



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

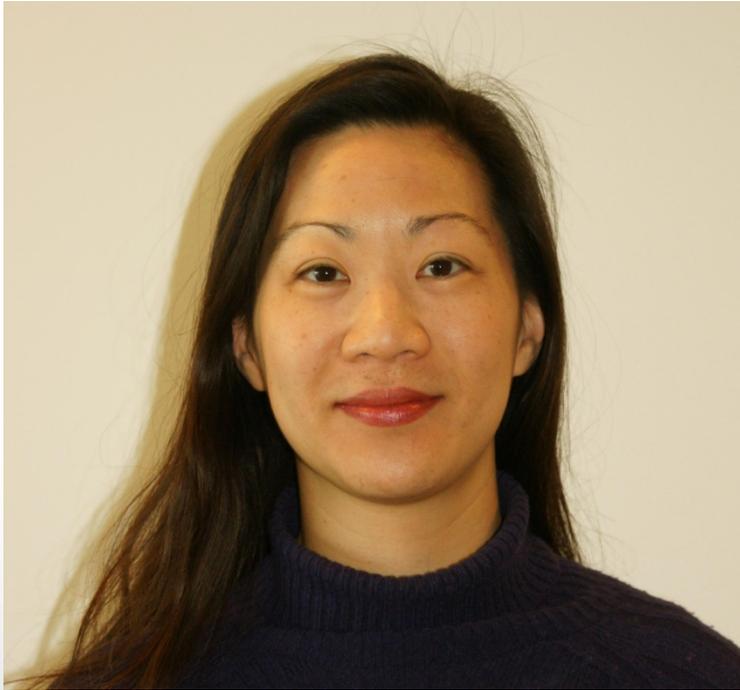
Melissa C. Chiu

Center for Administrative Records Research and Applications (CARRA)  
U.S. Census Bureau

August 24, 2016

Washington Evaluators Webinar

# Presenter



## **Melissa Chiu**

Data & Evaluation Outreach Manager  
Center for Administrative Records  
Research and Applications (CARRA)  
U.S. Census Bureau

# Outline

- Overview of existing Census Bureau Linkage Infrastructure
- Expansion to support program evaluation!
- Special opportunities:
  - HUD's Moving to Opportunity (MTO) data
  - Chapin Hall RFP, with Arnold Foundation support
- QnA

# Current data linkage infrastructure

- Datasets housed at the Census Bureau
  - Surveys, administrative records
  - Decades-long effort
- Record linkage methods
  - Steady state system to create unique person IDs, Protected Identification Keys (PIKs)
  - Use PIKs to link datasets
- Provisioning protocols enable secure access



# 13 U.S.C. § 6



## Access and Use of Administrative Records

(b) The Secretary may acquire, by purchase or otherwise, from States, counties, cities, or other units of government, or their instrumentalities, or from private persons and agencies, such copies of records, reports, and other materials as may be required for the efficient and economical conduct of the censuses and surveys provided for in this title.

# Data sources

## Federal Administrative Records

- Health and Human Services
- Housing and Urban Development
- Internal Revenue Service
- Social Security Administration
- Bureau of Justice
- U.S. Postal Service
- Office of Personnel Management
- Selective Service

## State Administrative Records

- Supplemental Nutrition Assistance Program
- Unemployment Insurance
- Women, Infants, and Children
- Low Income Energy Assistance Program
- Temporary Assistance for Needy Families
- Child Care Subsidy
- Permanent Fund

## Third Party Data

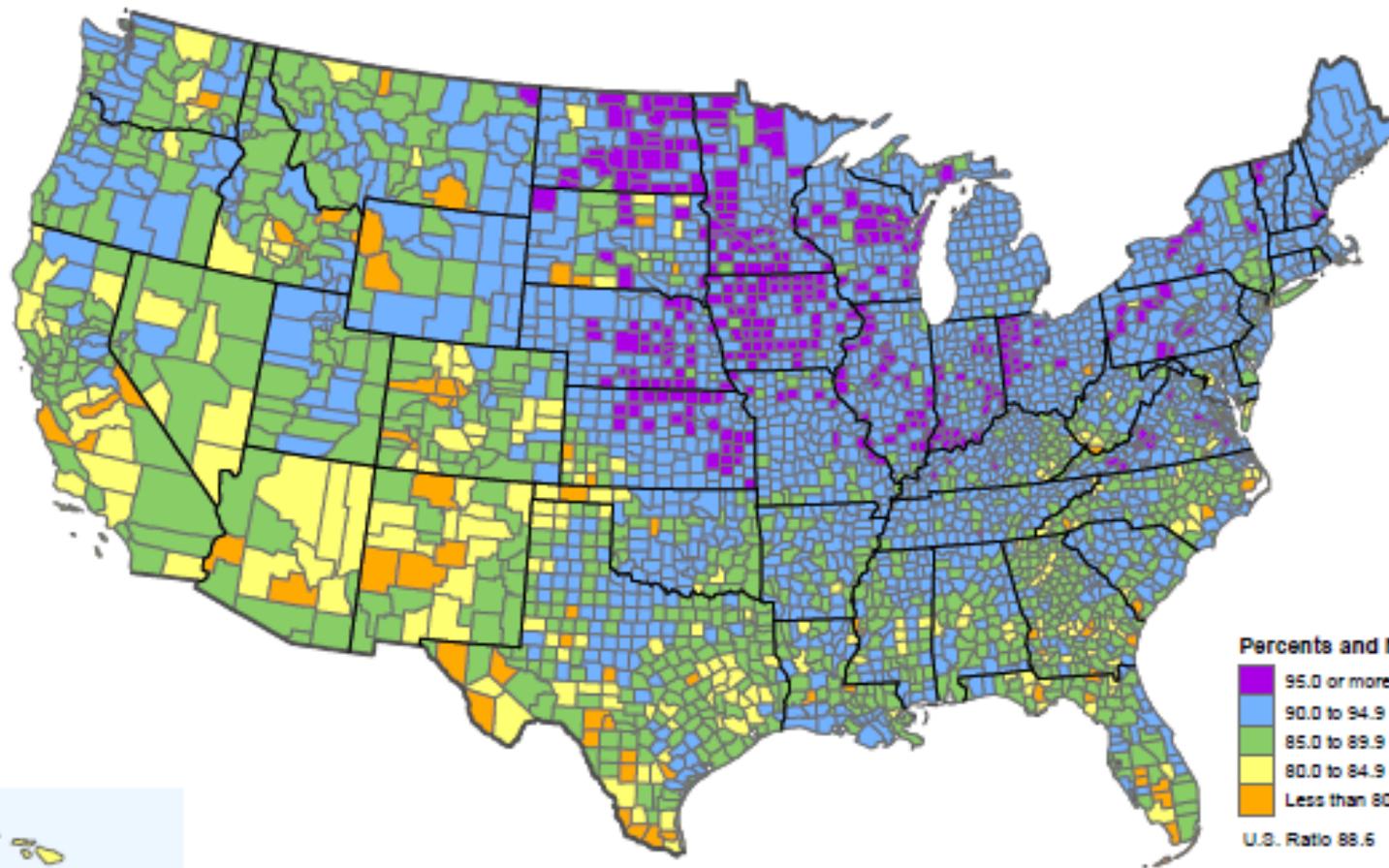
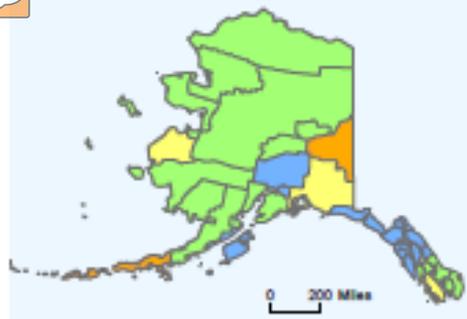
- Contact
- Household
- Housing



# Descriptive research products

- Census Bureau censuses, surveys, programs
- State programs- products used for outreach, understanding characteristics of the population they serve
  - Cross-domain
  - SNAP, WIC, or TANF linked with American Community Survey (ACS)
  - Table packages – household, sociodemographic, socioeconomic characteristics
  - Interactive visualization (SNAP and WIC)

# Match Ratio of 2010 Census and Administrative Records Persons by County



## Percents and Number of Counties

95.0 or more	N = 208
90.0 to 94.9	N = 1,611
85.0 to 89.9	N = 988
80.0 to 84.9	N = 249
Less than 80.0	N = 87

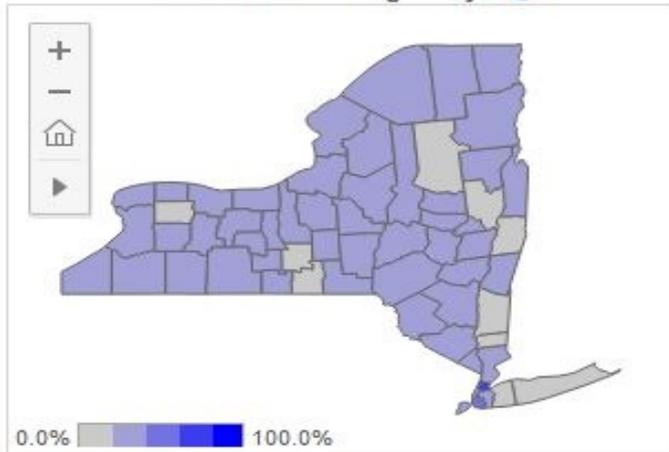
U.S. Ratio 88.6



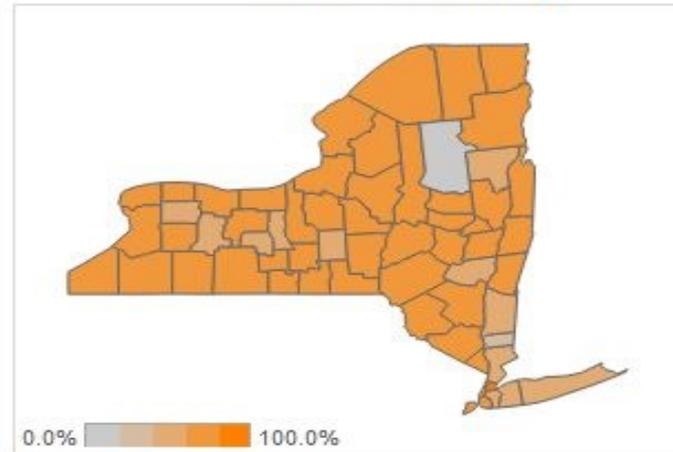
Source: 2010 Census and 2010 Census Match Study Administrative Records Data

# Interactive SNAP profile visualization

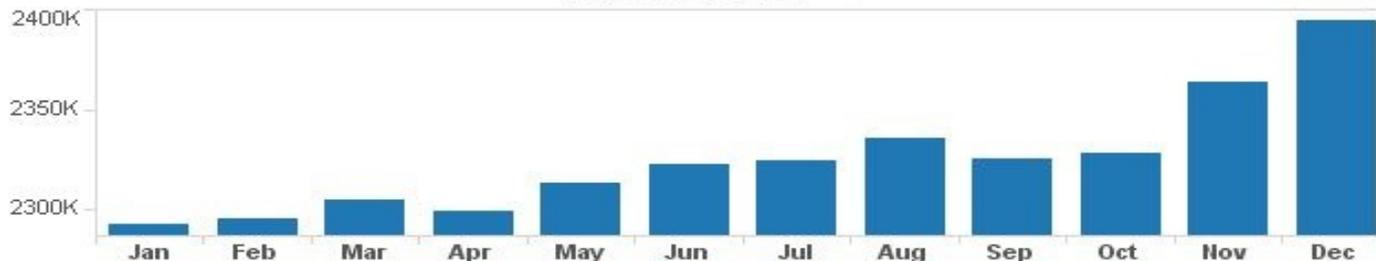
Estimated Eligibility [i](#)



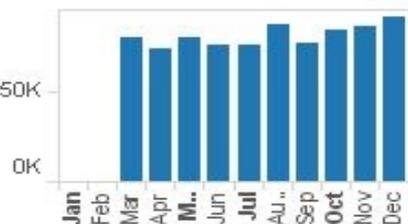
Estimated Access [i](#)



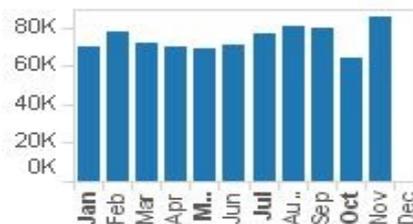
Number of Cases [i](#)



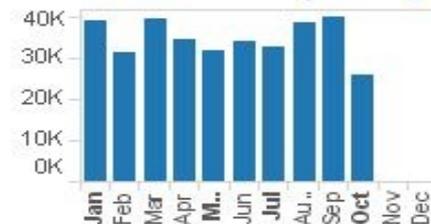
Number of Entries [i](#)



Number of Exits [i](#)



Number of Lapses [i](#)



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.

Note: Counties that do not meet minimum sample size requirements have missing values.

# More research products

- Modeled, cross domain
  - Do Doubled-up Families Minimize Household-level Tax Burden?
    - Current Population Survey Annual Social and Economic Supplement (CPS ASEC), 1040 tax data for Earned Income Tax Credit (EITC)
    - How complex households pool resources
  - The EITC over the business cycle: Who benefits?
    - CPS ASEC, EITC 2005 to 2010
    - Employment and eligibility patterns over the Great Recession
  - STEM Training and Early Career Outcomes of Female and Male Graduate Students
    - UMETRICS university administrative data, 2010 Census, W-2 forms
    - Gender differences in early STEM career and earnings
  - Local Labor Market Conditions & Program Participation Dynamics
    - NY SNAP, Census 2010
    - SNAP exit and county unemployment, industry employment growth

# Even more research products

- Methodological
  - Evaluating Race and Hispanic Origin Responses of Medicaid Participants Using Census Data
    - Medicaid, Census 2010, ACS, Census 2000
    - Quality of Medicaid responses of race and Hispanic origin
  - Medicare Coverage and Reporting: A Comparison of CPS and Administrative Records
    - Medicare Enrollment, CPS ASEC
    - Medicare undercount, response validation, insurance coverage
  - 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
    - Race and Hispanic origin reporting across sources

# And even more research products

- Longitudinal, long-term outcomes
  - National Longitudinal Mortality Study
    - CPS ASEC, death certificates
    - Differentials in demographic and socioeconomic characteristics on mortality
  - Does Childcare Setting Affect Grade Retention Risk? Using Linked Administrative Data to Assess U.S. Low-Income Children's Long-Term School Trajectories
    - Child Care and Development Fund, ACS, SIPP, CPS October School Enrollment Supplement (NCES)
    - Longer term effects of center-based versus home-based day care on child grade retention (research in early stages)
  - *Opportunities for your example here...*

# 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
  - Continue to enable **secure** access
  - Working on expanding **remote** access



# New data sources and infrastructure

## Data

- Federal, state & local survey and program data; third party data; data in the **wild**



## Tools

- Metadata, linkage and integration, analysis, maps, networks, visualizations



## Insight

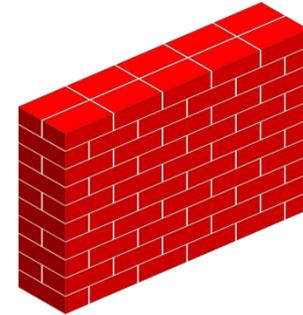
- Program administrators, policy makers, research community, public



*With secure access, privacy protections, and data stewardship!*

# Knowledge-making community

- Barriers to effective evidence-building
  - Knowledge gaps, data gaps
  - Fragmented, disconnected research
- Solution: Knowledge management by research and evaluation community, “knowledge-making”



# INTREPID

- “Innovation Network Transforming Research, Evidence, and Practice with Integrated Data”
- A community of knowledge makers innovating evidence-based solutions for a better society
- Census Bureau’s roles
  - Linkage infrastructure for integrating data
  - Space for sharing results, best practices
  - Initiate INTREPID in the near future
  - Partnering with HUD

# Do you have ideas?



We want to help!  
Here are some ways to work with us...

# Expansion to MTO intervention

- By Fall of 2016, expect to hold the Department of Housing and Urban Development's (HUD) **Moving to Opportunity (MTO)** data
  - Conducted in 1990s
  - Housing interventions
  - Multiple follow ups, studied outcomes across social domains
- First intervention data in the Census Bureau Linkage Infrastructure

# MTO project opportunities

- HUD is potentially supporting researchers' access fees through September 2020
- How to get started?
  - Write proposal and send to Census Bureau
  - Census Bureau will facilitate all approvals and reviews
- Plan to hold HUD's **Family Options Study** data by Fall 2017

# Pilot evaluation project opportunities

- Request for Proposals for evaluation studies that can benefit from Census' infrastructure
- With Chapin Hall, supported by the Laura and John Arnold Foundation
- **“Using Linked Data to Advance Evidence-Based Policymaking: Helping Projects Utilize the U.S. Census Bureau Linkage Infrastructure”**
- Deadline: **September 20, 2016**

# Goals

- Demonstrate an efficient way to **link state and local data to Census-held data** to answer important questions while protecting privacy
- Create **compelling use cases** for strengthening the Census linkage infrastructure to serve multiple levels of government
- **Inform** Federal, state, and local strategies for facilitating data linkage across programs

# Obstacles to use of data for evidence-based policymaking

- Fragmentation:
  - Intra-agency
  - Interagency
  - Intergovernmental
- Absence of compelling use cases
- Unreliable and incomplete data on program participants and results
- Interpretations of privacy requirements
- Lack of data infrastructure and analytical capacity

# Strategies to overcome resistance to data linkage

- Significant policy problem
- Strong executive leadership
- Strong partnerships among policymakers, program managers, researchers, and data experts
- Solution-oriented approach to privacy
- Prototypes that can be easily replicated

# Key features of Chapin Hall RFP

- Intergovernmental collaboration
  - State or local governments
  - Major non-profits supporting public sector programs
- Innovative partnerships
  - HUD
  - National Student Clearinghouse
- Researchers **bring in data** to link

# Submission types

## (I) Full Proposal

- Initial data-linkage pilots supported by Chapin Hall and the Census Bureau
- Projects are **ready for implementation** in 2016
- Full Proposal Submission:
  - Completed application form
  - Project summary
  - Principal Investigator CVs

## (II) Letter of Interest

- Inform Census Bureau Linkage Infrastructure expansion
- Strong LOIs may be invited to submit Full Proposal
- May be published and shared with the EBP Commission
- LOI Submission:
  - Cover letter
  - Project summary (one page!)

# Eligibility

## (I) Full Proposal

- Academic institutions, research or policy institutes, and governmental entities (except federal agencies)
  - Nonprofit service providers, federal agencies, and other entities are welcome to partner with researchers from eligible institutions to submit a full proposal

## (II) Letter of Interest

- **All institutions** are eligible to submit a LOI, including federal agencies and nonprofit service providers

# Services to selected projects

- Consultation on **preparation** of PI-held data and, if needed, **data processing** for incorporation into the infrastructure
- Consultation on Census Bureau data resources for analysis
- Facilitation of Special Sworn Status, background checks, and data sharing agreements
- Preparation of combined datasets, assignment of **unique identifiers**, and **documentation of data quality and data linkage**
- Facilitation of access to final datasets at a FSRDC, disclosure, and output review
- Note: Selected projects **do not** receive a monetary award

# Full proposal selection criteria

**Policy Importance and Relevance**

**Data and Research Design**

**Optimizing the Census Bureau Linkage Infrastructure**

**High Data Quality for Integration Success**

**Research or Evaluation Expertise\***

**Strong Commitment of Government or Nonprofit Partners**

**Ability to Meet Project Milestones**

\* “Researcher” broadly defined to include researchers, evaluators, investigators, and other similar personnel

# Example project categories

- Impact evaluations
- Long-term follow-up of prior RCTs
- Multi-generational studies
- Outcome measurement studies
- Needs assessment and descriptive studies
  
- Other project types are welcome

# Impact evaluation example

Intervention– Youth with Disabilities (PI-held)



Control Group



Treatment Group

Linked to Outcomes (federally-held)



Study impact of an RCT of an intervention serving youth with disabilities that is intended to improve educational and/or employment.

Study **post-transition outcomes** such as income, receipt of disability insurance, and receipt of housing assistance for people with disabilities.

# Long-term follow-up or multi-generation examples

Intervention— HUD's MTO (federally-held) or RCT (PI-held)



Linked to Outcomes (PI-held) and/or (federally held)



CRIMINAL RECORDS



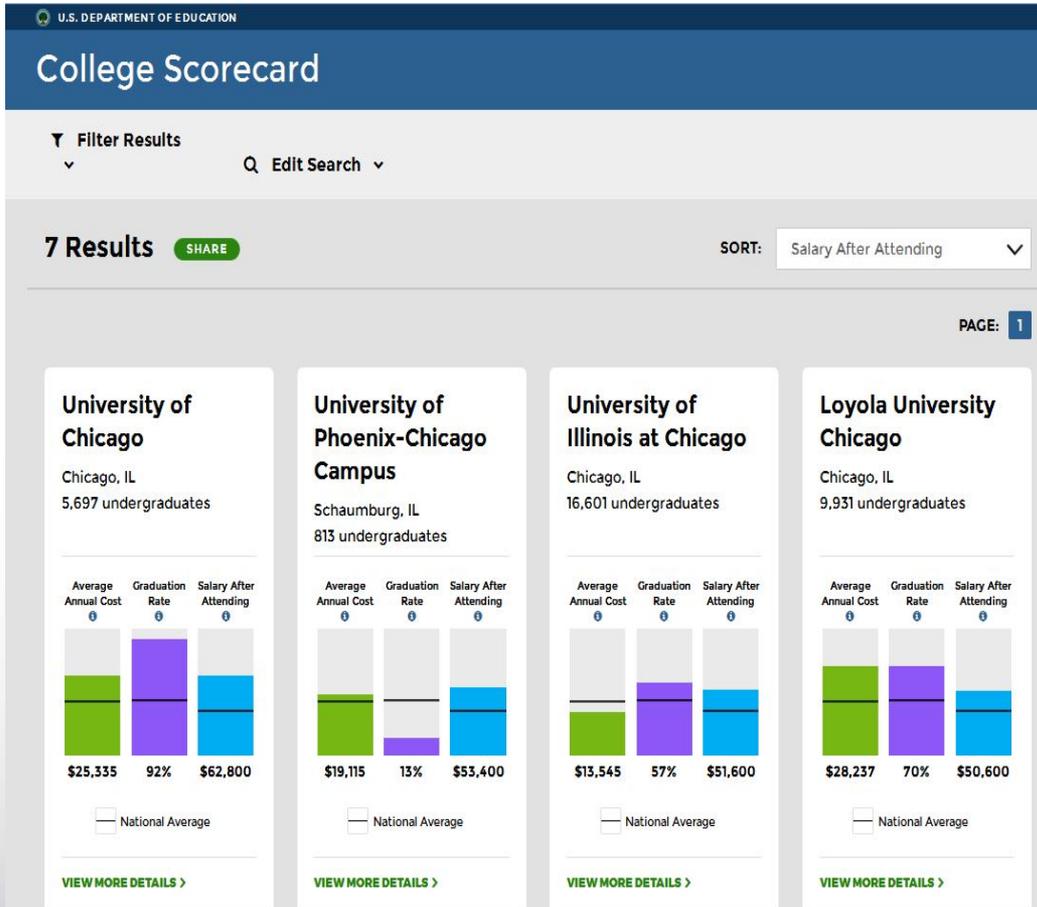
United States  
Census  
2000

United States™  
Census  
2010

Study **long-term impact** of HUD's Moving to Opportunity study across domains-- health, criminal convictions, postsecondary education, Neighborhood indicators: segregation, public health, crime rates.

**Two generation studies:** impact of an RCT on target group and their children

# Outcome measurement example



## State-level College Scorecards

- Average earnings of graduates by:

### PI-held data

- Campus
- Program of study
- Time to completion
- Participation in remedial programs
- GPA

# Needs assessment example

State or Local SNAP Enrollment (PI-held)



Linked to Population Characteristics (federally-held)



Study populations with high eligibility but low take-up rates for an entitlement, e.g., characteristics, geographic distribution. Use to plan outreach.

# Other examples

- Identifying geographic areas with:
  - High improper payment rates
  - High program utilization rates, where demand could be mitigated through prevention strategies

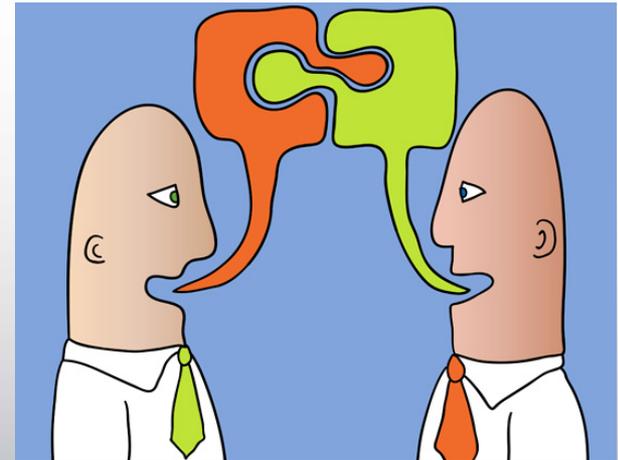
# RFP Information

**Due Date: September 20, 2016 - 11:59pm CDT**

- Main Page: [www.chapinhall.org/pages/RFP-Linked-Data-Evidence-Based-Policymaking](http://www.chapinhall.org/pages/RFP-Linked-Data-Evidence-Based-Policymaking)
  - Download [the RFP](#)
  - Check out [the FAQ](#)
  - Full Proposal [submission form](#)
  - LOI [submission form](#)
  - Specific questions? [lgjertson@chapinhall.org](mailto:lgjertson@chapinhall.org)

# Help us get the word out

- Please **disseminate** information about the MTO and pilot evaluations RFP to your colleagues and partners in state and local governments, research institutes, etc.
- If you are at a Federal agency and are interested in using the infrastructure, **talk with us**





# Contact us

## Melissa C. Chiu

Data and Evaluation Outreach Manager  
Center for Administrative Records Research and Applications  
U.S. Census Bureau

[Melissa.C.Chiu@census.gov](mailto:Melissa.C.Chiu@census.gov)

301-763-2421

[ADRM.CBLI@census.gov](mailto:ADRM.CBLI@census.gov)

## Census Bureau Linkage Infrastructure (CBLI):

<http://www.census.gov/about/adrm/data-linkage.html>

## CARRA Working Papers:

<http://www.census.gov/srd/carra/>

# Appendix Slides

# Evidence-Based Policymaking Commission

- Census infrastructure **distinct** from the Evidence-Based Policymaking Commission, which will:
  - **Study** “data inventory, data infrastructure, database security, and statistical protocols related to Federal policymaking and the agencies responsible for maintaining that data”
  - **Recommend:**
    - Whether a clearinghouse for program and survey data should be established
    - How to create such a clearinghouse
    - How governed
- Seated until September 2017



# Matching person records

## UNIQUE IDENTIFIER

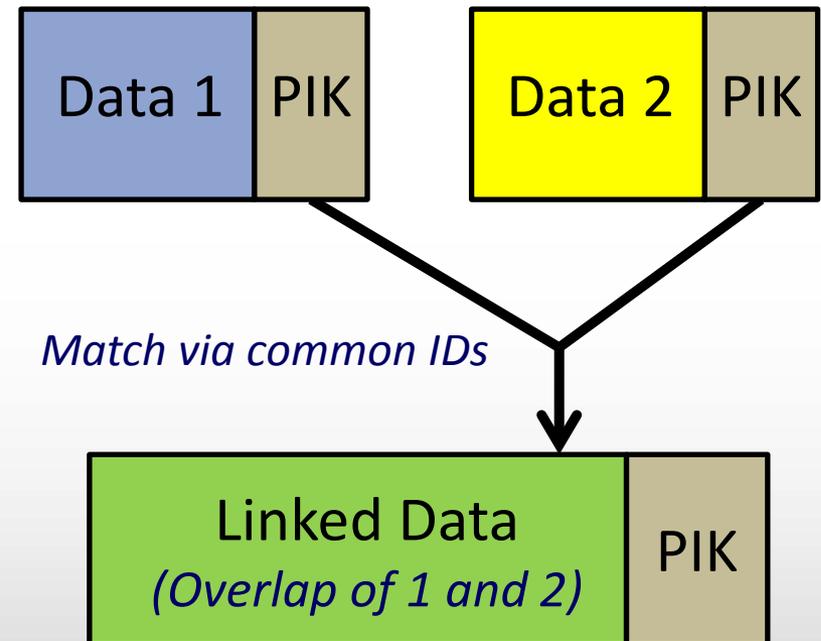
*Census Bureau uses its Person Identification Validation System (PVS) to assign a unique identifier, a Protected Identification Key (PIK), to records in a dataset. The PVS is conservative to ensure the same person receives the same PIK.*



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.*



# More on PVS and linkage

- See CARRA Working Papers:
  - <http://www.census.gov/srd/carra/>
- **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

# DATA ACCESS PROCESS FOR EXTERNAL RESEARCHERS

