

Federal Statistical Research Data Centers (FSRDC)

Samuel R. Bondurant

Dallas-Fort Worth Federal Statistical Research Data Center
Center for Enterprise Dissemination
US Bureau of the Census

Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau.



What is a Research Data Center?

**Federal Statistical Research
Data Centers (FSRDCs)**

United States[®]
Census
Bureau

What are Federal Statistical Research Data Centers (RDCs)?

- Partnerships between US Census Bureau, other federal statistical agencies and research institutions.
- Secure computing labs where qualified researchers conduct approved statistical analysis on restricted-access data.
- These data are collected by various government agencies (Census Bureau, NCHS, BEA, BLS, SSA, NSF, etc.).
- Managed by the Census Bureau.

UT SOUTHWESTERN
MEDICAL CENTER



UNT
UNIVERSITY OF NORTH TEXAS

T TEXAS TECH
UNIVERSITY.



A UNIVERSITY OF
TEXAS
ARLINGTON

UT DALLAS



The UNIVERSITY of OKLAHOMA

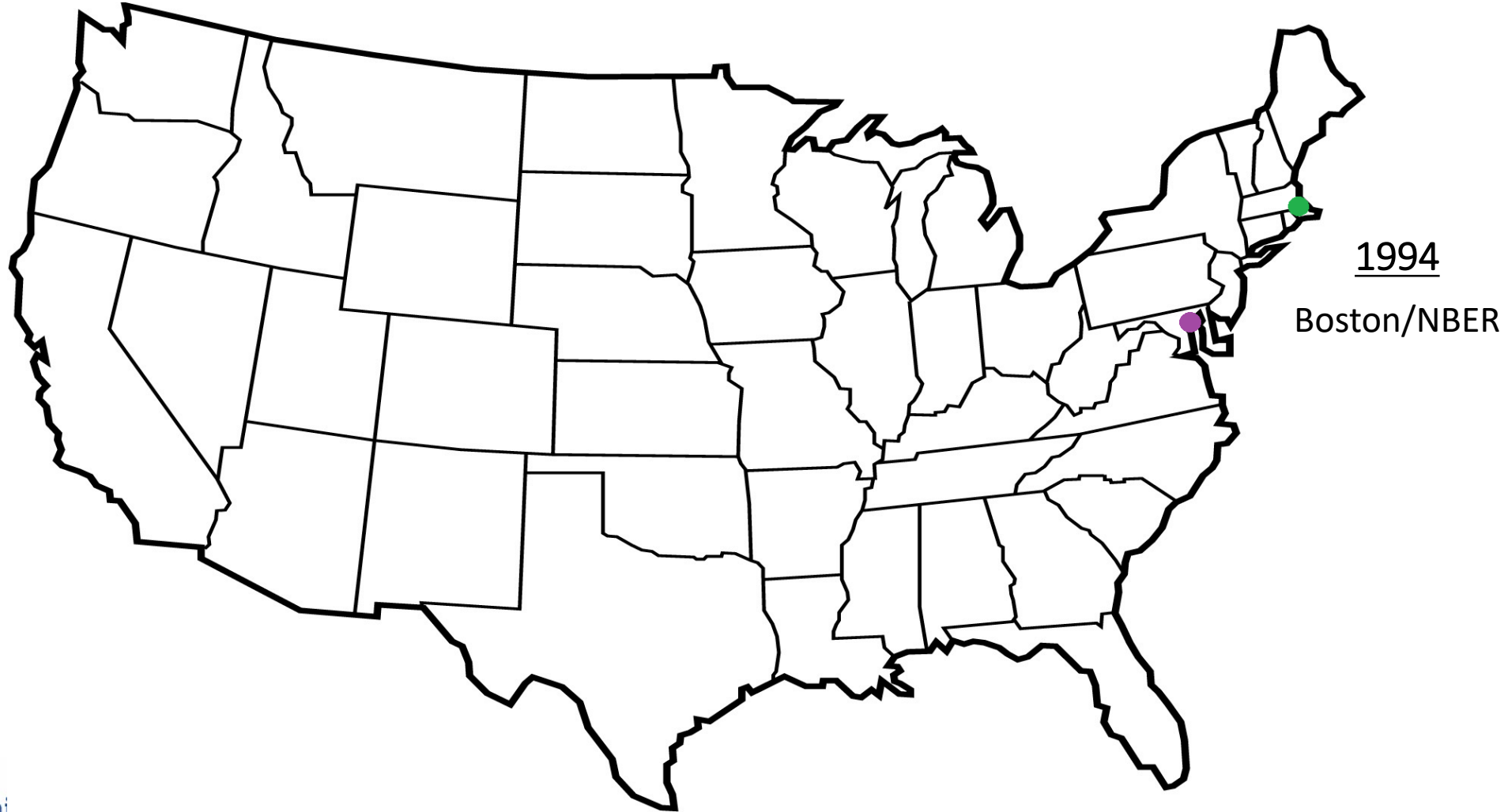
 **SMU**

United States®
Census
Bureau

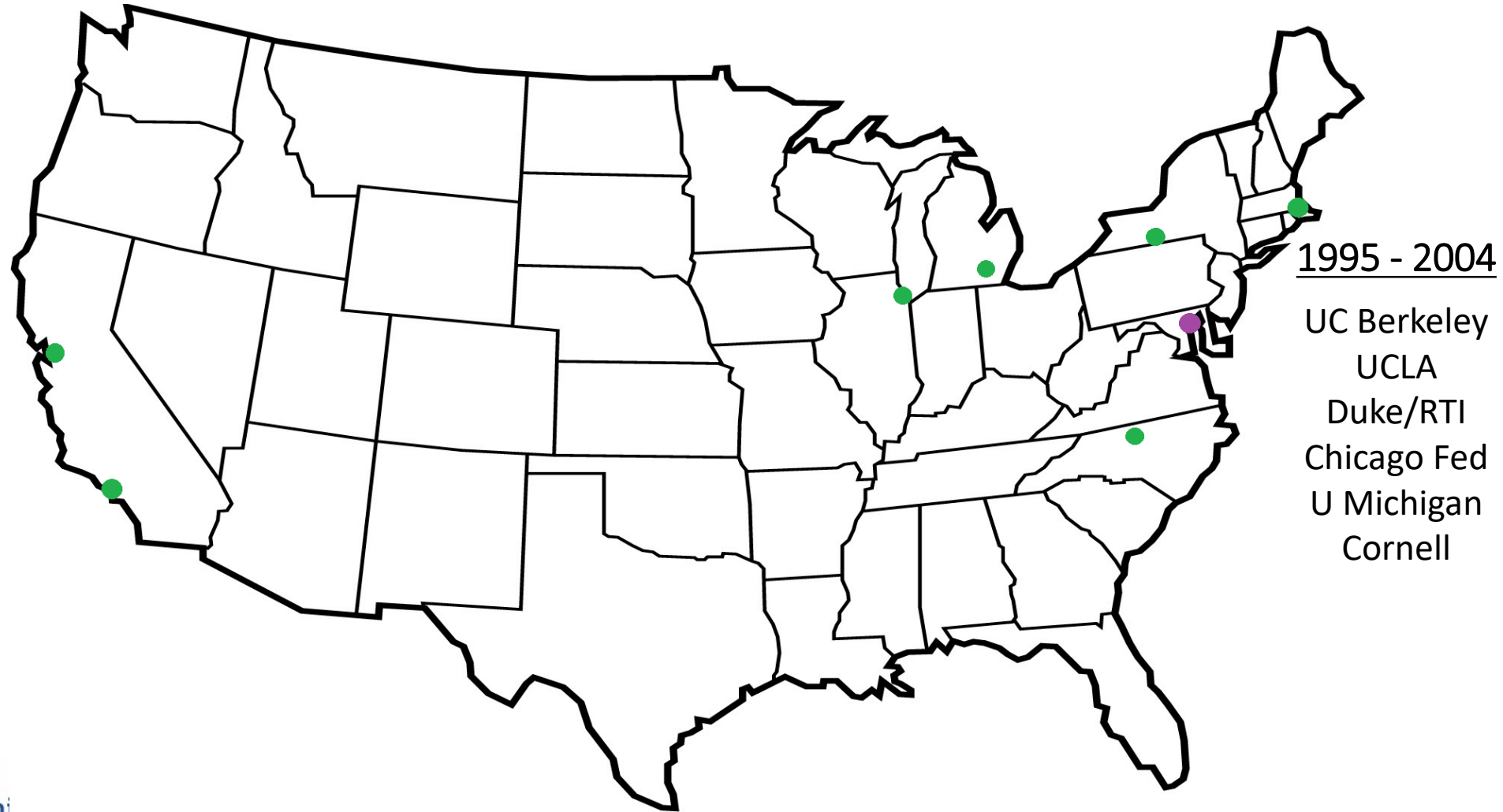
Research Data Centers



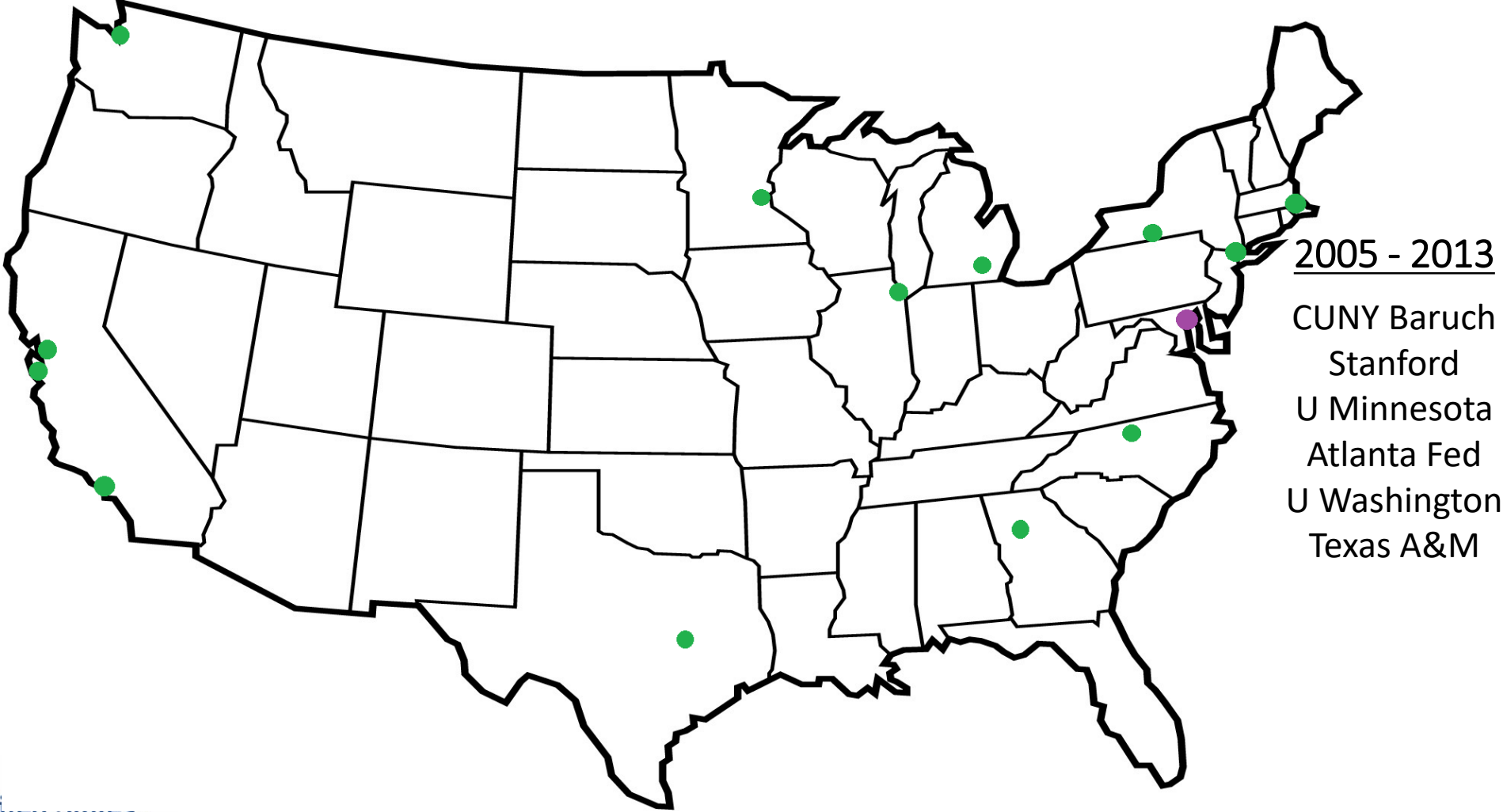
Research Data Centers



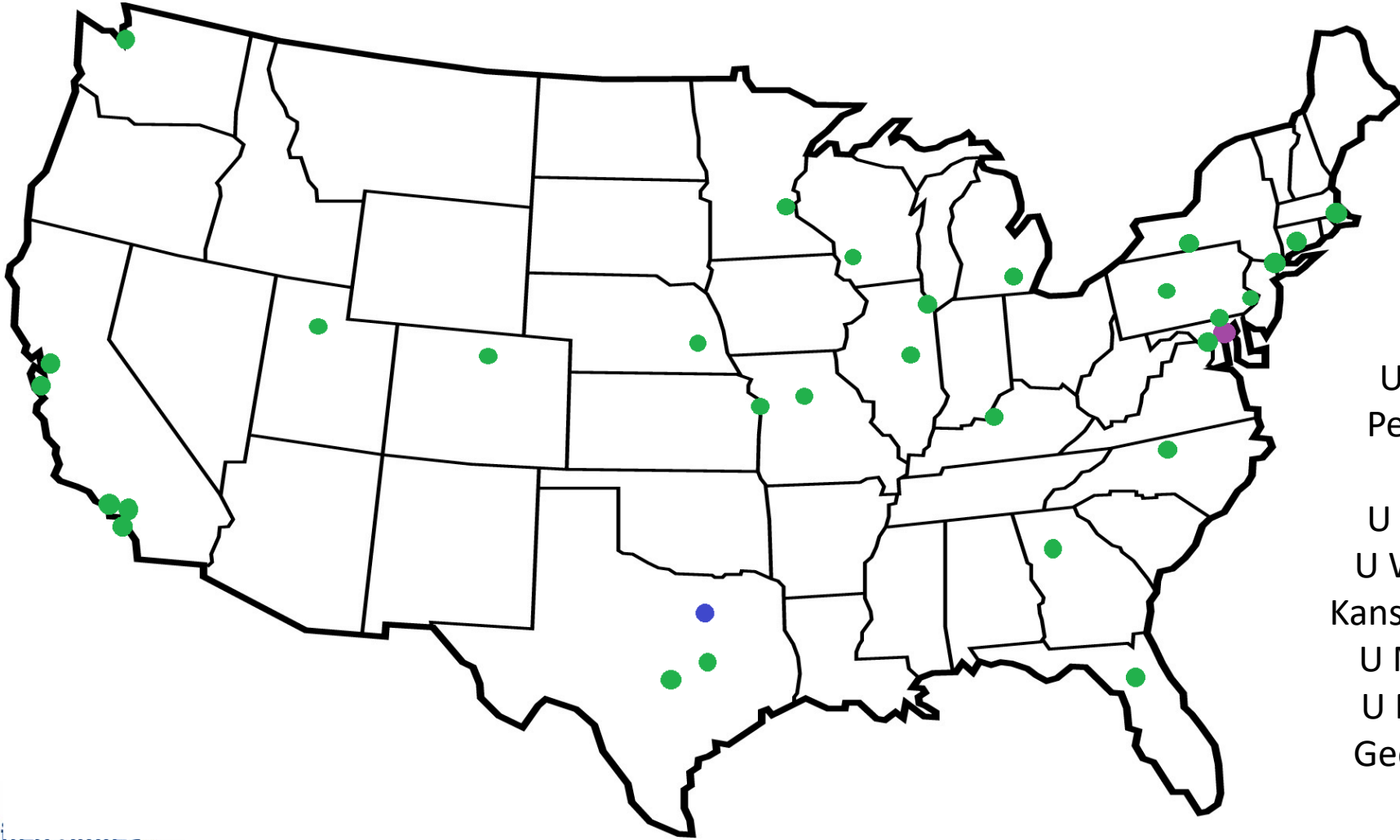
Research Data Centers



Research Data Centers



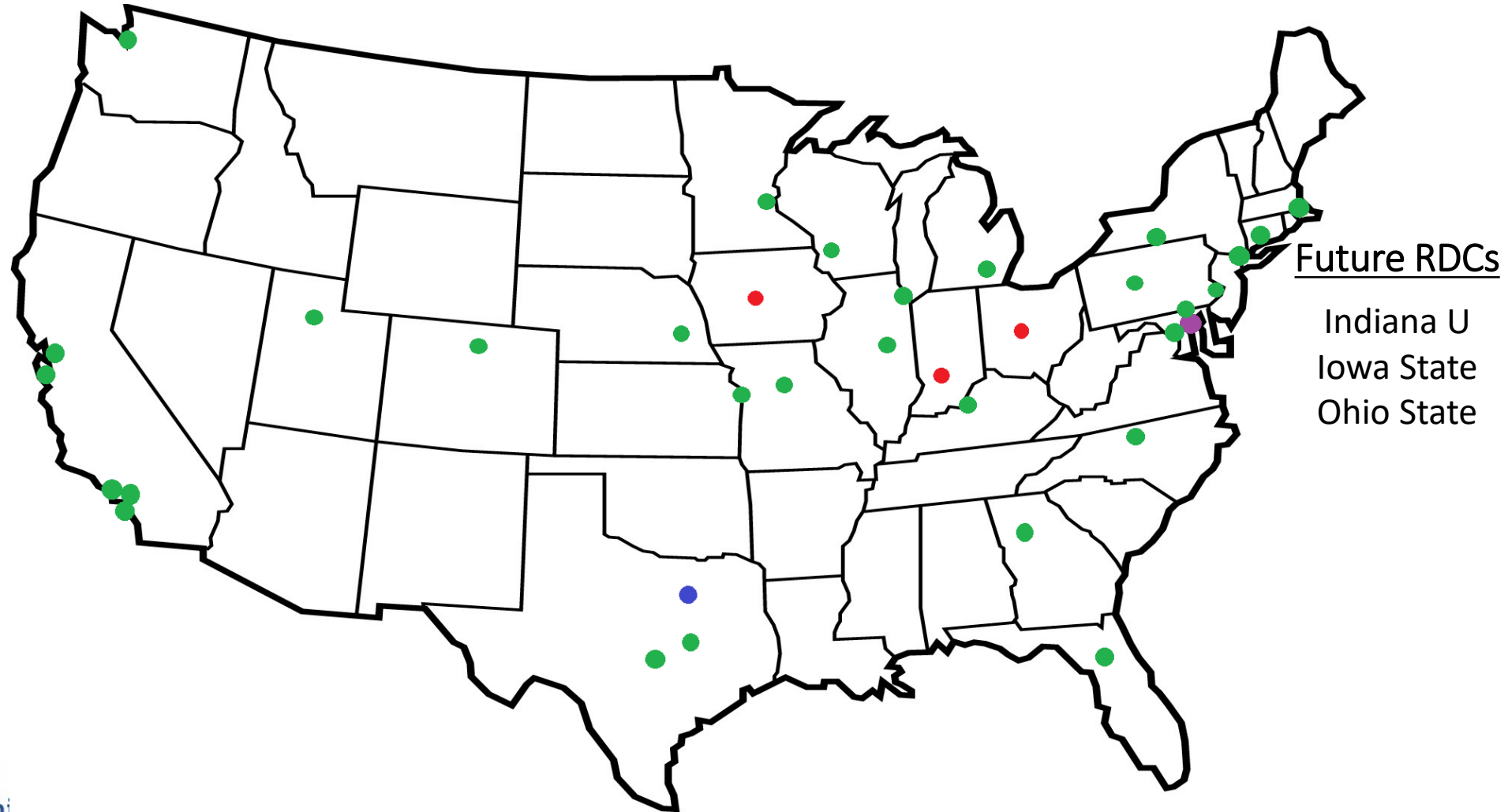
Research Data Centers

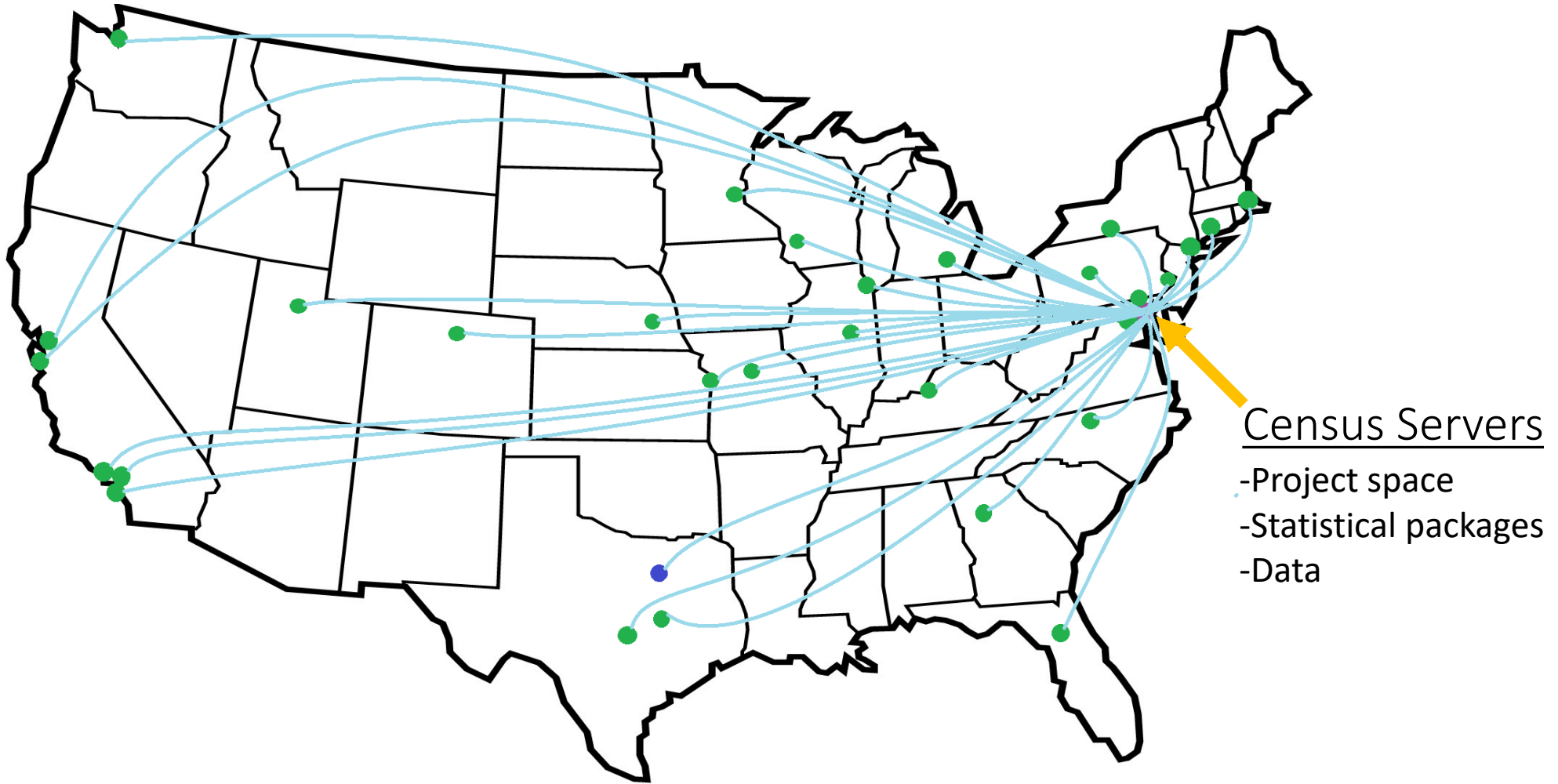


2014 - Present

- USC
- UC Irvine
- Penn State
- Yale
- U Missouri
- U Wisconsin
- Kansas City Fed
- U Maryland
- U Nebraska
- Georgetown
- U Kentucky
- Philadelphia Fed
- Colorado, Boulder
- UT Austin
- Dallas Fed
- Federal Reserve
- U Utah
- UIUC
- U Florida (2022)

Research Data Centers





Census Servers

- Project space
- Statistical packages
- Data

Why Is Census Required to Restrict Microdata Access?

- Titles 13 (Census) & 26 (IRS) U.S.C. and CIPSEA protect confidentiality
 - respondents cannot be identified
 - only Census employees and temp staff can access microdata
 - use limited to statistical purposes
 - access must benefit Census Bureau programs

Quick Poll

- Familiarity with the FSRDC program:
 1. Very knowledgeable of the FSRDC program and the process involved with working within an RDC
 2. Somewhat familiar/know of its existence
 3. This presentation is my first time learning about the FSRDC program

Data Availability

Census Bureau Data

- Economic Data
- Demographic Data
- Longitudinal Employer-Household Dynamics (LEHD)
- UMETRICS Data
- Criminal Justice Administrative Records System (CJARS)

Other Agency Data

- Agency for Healthcare Research and Quality (AHRQ)
- National Center for Health Statistics (NCHS)
- Bureau of Labor Statistics
- Bureau of Economic Analysis
- National Center for Science and Engineering Statistics (NCSES)
- Substance Abuse and Mental Health Services (SAMHSA)

Economic Data

Economic Censuses	Years
Auxiliaries (AUX)	1977-2012
Construction Industries (CCN)	1972-2017
Finance, Insurance, Real Estate (CFI)	1992-2017
Manufacturers (CMF)	1963-2017
Mining (CMI)	1987-2017
Retail Trade (CRT)	1977-2017
Services (CSR)	1977-2017
Trans., Comms., Utilities (CUT)	1987-2017
Wholesale Trade (CWH)	1977-2017

Business Register Data	Years
Compustat-SSEL Bridge (CSB)	1976-2016
Form 5500 Bridge File	1992-2018
Integrated Longitudinal Business Database (ILBD)	1977-2018
Longitudinal Business Database (LBD)	1976-2019
Ownership Change Database (OCD)	1963-2002
Standard Statistical Establishment List/ Business Register (SSEL/BR)	1974-2019

Transactions Data	Years
Commodity Flow Survey (CFS)	1993-2017
Foreign Trade Data - Export (EXP)	1992-2019
Foreign Trade Data - Import (IMP)	1992-2019
Longitudinal Foreign Trade Transactions Data (LFTTD)	1992-2019

Establishment Surveys	Years
Annual Survey of Manufacturers (ASM)	1973-2019
Current Industrial Reports (CIR)	1993-2011
Manufacturing Energy Consumption Survey (MECS)	1985-2018
Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)	1996-2020
National Employer Survey (NES)	1994-1998
Plant Capacity Utilization (PCU)	1974-2018
Survey of Manufacturing Technology (SMT)	1988-1993
Survey of Pollution Abatement Costs and Expenditures (PACE)	1979-2005

Firm Surveys	Years
Annual Business Survey	2017
Annual Capital Expenditures Survey (ACES)	1993-2019
Annual Retail Trade Survey (ARTS)	1999-2018
Business Expenditures Survey (BES)	1992-2007
Business R&D and Innovation Survey (BRDIS)	2008-2018
Enterprise Summary Report (ESR)	1977-1992
Exporter Database (EDB)	1992-2010
Quarterly Financial Report (QFR)	1977-2019
Service Annual Survey (SAS)	1991-2019
Survey of Business Owners (SBO)	1982-2012
Survey of Industrial R&D (SIRD)	1972-2007

Economic Censuses

- Quinquennial, years ending in 2 or 7
- Covers ~4 million business locations
 - Covers most industries and all geographic areas of the US
 - 9 different versions tailored to industry type
- Establishment level
- Variables included:
 - Name/address/ownership information/industry codes
 - Revenues, employment, payroll
 - Customer types (households, contractors, gov't, etc)
 - Sales by item code
 - Special inquiries depending on year and industry

Economic Census, Annual Surveys

- Intercensal years covered by other firm/establishment level datasets
 - CMF: Annual Survey of Manufactures (ASM)
 - CRT: Annual Retail Trade Survey (ARTS)
 - CSR: Services Annual Survey (SAS)
 - CWT: Annual Wholesale Trade Survey (AWTS)
- 2023 onward, the Annual Integrated Economic Survey will replace the following business surveys: ASM, AWTS, ARTS, SAS, ACES and M3UFO
- Link external business data using EIN or name and address.

Business Register (CBPBR) and Longitudinal Business Database (LBD)

- County Business Patterns Business Register
 - Covers all establishments in the United States
 - From IRS tax records and Census Surveys
 - # of employees, revenue, address,
- Longitudinal Business Database
 - Universe of all US businesses w/ paid employees
 - Links establishments across years/surveys
 - Payroll, size, birth/death, ownership, location
- Integrated Longitudinal Business Database (ILBD)
 - All employer and non-employer non-farm businesses

Additional Firm/Establishment Surveys

- Longitudinal Foreign Trade Transactions Data (LFTTD)
 - All transactions of \$2,500+
- Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)
- Annual Business Survey (ABS)
 - Info on selected economic and demographic characteristics for businesses and business owners
 - Survey of Business Owners
 - Annual Survey of Entrepreneurs
 - Business R&D and Innovation for Microbusinesses Survey

Demographic Data

Data Set	RDC Years	Geographic unit
American Community Survey (ACS)	1996-2020	Block
American Housing Survey (AHS)	1984-2019	Tract
Current Population Survey Supplements	(1997-2017)	MSA (mostly)
Current Population Survey (CPS ASEC) – March Supplement	1967-2020	Tract
Decennial Census	1950-2010	Block
Mortality Disparities in American Communities	2008-2015	Block
National Crime Victimization Survey (NCVS)	2006-2017	Block
NCVS School Crime, Police Public Contact Supplements	(2009-2017)	
National Longitudinal Mortality Study (NLMS)	1973-2011	County
National Longitudinal Survey (NLS) – Young/Mature Men/Women	1966-1999	Tract
Survey of Income and Program Participation (SIPP)	1984-2020	Tract/County

Demographic Administrative Datasets

- Numident (Social Security Administration)
 - Birth place, Birth date, Death Date (Social Security records)
- Census Household Composition Key
 - Links the PIKs of children born beginning in 1996 to the PIKs of their parents (from SSN registrations)
- Master Address File
 - MAF Extract (MAFX): Yearly snapshot of MAF for research purposes. Contains complete list of addresses, geography variables, and address ID (MAFID)
 - MAF Auxiliary Reference File (ARF): Links MAFID to individuals' PIKs so can track annual residence beginning in 2000

Demographic Administrative Datasets

- SNAP / TANF / WIC
 - Data availability varies by state
 - Must provide benefits to the Food and Nutrition Service
 - Evaluate and improve surveys
 - Provide estimates to improve understanding of poverty and food insecurity
 - Contribute to evidence-based policy and program evaluation
- Moving to Opportunity Survey (MTO)
 - Designed to help low-income families in public housing move to “opportunity” neighborhoods with vouchers and counseling.
 - Ran in five large cities between September 1994 and August 1998
 - Data collected on participants multiple times from 1994 through 2010
 - Individuals in RDC files have been assigned PIKs
- [Criminal Justice Administrative Records System \(CJARS\)](#)
 - Microdata from participating agencies following individuals through the criminal justice system
 - Includes arrestees, criminal defendants, inmates, and probationers/parolees

Demographic Administrative Datasets

- SSA data for CPS and SIPP respondents
 - Supplemental Security Record (SSR)
 - 831 Disability File
 - Master Beneficiary Record (MBR)
 - Summary Earnings Record (SER)
 - Detailed Earnings Record (DER)

RDC Demographic Data Advantages

- Individual and household data
- Less top-coding/censoring
- Additional variables (e.g. DOB)
- Larger sample
- More detailed geography
- Individual level linkages across time/surveys

Linking - the Power of the PIK

- PIK – a unique Census identifier assigned to individuals using name, birthdate, address information, and social security number.
- Many restricted Census demographic datasets contain PIKs
- PIK crosswalks make it possible to link restricted demographic datasets together:
 - 1940, 2000, and 2010 Decennial Census
 - American Community Survey
 - Survey of Income and Program Participation
 - Many administrative datasets

Protected Identification Key (PIK)

- Available in:

ACS	Decennial Census*
AHS	NCVS
CJARS	SSA (Numident)
LEHD data	SIPP
CPS-ASEC & Supplements	Administrative Data (SNAP, WIC, TANF, ETC)

- ~80-90% of the samples are PIK'd

*Not all years are PIK'd

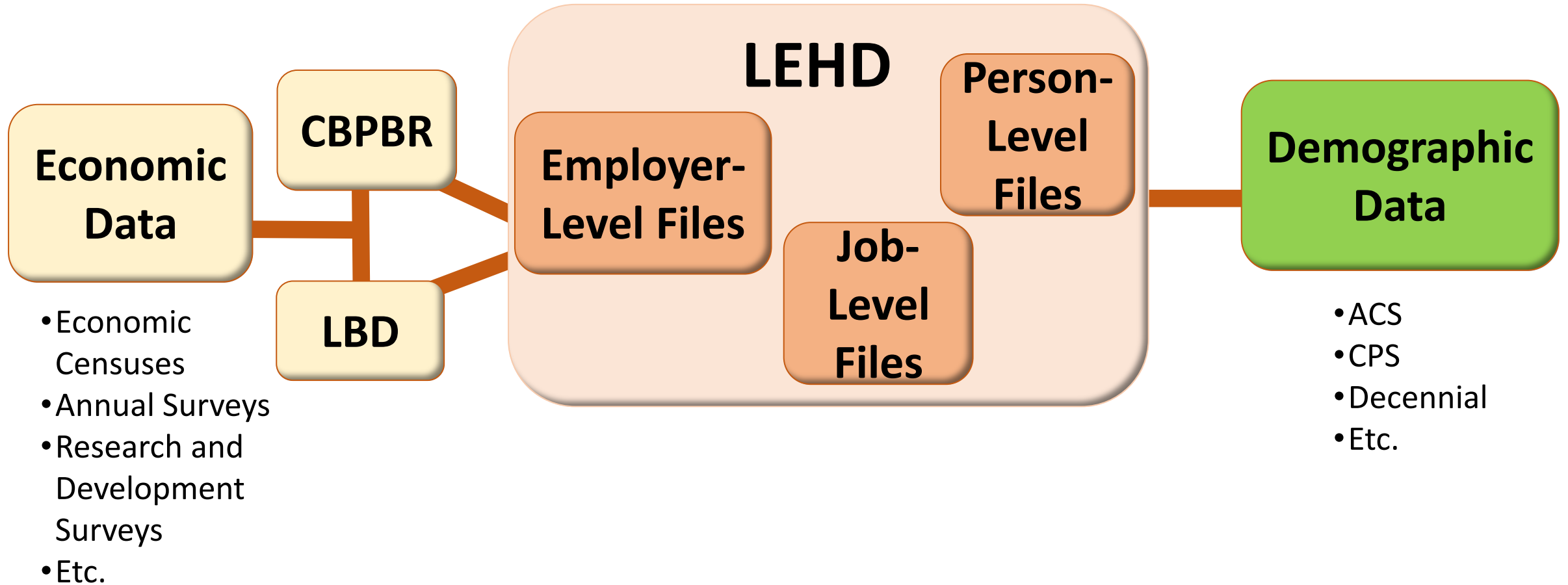
Linking by Geography

- Common to link researcher provided data to Census RDC data using geographic identifiers
- Geographic identifiers vary in the restricted data
- Common identifiers
 - State/County/FIPS
 - City
 - Zip
 - Address

Linked Data - LEHD

- Links individuals to place of employment
- Based on unemployment insurance administrative records
- Contains quarterly earnings for every company employee
- “Tracks” a person based on their place of employment
- Contains some demographic information on individuals
- Can link LEHD to other Census business data via firm identifier
- Can link LEHD to other Census household data via individual identifier
- Available on a state-by-state basis

LEHD Linkages



Criminal Justice Administrative Records System (CJARS)

- Longitudinal electronic records from criminal justice agencies
- Records are harmonized to track a criminal episode across all stages of the justice system from arrest to discharge
 - Bookings and arrests (from police/sheriffs' offices)
 - Trials, dispositions, sentences, and appeals (from local/state courts & DA offices)
 - Incarceration, probation, parole, and special programs (from state or local Department of Corrections)
- 150+ million criminal justice events & 30+ million individuals
- Anonymized records can be linked by PIKs to other restricted Census demographic data

UMETRICS

- Universities: Measuring the Impacts of Research on Innovation, Competitiveness and Science
 - Information on university grants (32 universities currently)
 - People paid on the grants
 - Businesses paid by the grants
- Can link grant recipients to internal Census Bureau data on individuals and firms

Data from Partnering Agencies Accessible in the RDC

- Agency for Healthcare Research and Quality (AHRQ)
- National Center for Health Statistics (NCHS)
- Bureau of Labor Statistics (BLS)
- Bureau of Economic Analysis (BEA)
- National Center for Science and Engineering Statistics (NCSES)
- Substance Abuse and Mental Health Services (SAMHSA)

- Aside from jointly approved projects, the proposal process is with individual agencies and not Census, though the background check is done by Census.

Health Data

Substance Abuse and Mental Health Services Administration (SAMHSA)

- Recently includes the National Survey on Drug Use and Health (NSDUH)
- [RDC - On-site at a FSRDC \(cdc.gov\)](https://www.cdc.gov/rdca/)
- Email: rdca@cdc.gov

Agency for Healthcare Research and Quality (AHRQ)

- [Medical Expenditure Panel Survey \(ahrq.gov\)](https://www.ahrq.gov/medexpend/)
- Email: CFACTDC@AHRQ.HHS.GOV

National Center for Health Statistics (NCHS)

Types of Restricted Variables

- Geographic Variables
- Linked Data Products
- Genetic Variables (NHANES phenotype data)
- 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

National Health Status Surveys

National Health and Nutrition Examination Survey (NHANES) I, II, and III

National Health Interview Survey (NHIS)

Longitudinal Study on Aging I and II (LSOA)

National Survey of Family Growth

National Survey of Children's Health

National Survey of Early Childhood Health

National Survey of Children with Special Health Care Needs

National Asthma Survey

National Health Care Surveys

National Ambulatory Medical Care Survey

National Hospital Ambulatory Medical Care Survey

National Survey of Ambulatory Surgery

National Hospital Discharge Survey

National Nursing Home Survey (NNHS)

National Home and Hospice Care Survey

National Employer Health Insurance Survey

National Health Provider Inventory

National Immunization Survey

Vital Statistics

Mortality and Multiple Mortality

Birth

Fetal Death

National Death Index

Marriage and Divorce

Agency for Healthcare Research and Quality (AHRQ)

- Medical Expenditure Panel Survey (MEPS) files include:
 - Household Component
 - Provider Component
 - Insurance/Employer Component
 - Nursing Home Component (1996 only)
 - Area Resource File
 - Two-year two panel file
 - MEPS-NHIS linked data
- Only Household Component and portions of Provider Component are publicly available

Agency for Healthcare Research and Quality (AHRQ)

- Fully specified ICD-9 medical condition codes
- Fully specified industry and occupation codes
- Lower levels of Geography
 - 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

Bureau of Labor Statistics Datasets

- Currently available in the FSRDCs:
 - National Longitudinal Surveys
 - Survey of Occupational Injuries and Illnesses
 - Census of Fatal Occupational Injuries (CFOI)
 - International Price Program
 - National Compensation Survey
 - Occupational Requirements Survey
 - Producer Price Indices

<https://www.bls.gov/rda/?#>
RDA_Admin@bls.gov

Bureau of Economic Analysis Data

- Allow researchers to analyze economic behavior of
 - Multinational Enterprises (MNEs)
 - Firms that trade in services
- Can be linked to Census firm level data
- U.S. citizenship required to access data
- <https://www.bea.gov/research/special-sworn-researcher-program>
- Email: SpecialSwornResearch@bea.gov

BEA Data

Data Set

Annual and Benchmark Surveys of U.S. Direct Investment Abroad (BE-10/11)

Annual and Benchmark Surveys of Foreign Direct Investment in the United States (BE-12/15)

Benchmark and Quarterly Survey of Transactions in Selected Services and Intellectual Property With Foreign Persons (BE-120/125), Annualized Data

Benchmark and Quarterly Survey of Financial Services Transactions between U.S. Financial Services Providers and Foreign Persons (BE-180/185), Annualized Data

Benchmark and Quarterly Surveys of Insurance Transactions by U.S. Insurance Companies with Foreign Persons (BE-45/140), Annualized Data

Annual Survey of New Foreign Direct Investment in the United States (BE-13)

Quarterly Survey of U.S. Direct Investment Abroad (BE-577), Annualized Data (expected soon)

Quarterly Survey of Foreign Direct Investment in the United States (BE-605), Annualized Data (expected soon)

National Center for Science and Engineering Statistics

- Survey of Earned Doctorates
- Survey of Doctoral Recipients
- PIK crosswalks will allow for links with other Census data

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Phases of an RDC Project

- **Proposal drafting**
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

The Standard Application Process (SAP)

- Proposals are now being submitted via newly launched Standard Application Process (SAP) Portal*.
- SAP Portal can be accessed here: <https://www.researchdatagov.org/>
- For more information visit here:
<https://www.census.gov/about/adrm/ced/apply-for-access.html>

**Assistance from your RDC admin will greatly reduce the need for future revisions to your application.*

Phases of an RDC Project

- **Proposal drafting**
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Phases of an RDC Project

- Proposal drafting
- **Submit proposal for review**
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Submit Proposal Through the SAP

- Census Review (no more than 12 weeks)
 - Overview of Research including research question, scientific merit, population of study, proposed data linkages, user provided data, etc.
 - Demonstrate need for non-public data
 - Methodology
 - Feasibility
 - Benefits to Census strength
- Additional Agency Review, if needed (up to 12 additional weeks)
 - Economic data (IRS), Numident (SSA), LEHD (states), CPS (BLS)

13 Pre-defined 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

Phases of an RDC Project

- Proposal drafting
- **Submit proposal for review**
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- **Security clearance (SSS)**
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Special Sworn Status (SSS)

- All persons working in the RDC require SSS
 - Census, NCHS, AHRQ, BLS, BEA, etc.
 - Census absorbs SSS cost, other agencies might not
- Couple months to process (longer for non-citizens)
 - Application process can start after Census project approval
 - Non-citizens must reside in the U.S. for 3 of last 5 years to qualify

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- **Security clearance (SSS)**
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- **RDC analysis**
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Working in the FSRDC Lab

- Majority of projects conducted entirely in the RDC lab
 - Data located on server in Maryland
 - Access data via thin client terminals located in cubicles
- No internet access or personal computers allowed in lab
- Statistical software available: SAS, Stata, R, Matlab, Python, QGIS, Agency reviews output before releasing
 - Penalty for disclosure is \$250,000 and/or 5 yrs in prison (inadvertent or otherwise)

Virtual RDC (VRDC) Access

- The 2018 Evidence Act requires an expansion of access to data to reduce geographic and financial barriers
- Eligibility
 - Approved researchers with active Special Sworn Status
 - Data owners approve remote access (i.e., Not all projects or datasets are approved for remote access).
 - ALL datasets used on the project must be remote access eligible.
 - Successful completion of remote access training and on-site FSRDC experience
- If you are interested in VRDC access, please consult with your local RDC Administrator.

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- **RDC analysis**
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- **Statistical output for Disclosure Avoidance Review**
- Provide working paper and benefit to Census

Disclosure Avoidance Review

- All output from an RDC projects goes through disclosure avoidance review
 - Ensures that “all required privacy-protecting steps have been taken and all disclosure standards are met.”
- Disclosure requirements vary based on which dataset is being used
- Census staff will email you approved output

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- **Statistical output for Disclosure Avoidance Review**
- Provide working paper and benefit to Census

Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

Timeline for Review Process

- Census review, (up to 12 weeks)
 - Other agency review, if needed (up to 12 additional weeks)
- After approval, SSS Packet
- Security clearance (12-16 weeks)
 - Additional time for non-citizens
- Project duration from 3-5 years

Thank you.

- My email: samuel.r.bondurant@census.gov
- Executive Director: anthony.murphy@dal.frb.org

More information:

- www.census.gov/fsrdc/about
- www.dfwrdc.org
- CES working paper series (census.gov or ideas.repec.org)