

# **Access to Innovation: Utilizing the Census Bureau Data Linkage Infrastructure for Evidence-Based Research**

**Melissa C. Chiu**

Evidence Building Staff, Economic Reimbursable Surveys Division  
U.S. Census Bureau

American Evaluation Association

**Moving Beyond the Barriers 2.0 - Effective Models for Using Administrative Data to  
Generate Evidence and Enhance Evaluation Capacity**

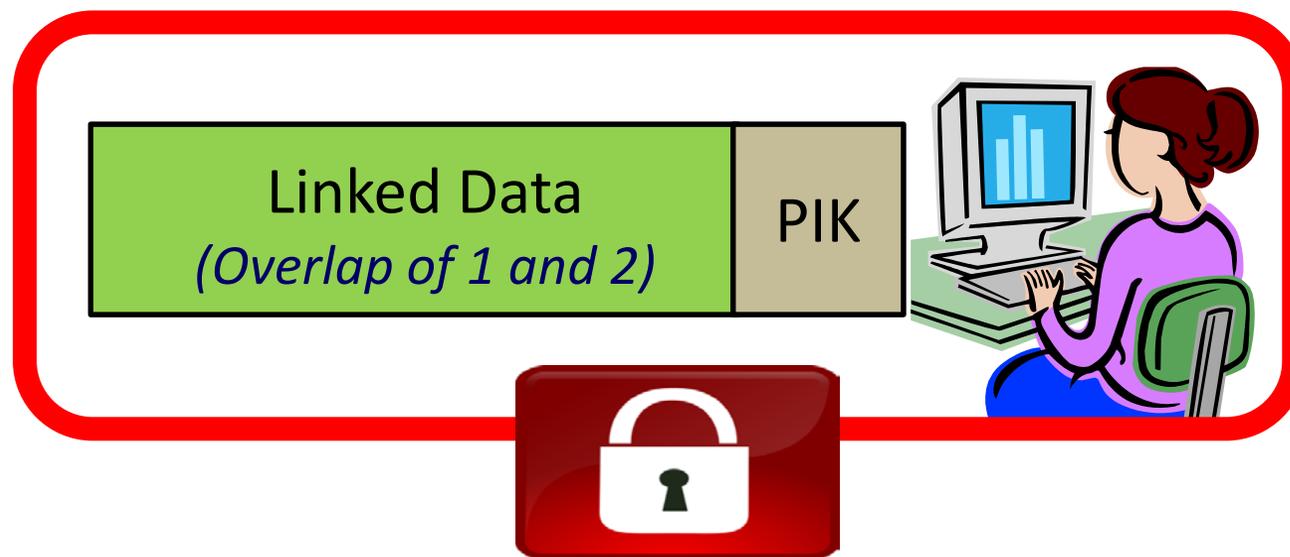
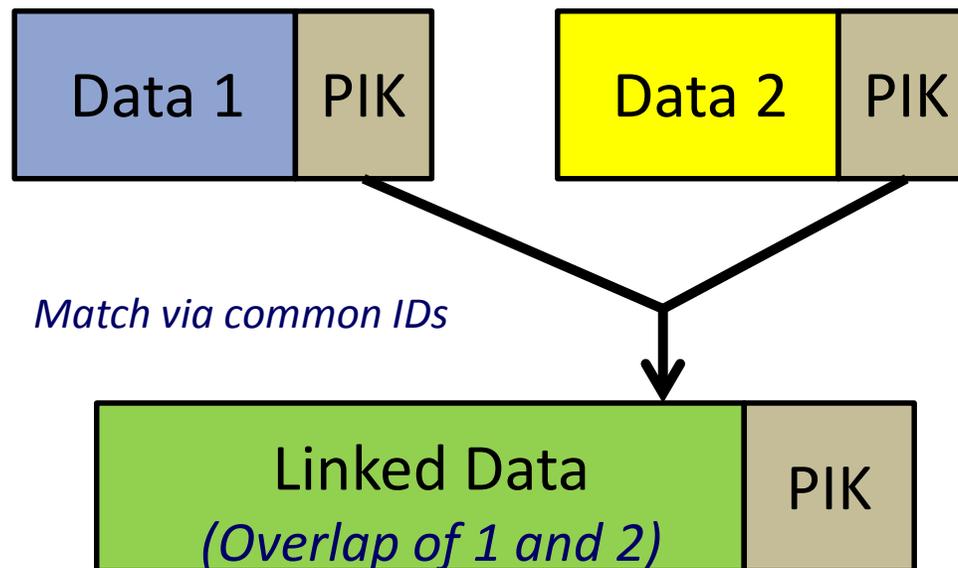
Nov. 9, 2017

# Data linkage infrastructure overview

**Acquire** datasets with  
13 U.S.C. §6 authority:

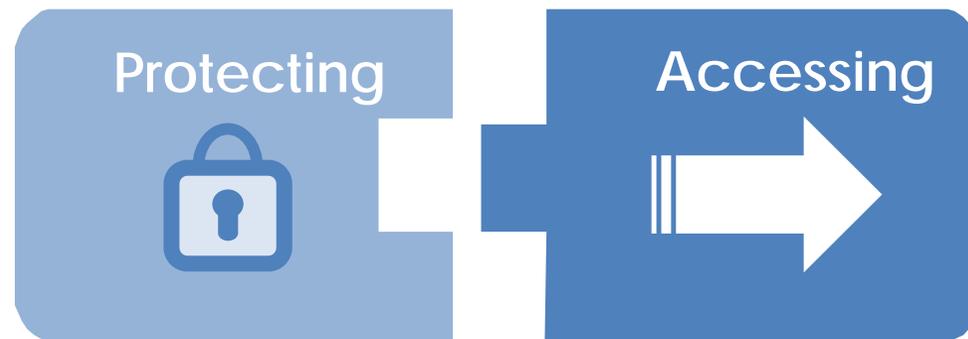
**Link** individuals or  
households across  
datasets:

**Provision** linked datasets  
to researchers in a secure  
environment for **analysis**  
and **statistical uses**:



# Legal framework and authority

*Title 13 provides authorities for protecting and accessing high value information about the nation's population and economy.*



**§ 6:** Acquire and utilize records to the greatest extent possible

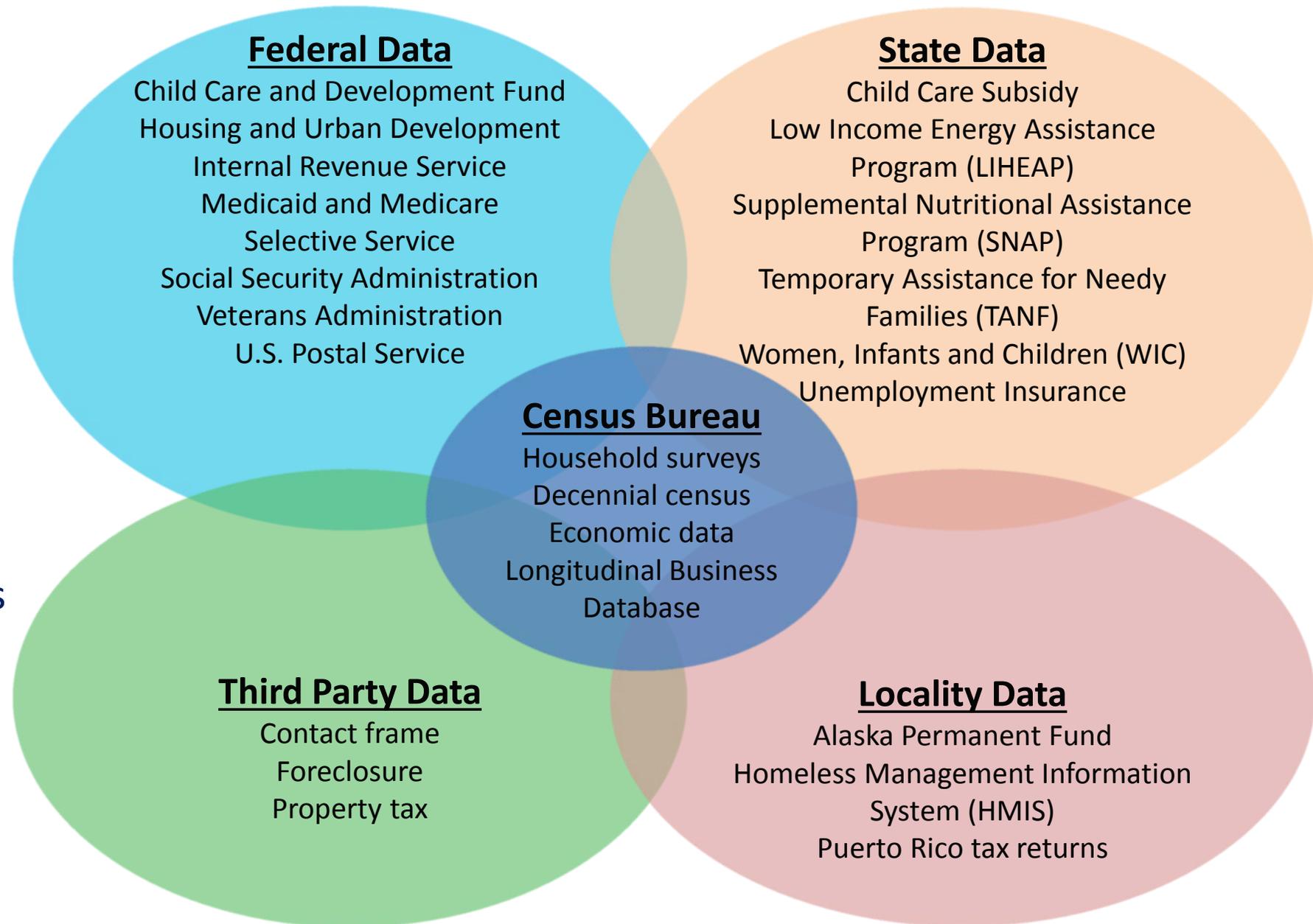
**§ 8:** Reimbursable studies and joint statistical projects

**§ 9:** Protect confidential individual and establishment data, limit access, and statistical uses

**§ 23(c):** Swear in researchers to assist the Census Bureau

## Research topics

People and households  
Employment  
Wages and earnings  
Education  
Public assistance disability  
Food security  
Health care  
Housing  
Public services and utilities  
Business



*For administrative data inventory, see: [www.census.gov/about/adrm/linkage/guidance.html](http://www.census.gov/about/adrm/linkage/guidance.html)*

*For Census Bureau restricted data, see: [www.census.gov/ces/dataproducts/restricted\\_data.html](http://www.census.gov/ces/dataproducts/restricted_data.html)*

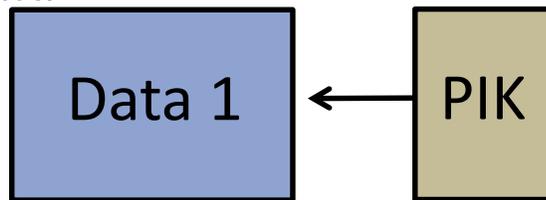
# PIK assignment

- Enhanced reference file based on Social Security Administration data
- **Exact** match of **SSN** with verification of name, DOB
- **Probabilistic** matching using **name, DOB, address, sex**
- See CARRA Working Papers:
  - <https://www.census.gov/about/adrm/linkage/working-papers.2014.html>
  - **2014-01**. The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software
  - **2014-02**. Estimating Record Linkage False Match Rate for the Person Identification Validation System

# Matching person records

## UNIQUE IDENTIFIER

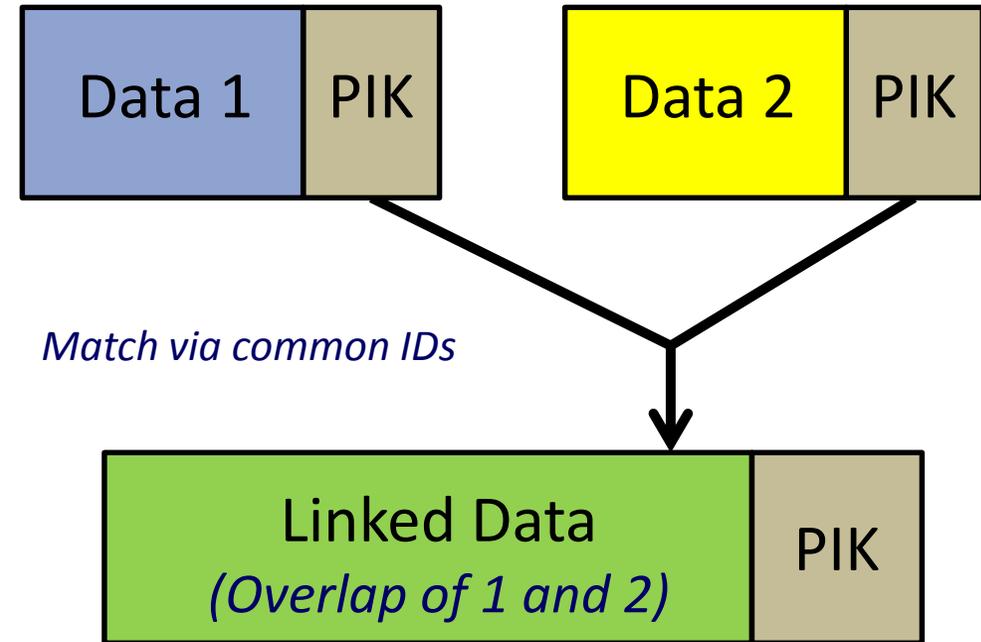
Census Bureau uses its Person Identification Validation System (PVS) to assign a unique identifier, a Protected Identification Key (PIK), to records in a dataset. The PVS is *conservative* to ensure the same person receives the same PIK.



All datasets must be used for *statistical purposes*. Content variables are stored without personally identifiable information (PII), resulting in added security.

## LINKING

Using the PIK, person records are matched across datasets.



# Secure environment

- Data stewardship
  - Users obtain Special Sworn Status, with background check
  - Data stewardship trainings, user agreement
  - Disclosure is punishable by law
- Federal Statistical Research Data Centers ([FSRDCs](#))
  - Physical and IT infrastructure
  - Restricted access rooms
  - Thin client monitors without data
  - NO: internet access, printers, etc.
- Disclosure review of output before release to user

# Infrastructure expansion

- Expanding infrastructure to support program evaluation, including performance
  - Especially interested in randomized controlled trials (RCTs)
  - Need identified data in order to link
- Streamlining access process and protocols
  - More timely
  - Census facilitates access, approvals
  - Continue to enable secure access
  - Working on expanding remote access



# Benefits for program evaluation and research

- Many datasets
  - Either cover one domain deeply or many topics shallowly
  - Selected universe, population coverage
- Integrated data can **fill in knowledge gaps**
  - Develop comprehensive understanding of outcomes across different social domains
  - Observe long-term and multi-generational outcomes
  - Provide comparison groups or benchmarks
  - Study methodological issues that affect inference and policy interpretation



# Operational and administrative benefits

- **Improve program administration**
  - Outreach
  - Before, during, and after implementation
- **Cost-effective:** Re-use of data is no or low cost
- **No additional burden** to respondents and program participants



# Selected state research products

- State programs- descriptive products used for outreach, understanding characteristics of the population they serve
  - Cross-domain
  - SNAP, WIC, or TANF linked with American Community Survey (ACS)
  - Survey data provide a comparison group of eligible nonparticipants, using model of eligibility
  - Table packages – household, sociodemographic, socioeconomic characteristics
  - Interactive visualization (SNAP and WIC)

# New York SNAP Profile 2012

## Interactive Visualization:

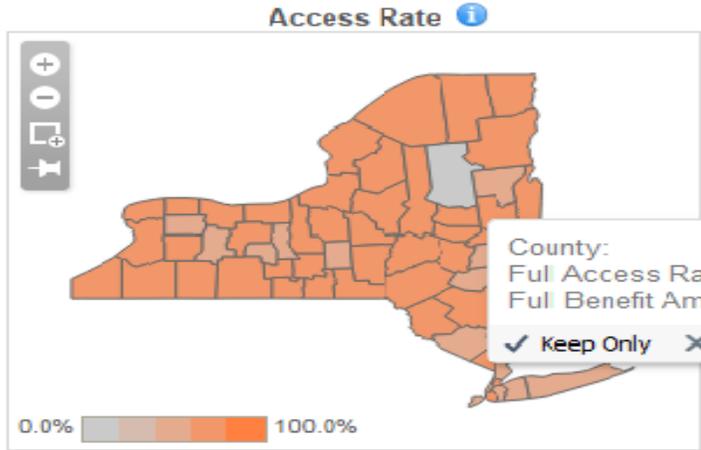
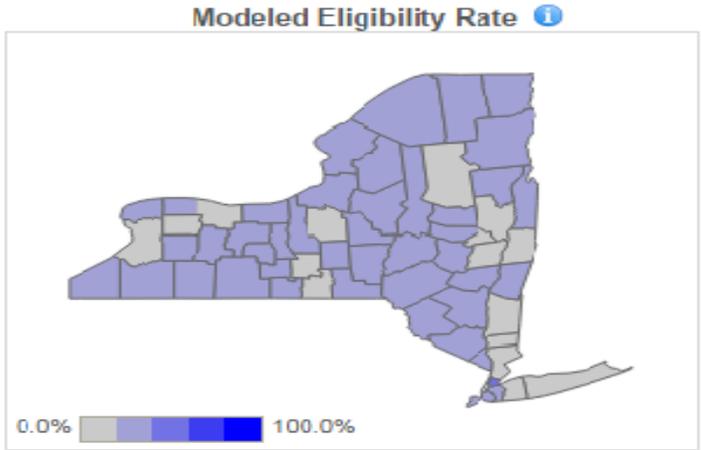
< Total Population Age Race Hispanic >

Tabs iterated for: total population, age, sex, race/ethnicity, veteran status

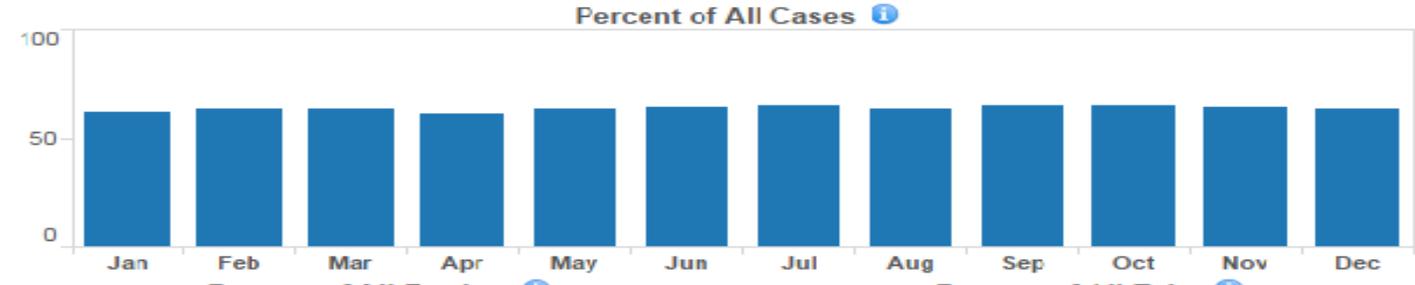
Hover over the apple for instructions: 🍏

Select a Race: White Alone

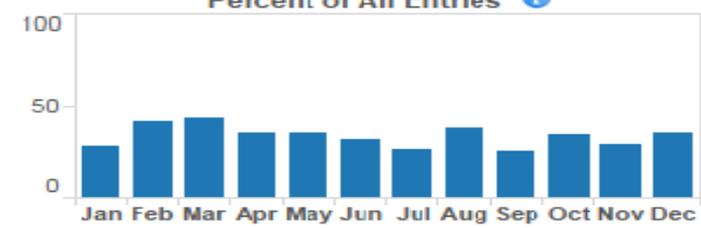
Race tab set to *White Alone*



Eligibility and take-up: Create own county groups and numbers recalculate



Percentage of participants, entries, and exits who reported race as "White Alone"



Source: Maps are generated using 2011-2012 New York SNAP AR linked to the 2012 1-year ACS. Charts are generated using 2012 New York SNAP AR linked to the 2012 1-year ACS.  
Note: Counties that do not meet minimum sample size requirements or that show no variation (i.e., where the value of the statistic is either 0 or 100 percent)

[www.census.gov/library/visualizations/interactive/snap-profiles.html](http://www.census.gov/library/visualizations/interactive/snap-profiles.html)

# Modeled, cross domain research products

- SNAP exit and county unemployment, industry employment growth
  - [Local Labor Market Conditions & Program Participation Dynamics](#)
  - NY SNAP, Census 2010
- Subsidized housing, mobility, high poverty neighborhoods
  - [The Effect of Low-Income Housing on Neighborhood Mobility: Evidence from Linked Micro-Data](#)
  - Low Income Housing Tax Credits (HUD), Census 2010 & 2000, public tract-level data
- Pollution exposure, adult wages, educational attainment, incarceration
  - [Air Quality, Human Capital Formation and the Long-term Effects of Environmental Inequality at Birth](#)
  - American Community Survey, SSA administrative data, IRS Form 1040s, EPA monitor data on pollution concentration, satellite remote sensing data on particulate matter
- Great Recession, minority-owned and women-owned businesses
  - [Small Business Growth and Failure during the Great Recession: The Role of House Prices, Race and Gender](#)
  - Longitudinal Business Database, Survey of Business Owners

# Methodological research products

- Methodological issues can have policy implications
  - Misreporting SNAP participation, spatial and temporal variation
    - [Within and Across County Variation in SNAP Misreporting: Evidence from Linked ACS and Administrative Records](#)
    - NY and TX SNAP, ACS
  - Medicare undercount, response validation, insurance coverage
    - [Medicare Coverage and Reporting: A Comparison of CPS and Administrative Records](#)
    - Medicare Enrollment, CPS ASEC
  - Quality of Medicaid responses of race and Hispanic origin
    - [Evaluating Race and Hispanic Origin Responses of Medicaid Participants Using Census Data](#)
    - Medicaid, Census 2010, ACS, Census 2000
  - Race and Hispanic origin reporting across sources
    - [When Race and Hispanic Origin Reporting are Discrepant Across Administrative Records and Third Party Sources: Exploring Methods to Assign Responses](#)
    - Census 2010, administrative data, third party sources

# Longitudinal research products

- Longitudinal, long-term outcomes
  - Differentials in demographic, socioeconomic characteristics on mortality
    - [National Longitudinal Mortality Study](#)
    - CPS ASEC, death certificates
  - Longer term effects of center-based versus home-based day care on child grade retention
    - [Is Subsidized Childcare Associated with Lower Risk of Grade Retention for Low-Income Children?](#)
    - Child Care and Development Fund, ACS
  - Environmental inequality, changes in individual pollution exposure
    - [Longitudinal Environmental Inequality and Environmental Gentrification: Who Gains From Cleaner Air?](#)
    - Satellite remote-sensing data of pollutants, IRS 1040s, Census 2000 & 2010
  - No multigenerational outcomes research examples yet
- ***Opportunities for your example here...***

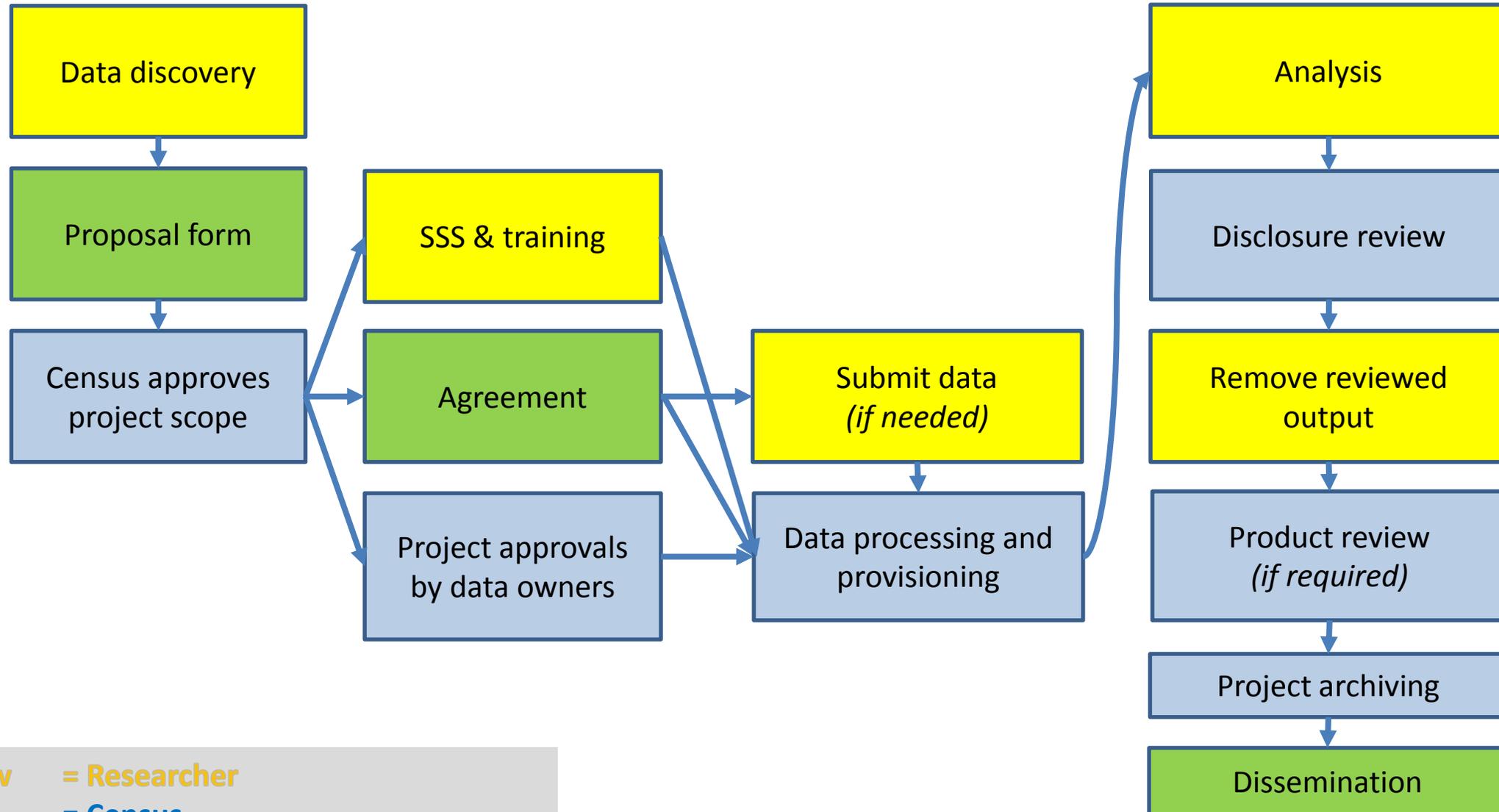
# Do you have ideas?



We want to help!  
There are ways to work with us...



# Access process overview



**Yellow** = Researcher  
**Blue** = Census  
**Green** = Both parties



# More information at [www.census.gov/datalinkage](http://www.census.gov/datalinkage)



Census.gov > About the Bureau > Data Linkage Infrastructure > Projects

## Data Linkage Infrastructure

About this Section

How do I access and use data?

Projects

Technical Documentation

Updates

Visualizations

Working Papers

### Projects

Census conducts a variety of research projects that use administrative records linked to census and survey data. Some of these projects generate new social and economic statistics, while others investigate ways to use linked data to improve sampling frames, survey measurement, and reduce respondent burden.

Additionally, Census collaborates with other stakeholders, including federal and state partners, to support the need for using statistical research to conduct evidence-driven program evaluation. Collaborations with researchers from academic and other institutions have led to the exploration of new avenues of research.

#### Census Longitudinal Infrastructure Project (CLIP)

Research building on Census' established infrastructure for linking records across data sources.

#### American Opportunity Study (AOS)

Collaborative effort to enable new research on social and economic mobility.

#### Social, Economic and Demographic Research

Exploratory research on a variety of topics including racial and ethnic identification and fluidity, income inequality, and social and economic mobility.

#### Program Participation and Policy-Related Research

Innovative research on program participation, including eligibility, coverage, and socioeconomic and well-being outcomes.

#### Record Linkage and the Person Identification Validation System (PVS)

Research on the quality of person matches to improve and expand record linkage with a variety of contemporary and historical data sources.

Contact us: [ERD.Data-Linkage@census.gov](mailto:ERD.Data-Linkage@census.gov)  
Melissa Chiu: [Melissa.C.Chiu@census.gov](mailto:Melissa.C.Chiu@census.gov)

CARRA Working Papers:  
[www.census.gov/about/adrm/linkage/working-papers.html](http://www.census.gov/about/adrm/linkage/working-papers.html)