–2010</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Survey of Business Owners (SBO)</td>
<td>Every 5 Years</td>
<td>Firm</td>
<td>1982–2007</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Survey of Industrial Research and Development (SIRD)</td>
<td>Annually (Discontinued)</td>
<td>Firm</td>
<td>1972–2007</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>See also <a href="#">BRDIS</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Economic datasets: Establishments

<table>
<thead>
<tr>
<th>Establishment Surveys and Data</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
<th>Sponsor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Survey of Manufactures (ASM)</td>
<td>Annually</td>
<td>Establishment</td>
<td>1973–2009</td>
<td>Census Bureau</td>
</tr>
<tr>
<td>Current Industrial Reports (CIR)</td>
<td>Annually</td>
<td>Establishment</td>
<td>1993–2008</td>
<td>Census Bureau</td>
</tr>
</tbody>
</table>
## Economic datasets: Transactions

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
<th>Sponsor</th>
</tr>
</thead>
</table>
### Economic datasets: BR

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compustat–SSEL Bridge (CSB)**</td>
<td>Annually</td>
<td>Establishment</td>
<td>1981–2005</td>
</tr>
<tr>
<td>Form 5500 Bridge File</td>
<td>Annually</td>
<td>Establishment</td>
<td>1992–2006</td>
</tr>
<tr>
<td><strong>Longitudinal Business Database (LBD)</strong></td>
<td>Annually</td>
<td>Establishment</td>
<td>1976–2011</td>
</tr>
<tr>
<td>Ownership Change Database (OCD)</td>
<td>Every 5 Years (Discontinued)</td>
<td>Establishment</td>
<td>1963, 1967–2002</td>
</tr>
</tbody>
</table>
Longitudinal Business Database

- Longitudinally linked Business Censuses
  - All non-farm establishments with paid employees in (almost) all industries
    - 24 million unique establishments
      - 8.5 million observations in 2011
    - Excludes airlines, agriculture, RR
Longitudinal Business Database

- LBD includes
  - Payroll
  - Employment
  - Ownership
  - Detailed geographic information
  - Industry at 6-digit NAICS (more detail in some cases)
  - Other variables available (e.g. sales) but coverage varies across sectors
Employer-Employee Linked Datasets

LEHD: Longitudinal Employer – Household Dynamics

- Quarterly data on employment and wages from state unemployment insurance agencies
  - Contains basic demographic data for all employees
  - Establishments linked to the LBD
  - Access requires state approvals
  - Multiple Datasets (independent access)
## LEHD Infrastructure

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Observation</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Register Bridge (BRB)</td>
<td>Establishment</td>
<td>1990–2008</td>
</tr>
<tr>
<td>Geocoded Address List (GAL)</td>
<td>Establishment</td>
<td>1990–2008</td>
</tr>
<tr>
<td>Quarterly Workforce Indicators (QWI)*</td>
<td>Establishment – Quarter</td>
<td>1990–2008</td>
</tr>
<tr>
<td>Unit-to-Worker (U2W)</td>
<td>Job (Person–Establishment)</td>
<td>1990–2008</td>
</tr>
</tbody>
</table>
Synthetic Data Server

The Synthetic Data Server (SDS) was set up to provide early access to new synthetic data products by the U.S. Census Bureau. These datasets are made available to interested researchers in a controlled environment, prior to a more generalized release.

At present, two datasets are made available on this server:

- SIPP Synthetic Beta (SSB) v4 and v5
- Synthetic LBD (SynLBD) v2

How to request access to the data is described on the next page. Access requests are reviewed for feasibility, but not otherwise restricted.

Once the data provider has signed off on the access, you will receive an account creation message from us with further instructions.

History

The server replaces the previous SIPP Synthetic Beta (SSB) server.

Funding acknowledgement

The SDS is funded through NSF grant SES-1042181.
Hosted Health Data

- We are now hosting research using confidential NCHS and AHRQ data in the CCRDC
- Rules for access and disclosure the same as those in their enclaves
  - http://www.meps.ahrq.gov
  - No requirement to demonstrate Census benefit.
- Long list of datasets – including NHIS, NHANES, NSFG, LSOA.…
New Data
National Center for Health Statistics

http://www.cdc.gov/rdc/

- **National Health and Nutrition Examination Survey (NHANES)**
  NHANES combines interviews and physical examinations to assess the health and nutritional status of adults and children in the United States.

- **National Health Care Surveys (NHCS)**
  A family of provider-based surveys that provide reliable information about health care providers, services, and patients. N

- **National Health Interview Survey (NHIS)**
  The NHIS collects data on a broad range of health topics through personal health interviews conducted in the home.

- **National Vital Statistics System (NVSS)**
  NVSS works with state vital registration systems to compile data on births, deaths, marriages, divorces, and fetal deaths.
Skip to end unless health questions
New Data

National Center for Health Statistics

National Health Care Surveys (NHCS)

- National Ambulatory Medical Care Survey (NAMCS)
- National Hospital Ambulatory Medical Care Survey (NHAMCS)
- National Survey of Ambulatory Surgery (NSAS)
- National Hospital Discharge Survey (NHDS)
- National Nursing Home Survey (NNHS)
- National Home and Hospice Care Survey (NHHCS)
- National Survey of Residential Care Facilities (NSRCF)
NHIS: Health Topics

- Demographics and SES
- Health status and disability
- Injury and poisonings
- Health insurance coverage
- Access to care
- Health services utilization
- Immunization
- Chronic conditions
- Health behaviors
- Height & Weight
New Data

National Center for Health Statistics

National Vital Statistics System (NVSS)

- Births (Natality)
- Deaths (Mortality)
- Fetal Death (Fetal Mortality)
- Linked Birth/Infant Death (Linked Fetal Mortality)
- Marriages and Divorces (Marital Status)
- National Maternal and Infant Health Survey (NMIHS)
- National Mortality Followback Survey (NMFS)
New Data
National Center for Health Statistics

Other NCHS Data Sources

- **Longitudinal Studies of Aging (LSOA)**
  The LSOA follows two cohorts of people 70 years of age and over to measure changes in their health, functional status, and health service use.

- **National Immunization Survey (NIS)**
  The NIS monitors immunization coverage of children between 19 and 35 months of age with a telephone survey and provider records.

- **National Survey of Family Growth (NSFG)**
  Collects information on family life, marriage and divorce, pregnancy, infertility, use of contraception, and men's and women's health.

- **State and Local Area Integrated Telephone Survey (SLAITS)**
  Collects health care information at the state and local levels to facilitate state and local area estimates to meet varied program and policy needs.

- **NCHS Data Linkage Activities (Linkage)**
  To enhance research value, NCHS links records from its population based surveys with other sources including Death Certificates (NDI), Medicare Claims (CMS), Social Security Benefits (SSA), Air Monitoring Data (EPA).
RDC Research Environment

“Thin Client” computing.
- Servers in Maryland, accessed via remote terminals
- Standard statistical software (SAS, Stata, Guass, Matlab, etc.)
- Standard Datasets kept on servers
- Other software/data coordinated by Administrator/CES staff

Secure Environment
- Restricted and monitored keycard access
- No Visitors
- No Laptops, internet
- Printing limited, RDC Administrator

Virtual RDC at Cornell (Synthetic Data, Zero Obs files)
http://www.vrdc.cornell.edu/news/
Fees

- $15,000 Standard Annual Project Fee
- Waivers available for UC Berkeley Faculty and Graduate Student Researchers (courtesy of D-Lab)
- Additional Fees for complex matching requiring CES staff
- Additional Fees for NCHS/AHRQ data – initial file creation and processing.
NEW

- Newly “recovered” historical household/population surveys and business/economic surveys.
  - Expedited access for evaluation purposes
  - Non-March CPS supplements, economic Censuses, ASMs…
    Write for details if you have questions.

- Kauffman Firm Survey Data Extension- Data Matching
  (http://www.kauffman.org/kfs/Travel-Grants-Program/Call-for-Proposals-%E2%80%93-KFS-Data-Extension-%E2%80%93-Data-Mat.aspx)

- CPS Food Security Supplement Data
Other

- CES Mentorship program
  (US citizens only)

- Virtual RDC

- INFO 7470: Spring Semester 2013 - archived
  http://www.vrdc.cornell.edu/info747/course_outline.html
Contact Information

- RDC web site: [http://www.ccrdc.ucla.edu/](http://www.ccrdc.ucla.edu/)
- email: [angela.andrus@census.gov](mailto:angela.andrus@census.gov)
- RDC phone: (510) 643-2262
- RDC administrator: Angela Andrus
- RDC executive director: Jon Stiles
- CES: [https://www.census.gov/ces/](https://www.census.gov/ces/)