Proposal Development in Federal Statistical Research Data Centers (RDCs)

Bethany S. DeSalvo, PhD
Federal Statistical Research Data Centers, Texas
Center for Economic Studies
United States Census Bureau

What are Research Data Centers (RDCs)?

- RDCs provide secure access to restricted data to qualified researchers with approved research projects.

- RDCs are restricted-access federal facilities, staffed by a Census Bureau employee, which meet all relevant security requirements.

- RDCs are a partnership between the local institution, the US Census Bureau and other federal statistical agencies.

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

RDCs as partnerships

For Academic Researchers:

- provides access to huge corpus of restricted data, supports cutting-edge research, and attracts and retains data-intensive faculty

For the Census Bureau:

- Extends pool of expertise on substantive, methodological, and statistical issues

Who can work in an RDC?

Researchers with an approved project, including:

- faculty and other researchers
- graduate students working with advisors
- foreign nationals with 3 of the past 5 years in the United States

Why Is Census Required to Restrict Microdata Access?

Titles 13 (Census); 26 (IRS) U.S.C.; CIPSEA protect confidentiality so that:

- the respondent cannot be identified
- only Census employees and temporary staff can access microdata
- access must potentially provide legitimate benefits to Census Bureau programs

The Nuts & Bolts of Doing Research in a RDC

- Research conducted on site

- Computing environment

- Restricted area with badge access

- No internet, phones or personal computers 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)

Demographic data: Restricted versus Public

- More geographic detail
- Additional variables
- More observations
- Variables "not" censored (income)
- Additional detail within variables

Data Available

Decennial Censuses

- 1950-2010 full count long and short form census data (when possible)

- Household and individual level demographic, socio-economic, program participation, education, household characteristics, etc

American Community Survey

- Yearly
- 1.5% of US population

Data Available

Current Population Survey Supplements

- Annual Social and Economic, Fertility, Food Security, School enrollment, Tobacco Use, Unbanked, Volunteer, Voter & Registration

American Housing Survey

- Core questions: Home condition, occupant characteristics, home improvements, housing costs, home values, characteristics of recent movers, etc

- Topical questions vary by year

Data Available

Survey of Income and Program Participation

- 2-4 year household panels; interviews ~every 4 months;; 14,000 to 52,000 households each wave
- Core: labor force, income dynamics, government transfers
- Topical modules vary

National Crime Victimization Survey

- Yearly; ~90,000 households
- Non-fatal and property crimes, reported and unreported; demographic information for respondent; demographic information of perpetrator, exp with CJ system

Data Available

National Longitudinal Mortality Study

- CPS-ASEC data linked to national death index
- CPS cohorts 1973-1998

National Longitudinal Survey (NLS)

- Original cohorts (1966, 1968)
- Labor market, demographic, and other data collected over 20 years
- ~5,000 respondents per cohort

Economic Data Advantages

Establishment and firm level characteristics

- Detailed industry and geography

Linking Data

- Consistent identifiers
- Business register
- Outside data

Economic Censuses

Data Set
Census of Auxiliaries (AUX)
Census of Construction Industries (CCN)
Census of Finance, Insurance, Real Estate (CFI)
Census of Manufacturers (CMF)
Census of Mining (CMI)
Census of Retail Trade (CRT)
Census of Services (CSR)
Census of Transportation, Communications, Utilities (CUT)
Census of Wholesale Trade (CWH)

Establishment Surveys

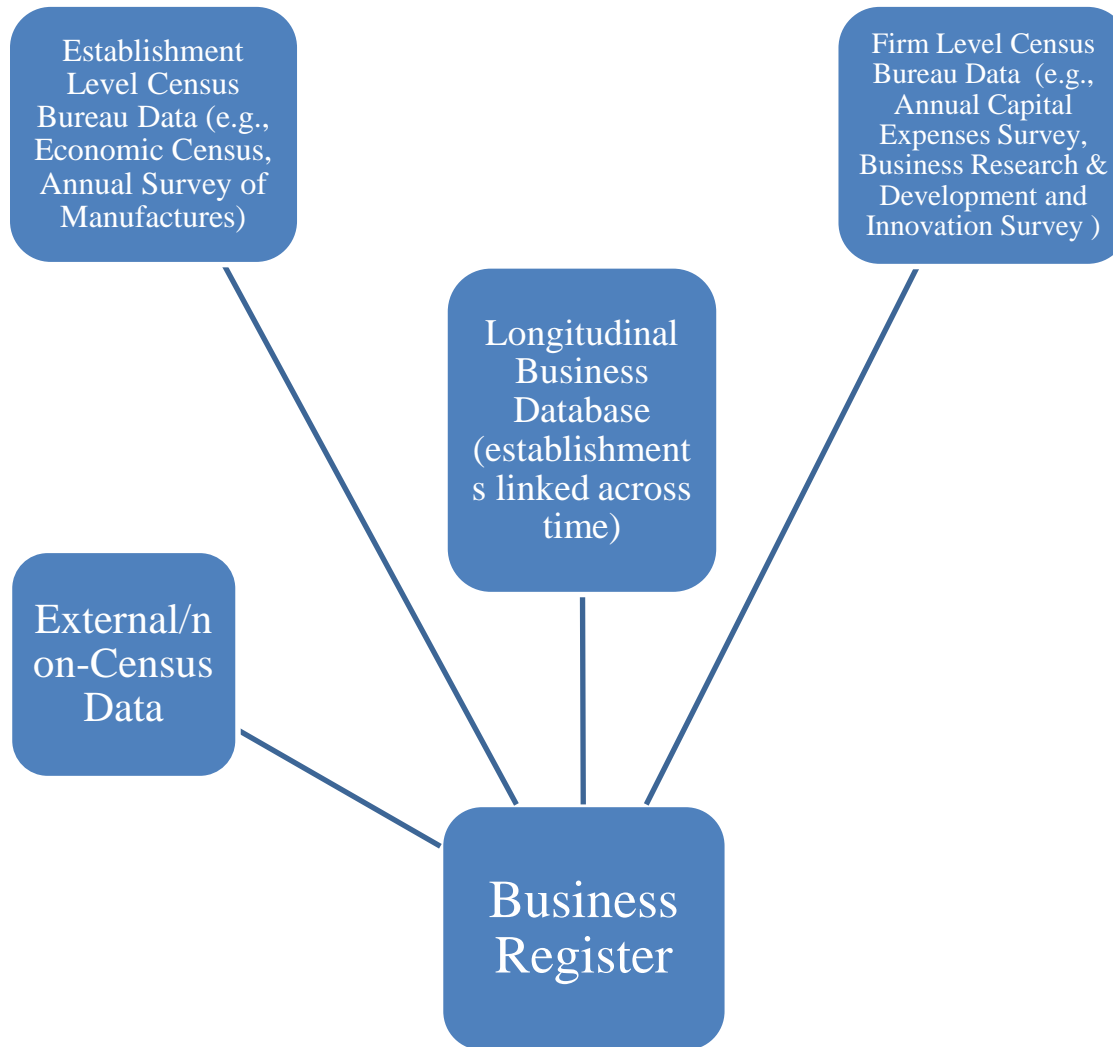| Data Set |
|---|
| Annual Survey of Manufacturers (ASM) |
| Current Industrial Reports (CIR) |
| Manufacturing Energy Consumption Survey (MECS) |
| Medical Expenditure Panel Survey – Insurance Component (MEPS-IC) |
| National Employer Survey (NES) |
| Quarterly Survey of Plant Capacity Utilization (QPC) |
| Survey of Manufacturing Technology (SMT) |
| Survey of Plant Capacity Utilization (PCU) |
| Survey of Pollution Abatement Costs and Expenditures (PACE) |

Firm Surveys

Data Set
Annual Capital Expenditures Survey (ACES)
Annual Retail Trade Survey (ARTS)
Business Expenditures Survey (BES)
Business Research & Development and Innovation Survey (BRDIS)
Enterprise Summary Report (ESR)
Exporter Database (EDB)
Quarterly Financial Report (QFR)
Service Annual Survey (SAS)
Survey of Business Owners (SBO)
Survey of Industrial Research and Development (SIRD)

Business Register Data

Data Set
Compustat-SSEL Bridge (CSB)
Form 5500 Bridge File
Integrated Longitudinal Business Database (ILBD)
Longitudinal Business Database (LBD)
Ownership Change Database (OCD)
Standard Statistical Establishment List / Business Register (SSEL)

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Transactions Data

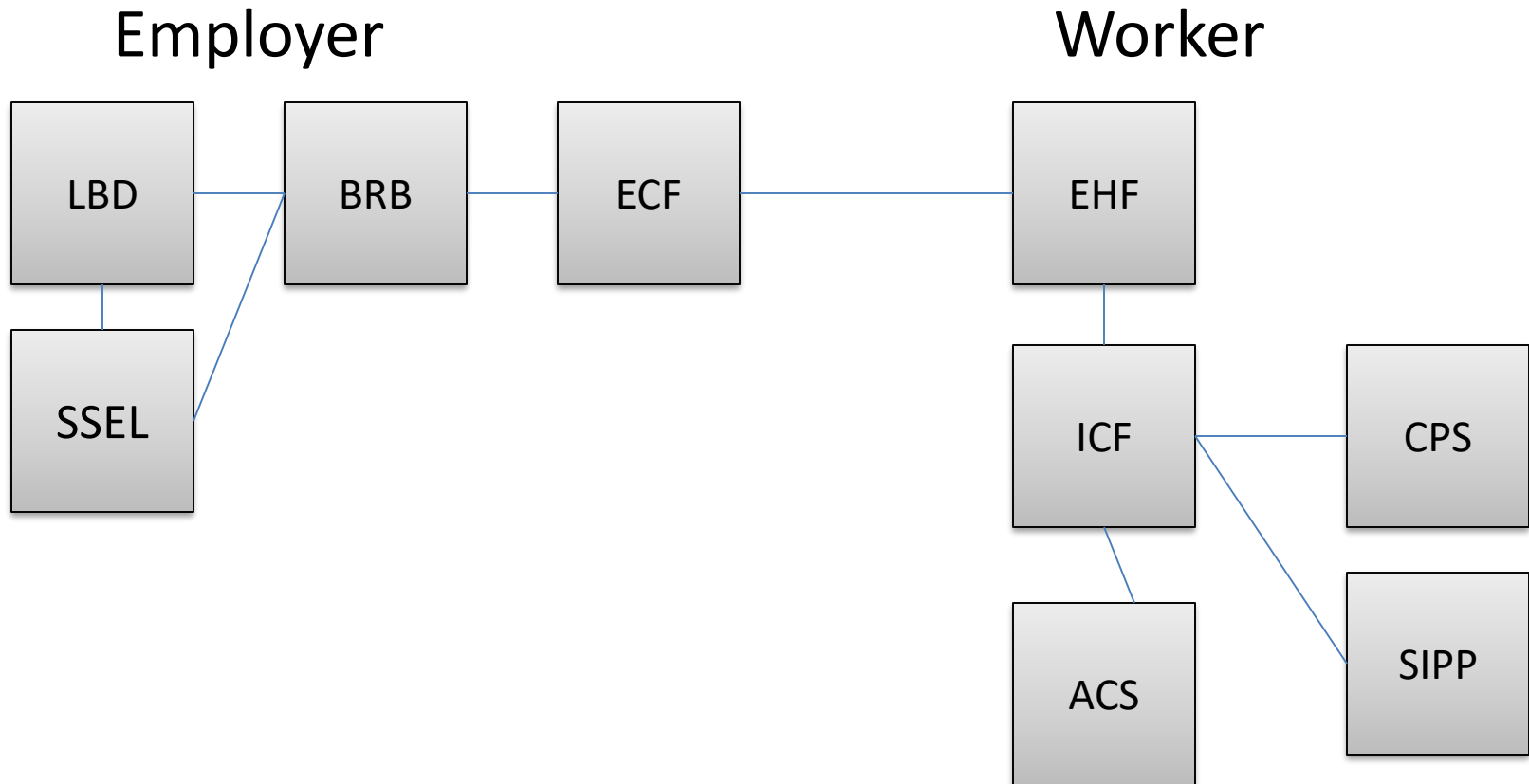| Data Set |
| --- |
| Commodity Flow Survey (CFS) |
| Foreign Trade Data - Export (EXP) |
| Foreign Trade Data - Import (IMP) |
| Longitudinal Foreign Trade Transactions Data (LFTTD) |

Linked Employer Household Dynamics (LEHD)

LEHD data combine administrative data from states' Unemployment Insurance systems with Census Bureau data.

1. **Workers:** Employer history and quarterly wages, Individual characteristics (sex, age, race), Point in time residence and place of birth

2. **Employers:** Industry, employment, total payroll, location

3. Linkages between workers and employers

4. Links to other Census data: Virtually any RDC data on businesses; SIPP; CPS March supplement; ACS

Longitudinal Employer-Household Dynamics (LEHD)

Recovered data

- Tapes from Unisys mainframe were recovered, providing data back to 1953 on all sectors of the economy
    - [Newly Recovered Microdata on U.S. Manufacturing Plants from the 1950s and 1960s: Some Early Glimpses." (3.7 MB) CES Discussion Paper CES-WP-11-29.](#)
- Recovered demographic data
    - CPS data back to 1962
    - Income Surveys Development Program data (old SIPP)
- Others

Important Web Sites

- Census Bureau Data: Center for Economic Studies

    - http://www.census.gov/ces/

- NCHS Research Data Center

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

- AHRQ

    - https://meps.ahrq.gov/data_stats/onsite_datacenter.jsp

Background Check

- Off-line paperwork and documentation
- On-line trainings and certifications
- Background check
  - Submitted online and followed with interview
    - Residential history
    - Foreign travel
    - Education and employment history
    - References
- Fingerprinting

Special Sworn Status

- SSS is authorized by Title 13 U.S.C. 23 (c) "to assist the Bureau of the Census in performing the work authorized by this title."

- The Census Bureau may provide SSS to an individual
  - **When an individual has expertise or specialized knowledge that can contribute to the accomplishment of Census Bureau projects or activities or engages in a joint project with the Census Bureau;**
  - When an individual is employed by an agency/organization performing a service for the Census Bureau under contract or providing information to the Census Bureau for statistical purposes;
  - When Federal law requires an individual to audit, inspect, or investigate Census Bureau activities.

Writing the proposal: perspective

- The perspective of your proposal is driven toward the predominant purpose or "the Census Bureau benefit."
- Your audience includes mostly data experts
- Your proposal is a request for data showing your project:
    - has 2 possible benefits to the Census Bureau
    - is feasible
    - emphasizes statistical models vs. tabular output
    - has scientific merit
    - clearly needs restricted use data
    - falls within the Census Bureau mandate
    - indicates an understanding of the appropriate disclosure avoidance protections

Proposal Package

1. Abstract

2. Proposal Description

3. Benefit to the Census Bureau (Predominate Purpose Statement/PPS)

Description

Sections:

- Introduction
- Background / Literature Review
- Data & Methods
- Output / Disclosure Risk
- Timeline / Project Duration
- Conclusion

Output / Disclosure Avoidance Review

- No output can leave the RDC without review
- Clear understanding of samples
  - No individual person or business can be identifiable in release
- Performed by Administrator and the Center for Disclosure Avoidance Review
  - 2-3 weeks (in general)
  - Intermediate output discouraged
  - Descriptive results may be problematic
  - Focus on statistical data for release

Timeline

Getting approved:

- Proposal development
- Review from Census
- Review from sponsoring agencies
- Background Investigation

Active Project:

- Data work
- Deliver the *Benefits to the Census Bureau*
- Disclose statistical output
- Build in time for R&R's

Benefits to the Census Bureau

- Predominant Purpose Statement
- Not a pro forma requirement
  - Legal basis on which researchers are allowed access to restricted use data
- Must provide 2 benefits

Benefits, examples

3. Developing means of increasing the utility of Census Bureau data for analyzing public programs, public policy, and/or demographic, economic, or social conditions

5. Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate

11. Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5

Thank you.

Bethany DeSalvo, PhD
Federal Statistical Research Data Center, Texas
Center for Economic Studies
US Census Bureau
Bethany.DeSalvo@census.gov
979-845-5618