

**Texas
Federal
Statistical
Research
Data
Center
(TXRDC)**

**Opportunities for
Research Using
Restricted Data**

**Faculty & Researcher
Session**

Mary Campbell, PhD

Director, TXRDC

Texas A&M University

Karin Johnson, PhD

Administrator,
TXRDC

US Census Bureau

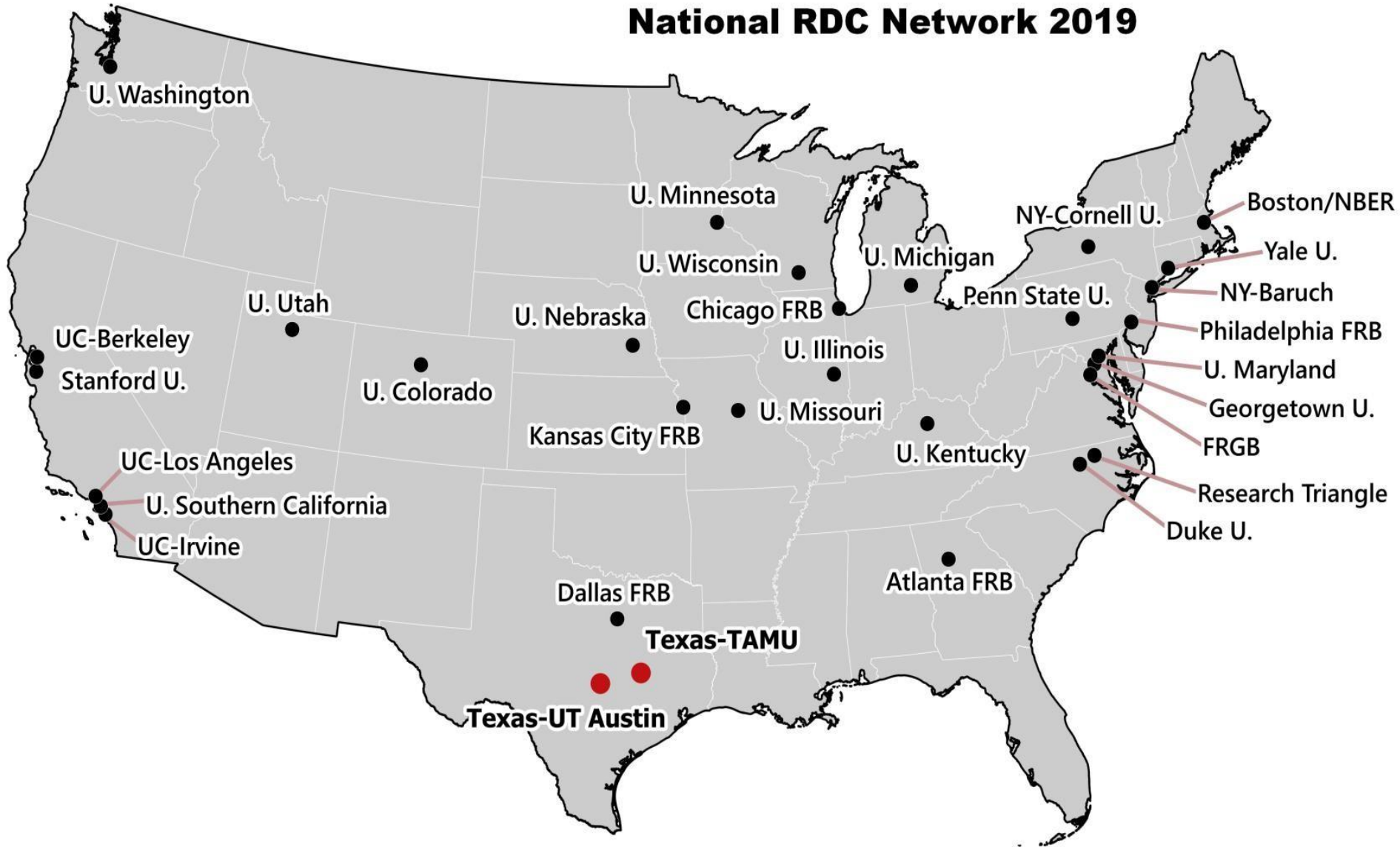
Federal Statistical Research Data Centers (FSRDCs)

- FSRDCs (or RDCs) are research facilities
- Collaboration among the Census Bureau, other federal statistical agencies, and institutions to provide researchers with approved projects access to restricted-use data

What do RDCs offer?

- Secure computing lab
- Joint Statistical Partnerships (JSP)
- Managed by the Center for Enterprise Dissemination (CED) at Census Headquarters
- Locally run by an on-site Census Bureau employee, the RDC Administrator

National RDC Network 2019



What is the Texas Consortium?

- RDCs are similar to research libraries and labs
- These universities are members, providing researcher access to valuable resources
- **We welcome other institutions and researchers!**

Administrative Home



Texas RDC Consortium Institutions



The TXRDC

Located in the D.L. Houston building
Moving to the Teague Research building
in Spring 2021!

Why we need RDCs

- Federal laws and regulations protect data confidentiality
- Federal employees and contractors access restricted data only under strict guidelines

→ **National FSRDC network is important for research analyzing restricted statistical data to benefit society**

Advantages of RDC Access to Researchers

- Restricted microdata
- Key variables not in public versions of datasets, or detailed versions of variables
- Full counts and/or larger samples
- Original responses
- Ability to link data

Research Communities RDCs Serve Best

Serve Well...

- Discipline-based substantive research
- Research plans with long time horizons

Serve Less Well ...

- Research with short time horizons
- Research producing special tabulations and maps
- Projects producing data products for public distribution

Available Data

- **Census Bureau Data**
 - **Economic Data**
 - Establishment, firm, transaction level
 - **Demographic Data**
 - household or individual level
 - **Combined Employer/Employee Data**
 - Longitudinal Employer-Household Dynamics (LEHD)
- **Health Data**
 - Agency for Healthcare Research and Quality (AHRQ)
 - National Center for Health Statistics (NCHS)
- **Other Agencies joining**

RDC Economic Data Advantages

- Restricted microdata available
 - Internal data at establishment and firm level
 - Detailed industry and geography
- Linking Data
 - Consistent identifiers
 - Business register
 - External data

Economic Census

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)

Transactions Data

Data Set

Commodity Flow Survey (CFS)

Foreign Trade Data - Export (EXP)

Foreign Trade Data - Import (IMP)

Longitudinal Foreign Trade Transactions Data (LFTTD)

Active Economic TXRDC Projects

- R&D expenditures, before/after policy
 - Business Research & Development and Innovation Survey, the Annual Survey of Manufactures, Mining, Construction Industries, LBD and Survey of Business Owners
- Mortgages and career decisions
 - LEHD, Annual Survey of Manufactures, Mining, Construction Industries

Demographic Data Available

- Individual and household data
- Restricted-use RDC data
 - More detailed geography
 - Less top-coding/censoring
 - Additional variables
 - Occupation codes
 - Possibly bigger samples
 - Metadata

More Detailed Geography

Data Set	Geography
Decennial Census	Block
American Community Survey (ACS)	Block
Survey of Income and Program Participation (SIPP)	Tract
Current Population Survey* (CPS) – March Supplement & Food Supplement	Tract
American Housing Survey (AHS)	Tract
National Longitudinal Survey (NLS) – Young/Mature Women	Lat/Lon Block Group
National Longitudinal Survey (NLS) – Young/Old Men	County
National Longitudinal Mortality Study (NLMS)	County

Active Demographic TXRDC Projects

- Internal, international migration
 - Decennial censuses, ACSs
- Veteran-Owned Small Businesses
 - Decennial censuses, ACSs, LBD, Survey of Business Owners, etc.

Health Data Partners

- Agency for Healthcare Research and Quality (AHRQ)
- National Center for Health Statistics (NCHS)



Agency for Healthcare
Research and Quality



Centers for Disease
Control and Prevention
National Center for
Health Statistics

Agency for Healthcare Research and Quality (AHRQ)

Data include:

- Fully specified ICD-9 medical condition codes
- Fully specified industry and occupation codes
- 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

AHRQ Data Sets

- Medical Expenditure Panel Survey – Household Component (MEPS-HC)
- Household Component-Insurance Component linked file
- Nursing Home Component
- 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)

National Center for Health Statistics (NCHS)

Types of Restricted Variables

- Geographic Variables
- Linked Data Products
- 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

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)

Active Health TXRDC Projects

- Foreign-born health outcomes
 - NHIS
- Drinking behaviors and mortality risks

New Datasets Frequently Become Available

See Census FSRDC web links for compilations

- [FSRDC Available Data](#)
- [Participating Agencies Data Sets](#)
- [Census CES Restricted Data Notes](#)

National Center for Health Statistics (NCHS)

- [NCHS Restricted Data Sets & Information](#)

Agency for Healthcare Research and Quality (AHRQ)

- [AHRQ Restricted Data Sets & Information](#)

Consult with the RDC Administrator to inquire about available datasets



Texas Research Data Center

COLLEGE OF LIBERAL ARTS

ABOUT TXRDC

FUNDING

PROPOSALS

EV

🏠 > Datasets

Datasets

A great many restricted federal datasets can currently be accessed in RDCs. These include datasets from the Census, the National Center for Health Statistics, the Agency for Healthcare Research and Quality, as well as other important federal datasets relevant to employment, income, wealth, taxes, health, disability, retirement, and establishments, to name just a few. More datasets are becoming available on a regular basis.

To clarify the availability of specific datasets and variables contact our director, Mary Campbell (m-campbell@tamu.edu).

Available Datasets in RDCs

The links below provide summary listings of datasets that are available in RDCs and links to other sources of information on the topic. Please be aware this is not a comprehensive list of available data and there may be other data sources that are available

- [NCHS Page Listing Available Datasets](#) ↗
- [CES Page Listing Available Datasets](#) ↗
- [AHRQ Page Listing Available Datasets](#) ↗
- [Business and Economic Data Sets in the Research Data Centers](#)

<https://liberalarts.tamu.edu/txrdc/datasets/>

Practical Considerations

- Project Approval by Agency and Researcher Security Approval
 - Agency review and approval
 - Demonstrated “benefits” to federal agencies
 - Special Sworn Status (3-year residency requirement for non-citizens)

Practical Considerations

- The RDC Computing Environment – Software and Hardware
 - Computing is conducted in-person in the secure lab
 - Census Bureau workstation hardware/software limitations

Practical Considerations

- Releasing Research Results Involves a Review Process
 - Access, analysis, and disclosure of results follow strict protocols
 - Sworn confidentiality for life

Contact us if you have an idea! We will:

- Brainstorm with you
- Help you to identify datasets and variables
- Assess the project's:
 - Feasibility and timeline
 - Benefits to the Census Bureau (if applicable)
 - Results disclosure
- Review your entire proposal package

Also....

Mary can discuss with you:

- Project goals
- External funding, graduate assistants, etc.
- RDC proposal development facilitation
- Consortium policies

Six Phases of RDC Projects

1. Proposal drafting
2. Submit for Agency Review
3. If approved, apply for Special Sworn Status
4. Begin working in the RDC
5. Prepare memo and output for Disclosure Review
6. Complete promised benefits, produce products for Center for Economic Studies Working Paper Series

Census Bureau Proposal Package

1. Project Description (with a template)
2. Benefit to the Census Bureau
("Predominate Purpose Statement" (PPS))
3. Abstract

Project Description

- State the importance of using restricted data!
- Sections (template):
 - Introduction
 - Literature Review
 - Methodology
 - Data Sources
 - Output/Disclosure Risk
 - Timeline/Project Duration
 - Conclusion

Project Description: Methods

- Describe empirical methodology, including equations
- Clarify the relationship between your specifications and the data
 - What are your hypotheses of interest?
 - What are your coefficients of interest and how do they relate to your stated hypotheses?
 - Is it a longitudinal or cross-sectional analysis?
 - How will you be linking data together?
 - What is your unit of analysis?
 - At what level are you estimating your equations (individual, household, etc?)

Project Description: Data Sources

You will provide:

- A detailed description of your data sources.
 - Public, restricted use (Census or other)
- A bulleted list of data sources with years.
- Information on how the data sources be linked (i.e. SIC code, county, etc.).
- A list of the software you wish to use.

Project Description: Output/Disclosure Risk

No individual person or business will be identifiable at release.

Describe:

- What you wish to take out of the lab (i.e., regression coefficients, standard errors, etc.)
- Different samples you will disclose

Limited descriptive statistics

- Intermediate output discouraged

Project Description: Timeline

- List expected major milestones, 3 to 5 years
 - When will you complete matching of datasets, construction of extracts, etc.
 - How do you expect the project to unfold
 - When will you request disclosure(s)
- Extensions often not granted
 - May be required to provide additional benefit

Benefits to the Census Bureau

- Predominant Purpose Statement (PPS)
- Not a *pro forma* requirement
 - Legal basis on which researchers are allowed access to restricted use data
- List each criterion verbatim with explanation
- Must provide 2 benefits
- **I will assist you with these**

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

Approval Process: 3 steps

- Step 1: RDC approval
- Step 2: Census Bureau (CED) approval
- Step 3: Sponsoring agency approval (if applicable)

Proposal Review Process: 5 standards

1. Benefit to Census Bureau programs (Title 13, U.S.C)
2. Scientific merit (soundness)
3. Clear need for restricted use data
4. Feasibility
5. Risk of disclosure / Falls within Census Bureau mandate

Approval Process Timeline*

- Title 13 datasets: about 120 days
- Sponsoring agency timing: varies
- Once approved, background check is initiated (possibly another 120 days)

* may change based on many factors

How you can speed up the process

- Adhere closely to all practices and procedures before proposal submission
- **Work closely with Karin on proposal development and on any requested revisions or clarifications**
- Provide the terms of use for any datasets you wish to bring to the lab
- Process Special Sworn Status (SSS) paperwork quickly once project is CED approved

Contact Us!

Mary Campbell, Director, Texas A&M University
m-campbell@tamu.edu

Karin Johnson, Administrator, U.S. Census Bureau
karin.johnson@census.gov

TXRDC website: <https://txrdc.tamu.edu>

FSRDC website: <https://census.gov/fsrdc>