

# 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

# CCRDC

## California Census Research Data Center Berkeley

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UC Berkeley, Stanford and UCLA **to enable qualified researchers** with approved projects to access confidential, unpublished Census Bureau data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)
- UCLA and Berkeley (1999) / (Stanford 2010) / USC (2014)  
UCI (2014)
- Duke (2000) / (RTI-2011)
- Chicago, Illinois (2002)
- 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) \*
- Survey of Income and Program Participation (1984-2008)
- American Housing Survey (1984-2012)
- National Longitudinal Survey (1966-1999)

# 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

## 1970, 1980, 1990 & 2000

### ■ 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 \(CES Discussion Paper Series\)](#)

You are here: [Census.gov](#) > [Business & Industry](#) > Center for Economic Studies

## Center for Economic Studies (CES)

[Main](#)
[Data Products](#)
[RDC Research Opportunities](#)
[Research Programs](#)
[Publications and Reports](#)
[FAQs](#)

### Who Are We?

The Center for Economic Studies (CES) partners with stakeholders within and outside the Census Bureau to improve measures of the economy and people of the United States through research and the development of innovative data products.

### Related Links

- [Annual Research Report](#)
- [Conferences](#)
- [Discussion Papers](#)
- [Dissertation Mentorship](#)
- [Seminars](#)

### Useful Links

- [Acronyms and Definitions](#)
- [Data Protection Policy](#)
- [Economist Positions](#)
- [LEHD](#)

### 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 Employer-Household Dynamics \(LEHD\) data](#)

### Data Products

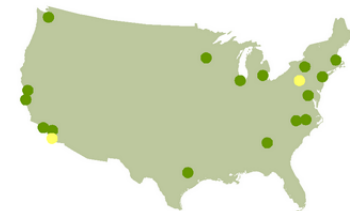
CES creates innovative [public-use data products](#) from existing census, survey, and administrative data.

- [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 \(SynLBD\)](#)  
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

## Restricted-Use Economic Census\* Microdata

Data Set	Frequency	Unit of Enumeration	Availability
<a href="#">Census of Auxiliary Establishments (AUX)</a>	Every 5 Years	Establishment	1977–2007
<a href="#">Census of Construction Industries (CCN)</a>	Every 5 Years	Establishment	1972–2007
<a href="#">Census of Finance, Insurance, and Real Estate (CFI)</a>	Every 5 Years	Establishment	1992–2007
<a href="#">Census of Island Areas – Puerto Rico (CIAPR)</a>	Every 5 Years	Establishment	1992–2007
<a href="#">Census of Manufactures (CMF)</a>	Every 5 Years	Establishment	1963, 1967–2007
<a href="#">Census of Mining (CMI)</a>	Every 5 Years	Establishment	1987–2007
<a href="#">Census of Retail Trade (CRT)</a>	Every 5 Years	Establishment	1977–2007
<a href="#">Census of Services (CSR)</a>	Every 5 Years	Establishment	1977–2007
<a href="#">Census of Transportation, Communications, and Utilities (CUT)</a>	Every 5 Years	Establishment	1987–2007
<a href="#">Census of Wholesale Trade (CWH)</a>	Every 5 Years	Establishment	1977–2007

# Economic datasets: Firms

## Restricted-Use Firm Microdata

Data Set	Frequency	Unit of Enumeration	Availability	Sponsor
<a href="#">Annual Capital Expenditures Survey (ACES)</a> <i>Includes Information and Communication Technology (ICT) Supplement starting in 2003</i>	Annually	Firm	1993–2010	Census Bureau
<a href="#">Annual Retail Trade Survey (ARTS)</a>	Annually	Firm	1999–2010	Census Bureau
<a href="#">Annual Wholesale Trade Survey (AWTS)</a>	Annually	Firm	1995–2010	Census Bureau
<a href="#">Business Expenditures Survey (BES)</a> <i>Also known as Business Expenses Supplement</i>	Every 5 Years	Firm	1992–2007	Census Bureau
<a href="#">Business Research &amp; Development and Innovation Survey (BRDIS)</a> <i>See also <a href="#">SIRD</a></i>	Annually	Firm	2008–2011	<a href="#">National Science Foundation</a>
Enterprise Summary Report (ESR)	Every 5 Years	Firm	1977–1992	Census Bureau
<a href="#">Exporter Database (EDB)</a>	Annually	Firm	1997, 2002, 2004–2010	<a href="#">International Trade Administration</a>
<a href="#">Kauffman Firm Survey (KFS)</a>  <i>Must be used in conjunction with Census Bureau restricted-use microdata</i>	Annual	Firm	2004–2010	<a href="#">Ewing Marion Kauffman Foundation</a> 
<a href="#">Manufacturers' Shipments, Inventories, and Orders (M3)</a>	Monthly	Firm	1992–2011	Census Bureau
<a href="#">Quarterly Financial Report (QFR)</a>	Annually	Firm	1977–2011	Census Bureau
<a href="#">Service Annual Survey (SAS)</a>	Annually	Firm	1991–2010	Census Bureau
<a href="#">Survey of Business Owners (SBO)</a>	Every 5 Years	Firm	1982–2007	Census Bureau
<a href="#">Survey of Industrial Research and Development (SIRD)</a> <i>See also <a href="#">BRDIS</a></i>	Annually (Discontinued)	Firm	1972–2007	<a href="#">National Science Foundation</a>



# Economic datasets: Establishments

Establishment Surveys and Data				
Data Set	Frequency	Unit of Enumeration	Availability	Sponsor
<a href="#">Annual Survey of Manufactures (ASM)</a>	Annually	Establishment	1973–2009	Census Bureau
<a href="#">Current Industrial Reports (CIR)</a>	Annually	Establishment	1993–2008	Census Bureau
<a href="#">Manufacturing Energy Consumption Survey (MECS)</a>	Various	Establishment	1985, 1988, 1991, 1994, 1998, 2002, 2006	Energy Information Administration (EIA)
<a href="#">Medical Expenditure Panel Survey – Insurance Component (MEPS–IC)</a>	Annually	Establishment	1996–2006, 2008–2010	Agency for Healthcare Research and Quality (AHRQ)
<a href="#">National Employer Survey (NES)</a>	Various	Establishment	1994, 1996, 1997, 1998	Department of Education (DOE)
<a href="#">Quarterly Survey of Plant Capacity Utilization (QPC)</a>	Quarterly	Establishment	1974–2007	Federal Reserve (FED)
<a href="#">Survey of Manufacturing Technology (SMT)</a>	Various	Establishment	1988, 1991, 1993	Census Bureau
<a href="#">Survey of Plant Capacity Utilization (PCU)</a>	Annually	Establishment	1974–2005	Federal Reserve (FED)
<a href="#">Survey of Pollution Abatement Costs and Expenditures (PACE)</a>	Various	Establishment	1974–1982, 1984–1986, 1988–1994, 1999, 2005	Environmental Protection Agency (EPA)

# Economic datasets: Transactions

Data Set	Frequency	Unit of Enumeration	Availability	Sponsor
<a href="#">Commodity Flow Survey (CFS)</a>	Various	Establishment	1993, 1997, 2002, 2007	<a href="#">Department of Transportation (DOT)</a>
<a href="#">Foreign Trade Data – Exports (EXP)</a>	Monthly	Transaction	1992–2011	Census Bureau
<a href="#">Foreign Trade Data – Imports (IMP)</a>	Monthly	Transaction	1992–2011	Census Bureau
<a href="#">Longitudinal Firm Trade Transactions Database (LFTTD)</a>	Monthly	Transaction	1992–2011	Census Bureau

# Economic datasets: BR

Data Set	Frequency	Unit of Enumeration	Availability
Compustat-SSEL Bridge (CSB)**	Annually	Establishment	1981– 2005
Form 5500 Bridge File	Annually	Establishment	1992– 2006
Integrated Longitudinal Business Database (ILBD)	Various	Establishment	1977, 1982, 1987, 1992, 1994–2010
<a href="#"><u>Longitudinal Business Database (LBD)</u></a>	Annually	Establishment	1976–2011
Ownership Change Database (OCD)	Every 5 Years (Discontinued)	Establishment	1963, 1967–2002
Standard Statistical Establishment List/Business Register (SSEL)	Annually	Establishment	1974–2012

# 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

Data Set	Unit of Observation	Years
Business Register Bridge (BRB)	Establishment	1990–2008
Employer Characteristics Files (ECF)	Establishment – Quarter	1989–2008
Employment History Files (EHF)	Job (Person–Firm)	1985–2008
Geocoded Address List (GAL)	Establishment	1990–2008
Individual Characteristics Files (ICF)	Person	1985–2008
Quarterly Workforce Indicators (QWI)*	Establishment – Quarter	1990–2008
Unit-to-Worker (U2W)	Job (Person–Establishment)	1990–2008

# Synthetic products



Cornell University

SEARCH CORNELL:

go

Pages People more options



## VirtualRDC @ Cornell

Synthetic Data Server, Econ Compute Cluster, and more

INFORMATION ▾ ECCO ▾ SYNTHETIC DATA SERVER ▾ DATA ▾ DOCUMENTATION ▾ HELP ▾ LABOR DYNAMICS INSTITUTE ▾

### Access to ECCO

Economics Compute Cluster (ECCO):

Account request procedures  
Usage instructions

### Synthetic Data Server

Info on Synthetic Data Server (SDS):

Account request procedures  
Codebooks on CED<sup>2</sup>AR

### Where is the Social Science Gateway?

The Social Science Gateway (SSG) grant has ended, please read here about

## 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.cdc.gov/nchs/r&d/rdc.htm>
  - <http://www.meps.ahrq.gov>
  - [http://www.ciser.cornell.edu/NYCRDC/documents/NCHS\\_RDC\\_Data.pdf](http://www.ciser.cornell.edu/NYCRDC/documents/NCHS_RDC_Data.pdf)
  - No requirement to demonstrate Census benefit.
  - Long list of datasets – including NHIS, NHANES, NSFG, LSOA....
  - [http://www.ciser.cornell.edu/NYCRDC/documents/NCHS\\_RDC\\_Data.pdf](http://www.ciser.cornell.edu/NYCRDC/documents/NCHS_RDC_Data.pdf)

# 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>)
- National Crime Victimization Survey, 2008 - 2012
- CPS Food Security Supplement Data

# Other

- CES Mentorship program  
(US citizens only)

- Virtual RDC

<http://www.vrdc.cornell.edu/news/synthetic-data-server/>

- INFO 7470: Spring Semester 2013 - archived

[http://www.vrdc.cornell.edu/info747/course\\_outline.html](http://www.vrdc.cornell.edu/info747/course_outline.html)

# Contact Information

- RDC web site: <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/>