

# From Ingest to Extract: Standardizing Data Preparation

Amy O'Hara

U.S. Census Bureau

December 16, 2014

# Overview

- Supplemental Nutrition Assistance Program (SNAP) Data Acquisition
- Census Bureau Plans with the Data
- Data Sharing Agreements
- Routinized Process for Data Preparation
- Output of Process
- Conclusion

# Supplemental Nutrition Assistance Program (SNAP)

- SNAP is the largest program in the domestic hunger safety net
- 46.5 million participants in 2014
- From *Building a Healthy America* (April 2012):
  - Nearly half of participants are children
  - Many elderly and disabled
  - Many participants work

# Census Bureau Plans with SNAP Data

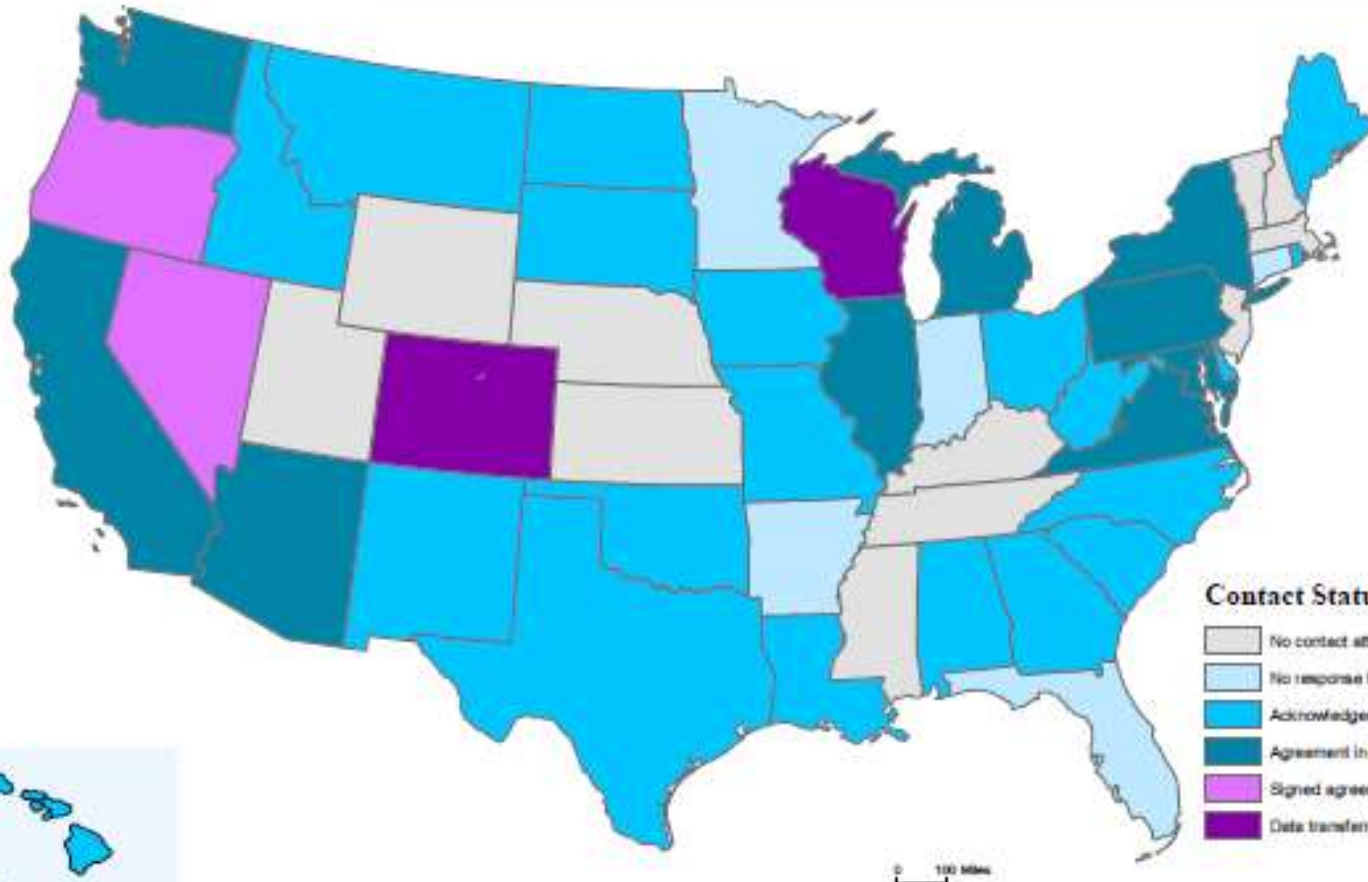
- Evaluate and improve survey data
- Improve record linkage methods
- Research and planning for the 2020 Census
- Understand program participation, characterize likely eligible non-participants
- Joint statistical research with Economic Research Service (ERS)

# Data Sharing Agreements

- Letter of support between USDA Food and Nutrition Service and Census Bureau
- Agreement template
- Statistical product for state program officials
  - Insights by matching to the American Community Survey (ACS) to study program eligibility and take-up
  - Exploring options to make aggregated data available on a public facing website for tabulation and visualizations

# Data Sharing Agreements (2) –

## State SNAP & WIC Data Acquisition: Contact Status



# Routinized Process for Data Preparation

- Products needed for every state where we enter an agreement
- No standardized data content across states
- Need a process for receiving SNAP data
- Need research on record validation and unit composition
- Need to produce reports on harmonized data

# Routinized Process (2) – Data Preparation Handbook

- Outline our process for receiving SNAP data
  - Data discovery
  - Data intake
  - Data clearance
  - Initial quality control (QC)
  - Data preparation for Person Identification Validation System (PVS) processing
  - Post PVS QC
  - Creating a state research file
  - Generating tables



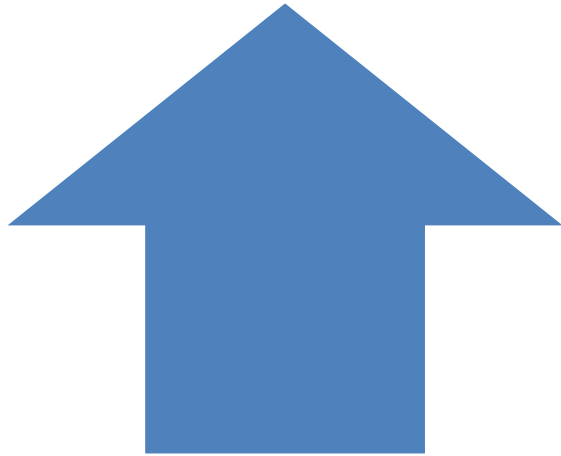
# Routinized Process (3) – Using State Research File

- Which records were validated and eligible for matching
- Research using the program data
- Research on ACS linked data including SNAP variables, demographics, and unit composition

# Output of Process

- Uniformly documented, processed state data
- Data quality and linkage information
- Enhanced understanding of program
- Information for states

# Output of Process (2) – Who gets a PIK

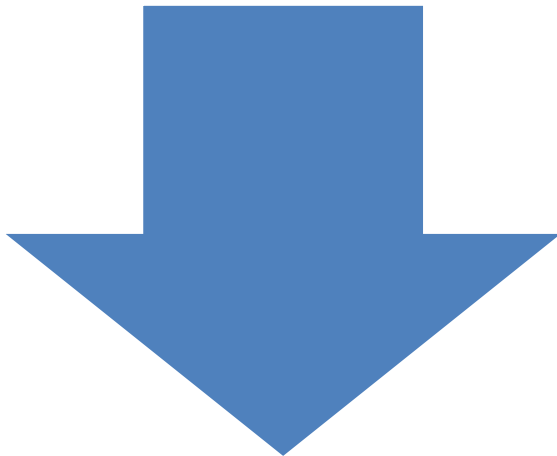


Married

More months of benefits

White

Non-Hispanic



Younger

Fewer years of completed  
education

Missing data (e.g., date of birth)

# Output of Process (2) – Matching to ACS

## 2012 New York

- 3,988,623 SNAP recipient records
  - 99.2% received a PIK
- 344,654 ACS persons
  - 89.4% received a PIK
- 48,823 linked persons

## 2012 Colorado

- 727,645 SNAP recipient records
  - 99.9% received a PIK
- 87,741 ACS persons
  - 91.6% received a PIK
- 8,346 linked persons

# Value-added for States

- 33% of New York residents were estimated to be eligible; of those, 65% participated in SNAP
- Eligibility and participation varied across counties and by race, Hispanic origin, English language ability, age, and employment status
- Erie County had a 71.5% participation rate, Bronx County had a 80.4% participation rate
  - Erie County had 48.4% of eligible persons 70 and older participating, but in Bronx County 70.6% of its eligible persons 70 and over were participating

# Conclusion

- Handbook, processing checklist, evaluation of data quality, and standardized statistical products achieved through our process
- Helps us satisfy the terms of our data use agreements
- Increases efficiency in the production of statistics that help program administration
- Primary costs are incurred in establishing and maintaining the data use agreements; initial data receipt is labor intensive
- Expand to other programs