Metadata for Confidential Microdata

Todd Gardner
Center for Economic Studies
U.S. Census Bureau

Federal Committee on Statistical Methodology
September 14, 2018

Overview

- Federal Statistical Research Data Centers (FSRDC)

- Data Linkage at the Census Bureau

- Census Longitudinal Infrastructure Project (CLIP)

- ICPSR Census Bureau Data Repository
  - The Inter-University Consortium for Political and Social Research at the University of Michigan

Challenges When Using Metadata for Confidential Microdata

- Program areas collect data for specific purpose

- Future research use not a consideration

- Documentation
  - Not in standard format (unstructured or semi-structured)
  - Limited information about variables
  - Protected by Title 13

The Federal Statistical Research Data Center Program

- FSRDCs are partnerships between federal statistical agencies and leading research institutions

- Secure facilities managed by the Census Bureau to provide access to restricted-use microdata for statistical purposes only

Current FSRDC Program
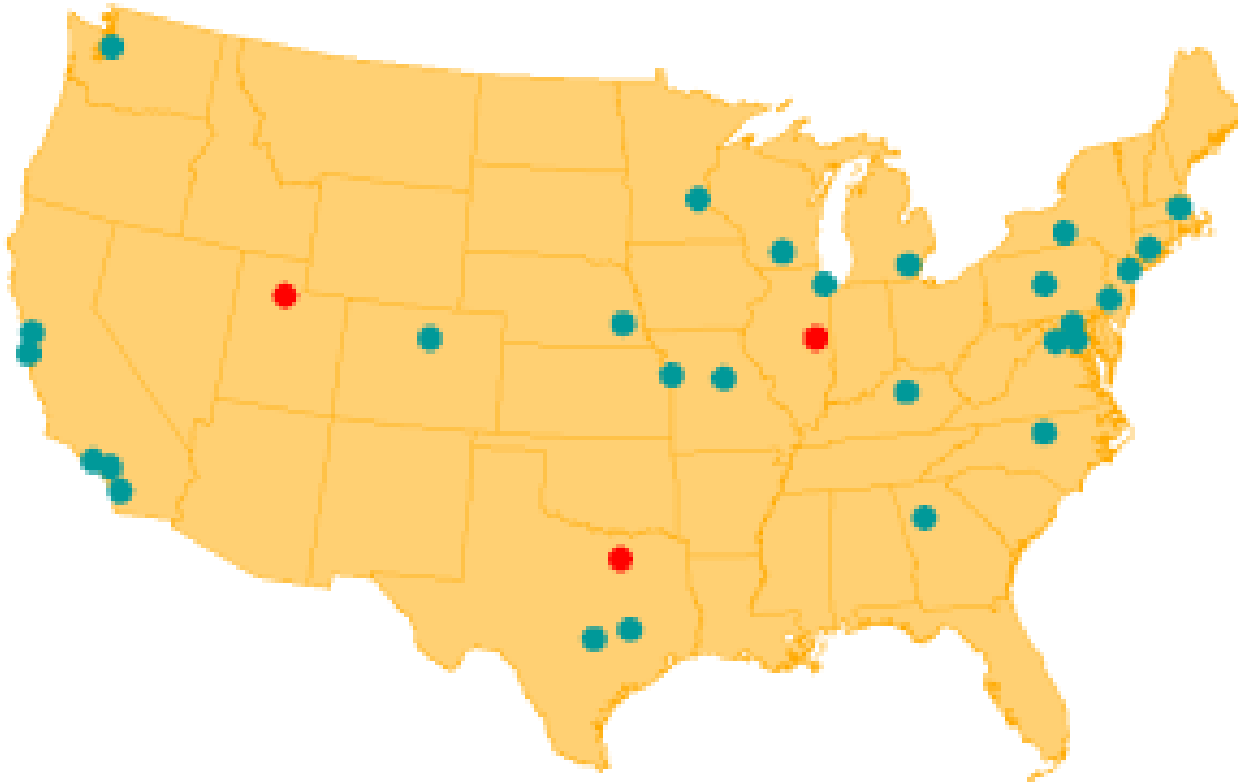
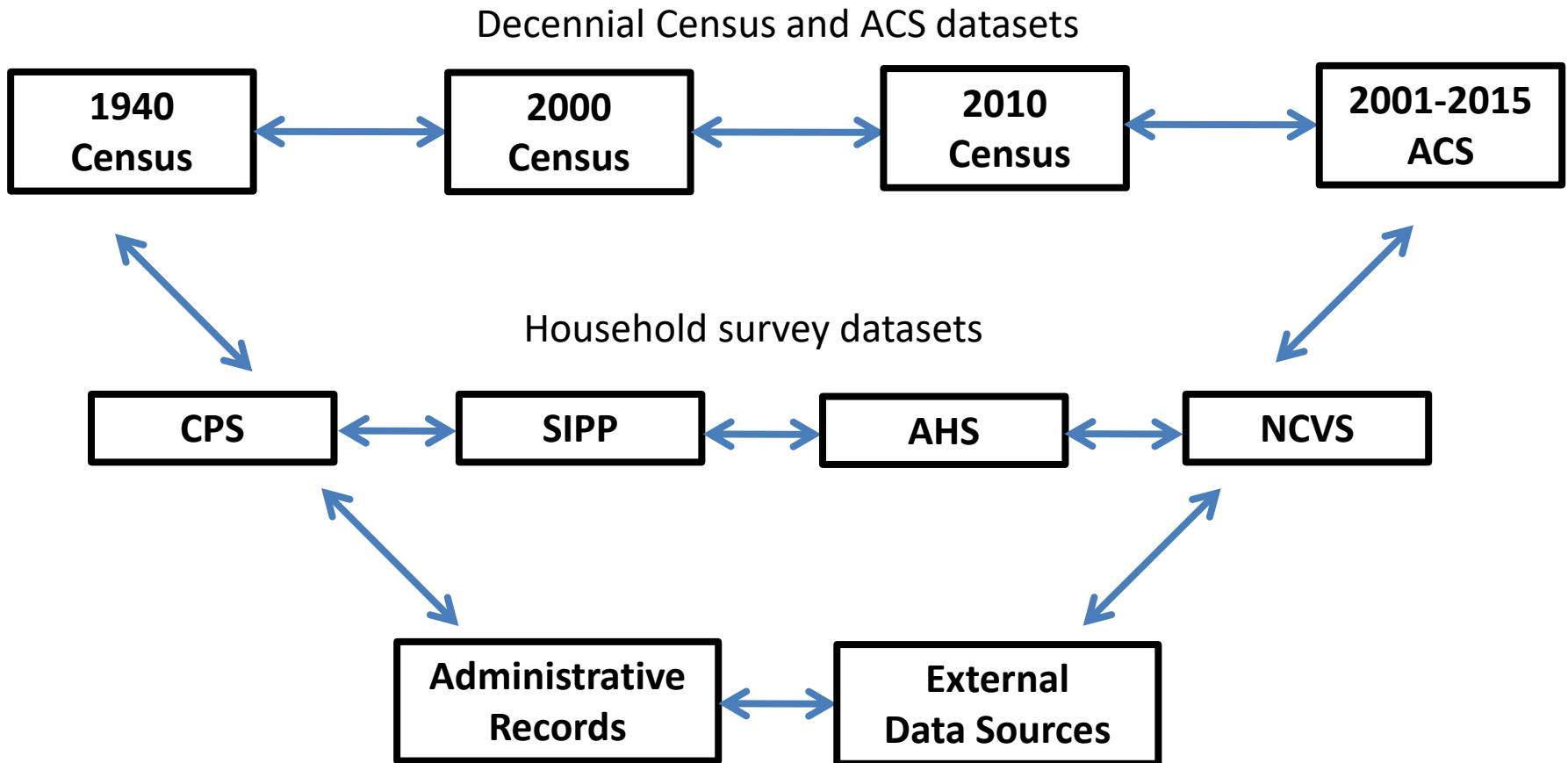28 FSRDCs (blue) with 3 (red) opening later this year

Available Data

- Economic Data
  - Economic Census, Longitudinal Business Database (LBD), and the Annual Survey of Manufactures (ASM), and many others
- Individual and Household Survey Data
  - Decennial Census, American Community Survey (ACS), Current Population Survey (CPS), and others
- Longitudinal Employer-Household Dynamics (LEHD) Data
- Administrative Records

Data Linkage

- The Census Bureau relies on record linkage for the strategic reuse of administrative data

- The Census Longitudinal Infrastructure Project (CLIP) uses the Census Bureau's record linkage and administrative records infrastructure to create longitudinal datasets for demographic analyses

United States™
Census
Bureau

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

Potential Linkages

Decennial Census and ACS datasets

| 1940 Census | ⟷ | 2000 Census | ⟷ | 2010 Census | ⟷ | 2001-2015 ACS |

Household survey datasets

| CPS | ⟷ | SIPP | ⟷ | AHS | ⟷ | NCVS |

| Administrative Records | ⟷ | External Data Sources |

CLIP Goals

- Data linkage
  - Refine matching procedures to improve accuracy of linkages
  - Understand biases introduced by matching procedures
- Data products
  - Enhance existing data
  - Create new data products
- Research
  - Expand academic collaboration and research
  - Improve social and economic measurement

Working Environment for Researchers
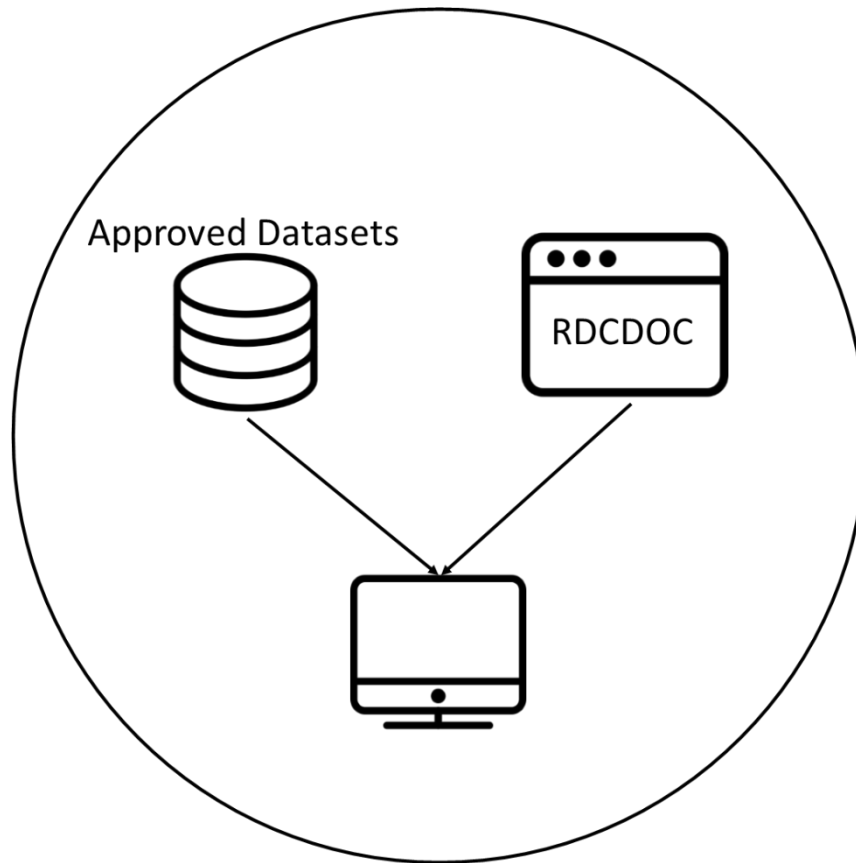
Each research project has access only to approved datasets in the data warehouse and the RDCDOC

The RDCDOC is an internal Census Bureau website that contains information for researchers, including documentation about the restricted-use datasets

The RDCDOC

- The RDCDOC collects information for researchers working in the FSRDCs and inside the Census Bureau
- The RDCDOC includes:
  - Dataset-specific documentation as provided by the program areas
  - Survey forms archive
  - The CES Technical Notes Series
  - Geographic reference materials
  - Software documentation
  - Information about updates to the data warehouse
- Problems
  - Difficult to maintain
  - Not available to the public

ICPSR's Census Bureau Data Repository
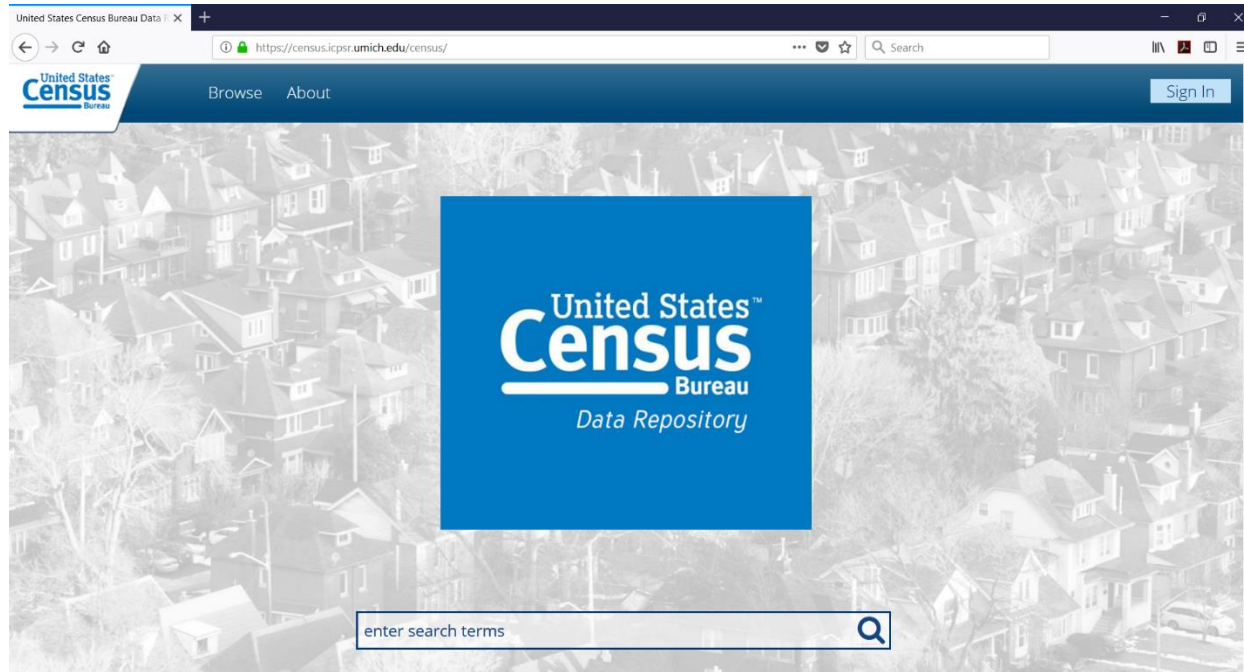
ICPSR has established a repository for U.S. Census Bureau data and metadata cleared for public release

ICPSR's Census Bureau Data Repository

- Preserves and disseminates survey instruments, specifications, data dictionaries, codebooks, and other materials provided by the U.S. Census Bureau

- Also lists additional Census-related data collections from its larger holdings

- Currently the repository contains metadata for only a few internal datasets

Metadata for Restricted-Use Files on ICPSR's Census Bureau Data Repository

ICPSR's Census Bureau Data Repository

- Helps fulfill key recommendations made by the 2017 "Report of the Commission on Evidence-Based Policymaking"

- Goals:
  - Improve transparency by providing a searchable inventory of datasets
  - Enable the public to learn about the data the government collects
  - Allow researchers to identify what data are available
  - Make proposal process more efficient

Challenges

- Disclosure review

- Converting metadata to a structured format

- Harmonization of data and documentation

- Incorporating user notes into metadata

Thank you!

Todd Gardner

Todd.K.Gardner@census.gov

Center for Economic Studies

www.census.gov/ces