

Research Opportunities Using Restricted Data: Researcher Session

Karin Johnson, PhD

Administrator, TXRDC

US Census Bureau

Mary E. Campbell, PhD

Director, TXRDC

Texas A&M University

Kelly Raley, PhD

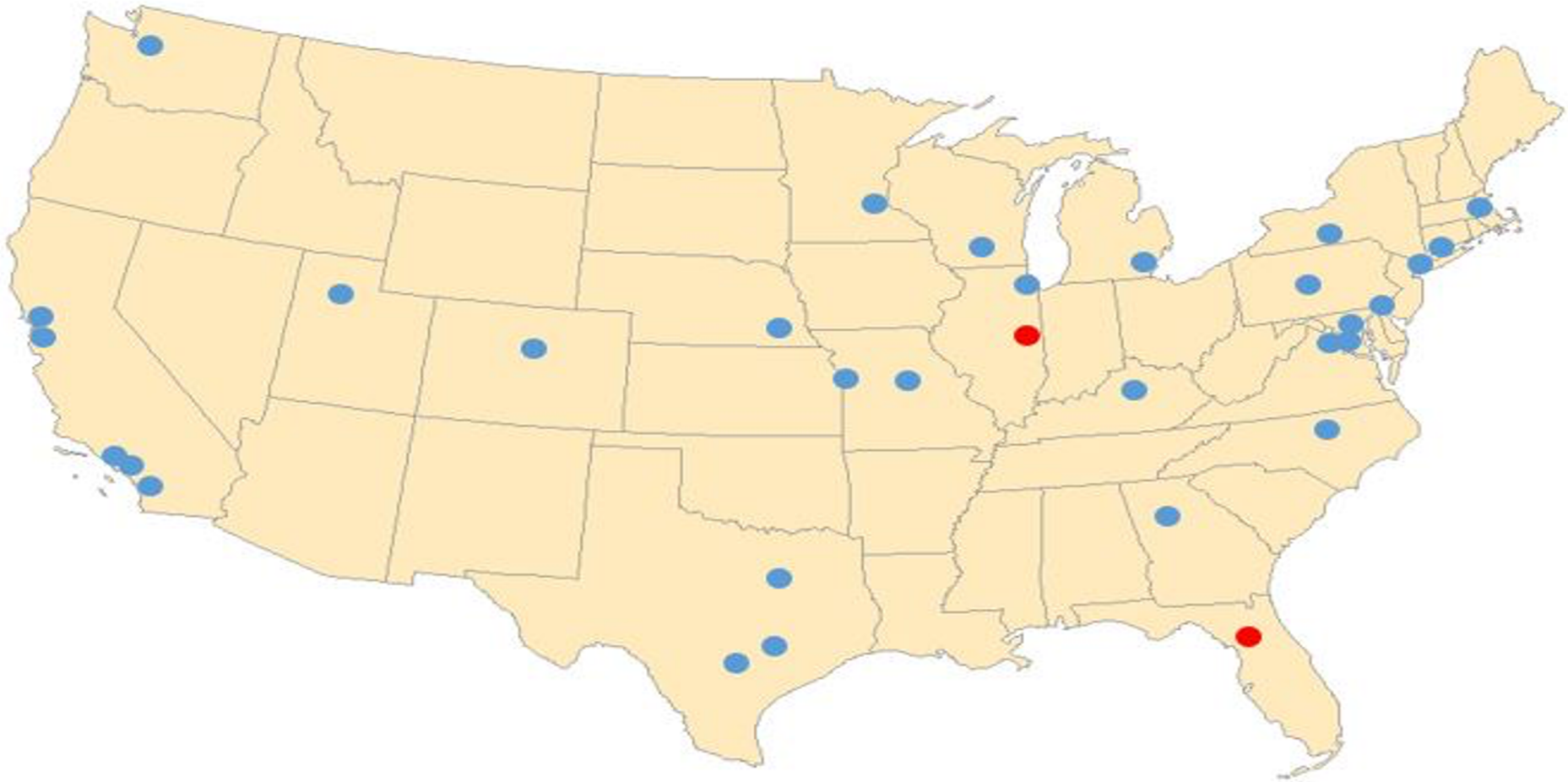
Director, UTRDC

University Texas Austin

Federal Statistical Research Data Centers (FSRDCs)

- FSRDCs are **secure research computing labs**
- Collaboration among the **Census Bureau**, other **federal statistical agencies**, and **institutions** to provide researchers with approved projects access to restricted-use data
- Federal laws and regulations protect data confidentiality
- Federal employees and contractors access restricted data only under strict guidelines

National RDC Network 2021



What is the Texas Consortium?

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

Administrative Home



Texas RDC Consortium Institutions



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: Examples

- Census Bureau Data
 - Demographic Data
 - Household or individual level
 - Economic Data
 - Establishment, firm, transaction 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 (e.g, BEA, BLS)

RDC Demographic Data: Advantages

- Individual and household data
- Restricted-use microdata
 - 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, Harmonized Decennial censuses, ACSs*
- Veteran-Owned Small Businesses
 - *Decennial censuses, ACSs, LBD, Survey of Business Owners, etc.*

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 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)

Business Registrar 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)

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)

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*

Health Data Partners

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



AHRQ Data Examples

- 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)

NCHS Data Examples

- National Health and Nutrition Examination Survey (NHANES)
- National Health Care Surveys
- National Health Interview Survey (NHIS)
- National Survey of Family Growth (NSFG)
- NCHS Data Linkage Activities
- National Vital Statistics System (NVSS)

Active Health TXRDC Projects

- Foreign-born health outcomes
 - *NHIS*
- Drinking behaviors and mortality risks
 - *Coming soon!*

Where to scout data

ResearchDataGov

[Search datasets and variables](#)

See Census FSRDC web links for compilations

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



Consult with the RDC Administrator to inquire about available datasets



Texas Research Data Center

COLLEGE OF LIBERAL ARTS

[ABOUT TXRDC](#)

[FUNDING](#)

[PROPOSALS](#)

[EVENTS](#)

[Home](#) > [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/>

Six Phases of RDC Projects

1. Proposal drafting
2. Submit for Agency Review(s)
3. If approved, apply for Special Sworn Status (moderate security clearance)
4. Begin working in the RDC
5. Prepare memos 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 (using a template)
2. Benefits 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 (restricted and, if applicable, public-use)
 - Expected Output/Disclosure Risk
 - Timeline/Project Duration (3 to 5 years)
 - Conclusion

Benefits to the Census Bureau

- “Predominant Purpose Statement” (PPS)
- NOT a courtesy requirement
 - **Legal basis on which researchers are allowed access to restricted use data**
- Lists each criterion verbatim with explanation
- Must provide 2 benefits
- Karin 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

Proposal Review Process: 5 standards

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

Approval Process Timeline*

About 6 months to 1 year

- Census Bureau (Title 13) approval:
 - No more than 3 months
- Sponsoring agency approval (e.g., IRS, SSA), if applicable:
 - Varies
- Once approved, security clearance application is initiated:
 - Between 3 and 6 months
 - Possibly longer for Foreign National residents (3-year residency requirement for non-citizens)

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 Census approved

Practical Considerations

The RDC Computing Environment

- Census Bureau workstation hardware/software limitations
- Computing is conducted in-person in the secure lab
 - ***Only Census (Title 13) projects may apply for virtual access***

Practical Considerations

Releasing Research Results Involves a Disclosure Risk Review Process

- Releasing results follows strict protocols
- About 2-3 weeks; up to 12 weeks
- 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 and Kelly can discuss with you:
 - Project goals
 - External funding, graduate assistants, etc.
 - RDC proposal development facilitation
 - Consortium policies

Contact Us!

Mary E. Campbell, Director, Texas A&M University m-campbell@tamu.edu
Kelly Raley, Director, University Texas Austin kelly.raley@austin.utexas.edu

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

Links:

[FSRDC website](#)

[TXRDC website](#)

[UTRDC website](#)