Data Infrastructure for Studying Mobility: The Use of Administrative Data to Study Mobility

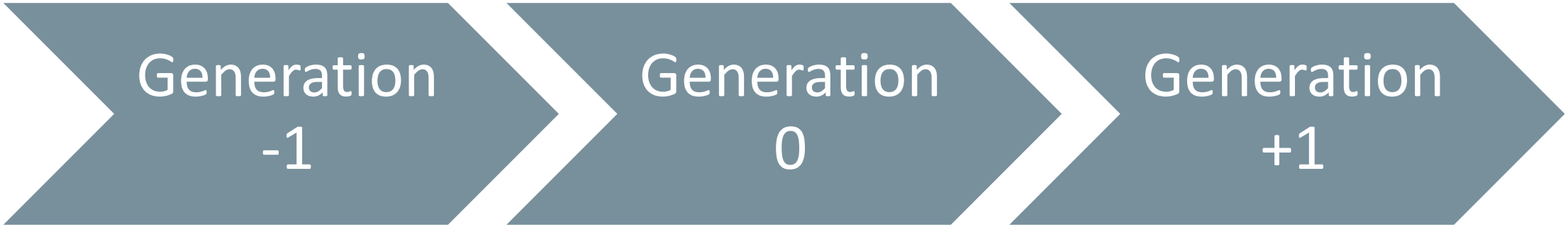
February 15, 2021

Katie Genadek
U.S. Census Bureau
University of Colorado – Boulder

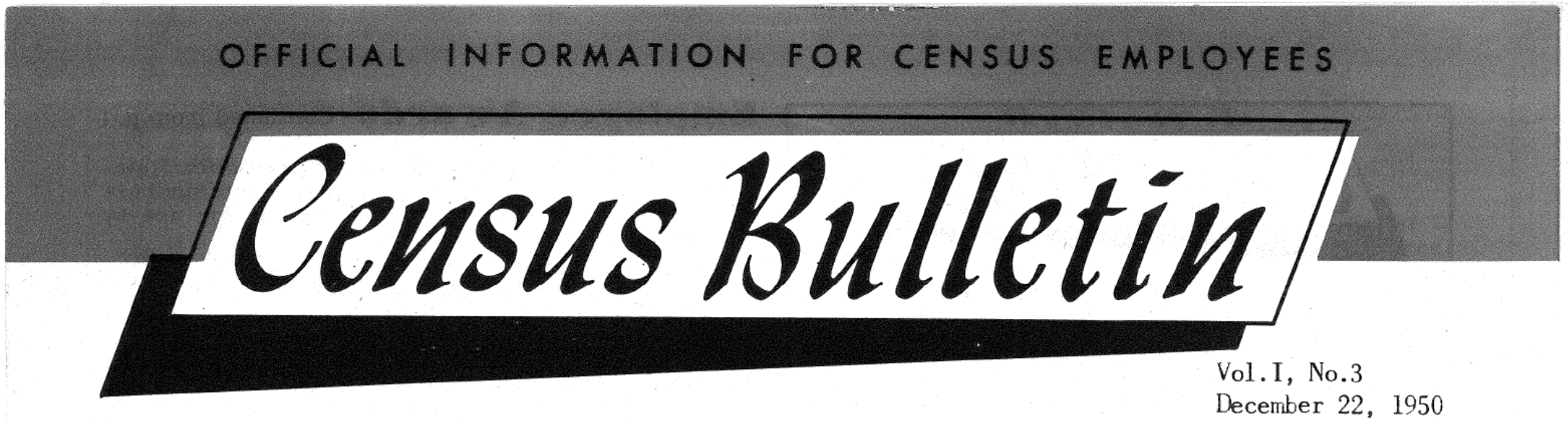
Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

Data needed to study social and economic mobility

For a representative population:



Long history of studying mobility



OFFICIAL INFORMATION FOR CENSUS EMPLOYEES

Census Bulletin

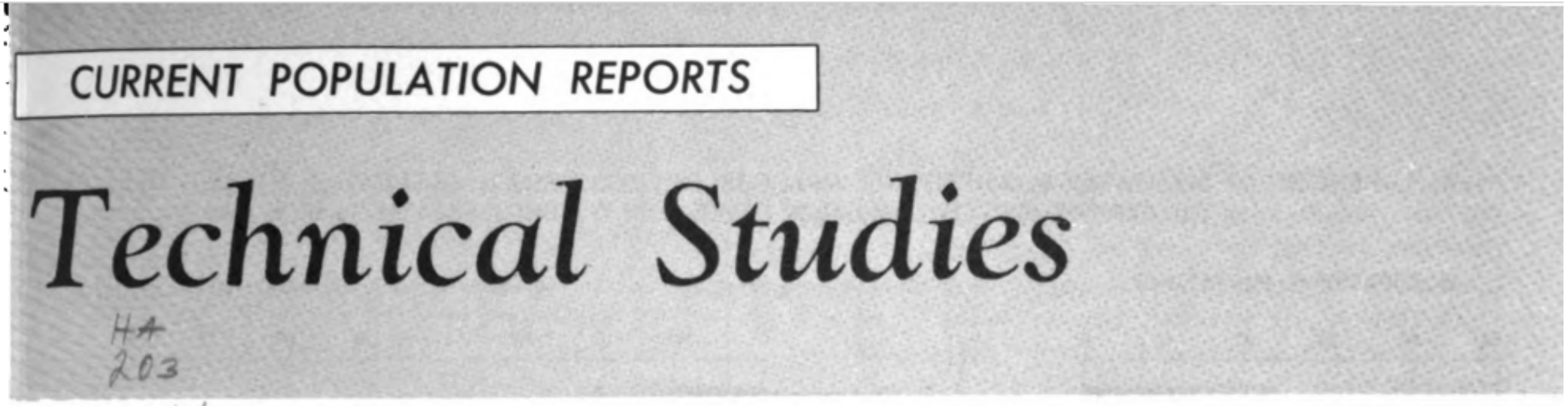
Vol. I, No. 3
December 22, 1950

Occupational Mobiliy Survey Is Planned

At the request of the Social Science Research Council, a private research organization, the Population and Field Divisions are making plans for an Occupational Mobility Survey to find out what kinds of jobs people have and how they have obtained them. This survey will start early in January in Chicago, St. Paul, Philadelphia, New Haven, Los Angeles and San Francisco. It is expected that 30 enumerators in each city will be able to complete the job by February 28.

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Long history of studying mobility



CURRENT POPULATION REPORTS

Technical Studies

HA 203

Series P-23, No. 11
May 12, 1964

LIFETIME OCCUPATIONAL MOBILITY OF ADULT MALES MARCH 1962

The proportion of men 25 to 64 years old whose occupation is in the same major occupation group as that of their fathers was not quite 1 in 4, i.e., 23 percent, in 1962. The occupational inheritance and occupational self-recruitment for a number of reasons including shifts in occupational composition over time, supply and demand factors, and differen-

Occupational Mobility

THE TREND OF OCCUPATIONAL MOBILITY IN THE UNITED STATES *

OTIS DUDLEY DUNCAN

University of Michigan

Inter- and intragenerational mobility matrices for 1962 may be used to infer the occupation distributions of certain cohorts in 1952, 1942, and 1932, on the assumption that patterns of mobility to these dates were the same, for men of comparable age, as the patterns observed in 1962. Comparison of the actual with the inferred distributions for the earlier years reveals net differences between recent and earlier patterns of mobility. The recent patterns feature more movement into salaried (though not self-employed) jobs as professional, technical, and kindred workers and salaried managers and officials, and less movement into manual and farm occupations. As of 1962 there was no immediate cause for concern that the American occupation structure was becoming more rigid. Projection of the present trend indefinitely far into the future is, however, not warranted.

Marriage Mobility

PATTERNS OF INTERGENERATIONAL MOBILITY OF FEMALES THROUGH MARRIAGE*

NORVAL D. GLENN
ADREAIN A. ROSS
JUDY CORDER TULLY

The University of Texas at Austin

American Sociological Review 1974, Vol. 39 (October): 683-99

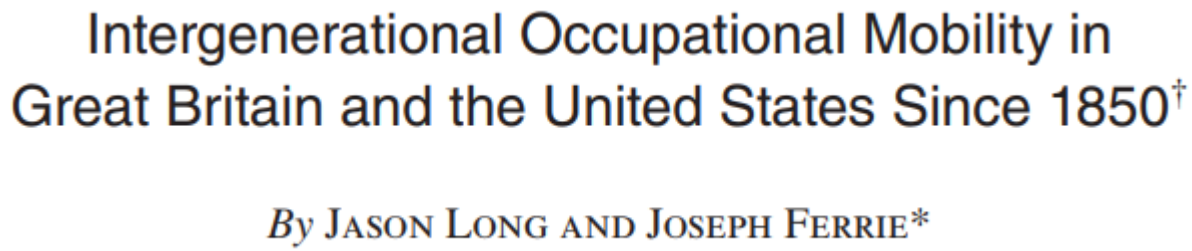
A comparison, based on data from four U.S. national surveys, of female mobility through marriage with male mobility through occupational attainment indicates (a) no tendency toward hypergamy not explained by the favorable balance of upward over downward mobility among males, (b) a weaker relationship of origin to destination among females, (c) a pronounced male mobility advantage among middle-origin persons, and (d) substantially more downward mobility into the manual-farm class by females than by males. Although the data indicate in general that the American status structure is more fluid than the male mobility data alone would indicate, there apparently is a condition of relative stasis among middle-origin females.

Income mobility

INTERGENERATIONAL EARNINGS MOBILITY IN THE UNITED STATES: SOME ESTIMATES AND A TEST OF BECKER'S INTERGENERATIONAL ENDOWMENTS MODEL

Jere Behrman and Paul Taubman*

Mobility research in part driven by available data



Intergenerational Occupational Mobility in
Great Britain and the United States Since 1850†

By JASON LONG AND JOSEPH FERRIE*

"Intergenerational occupational change was adopted as the metric for mobility for reasons of convenience as well: it is the only economic outcome that can be examined throughout the period since 1850. It is in some ways superior to income as a measure of mobility, and in some ways inferior."

But not completely…

Bringing Intergenerational Social Mobility Research into the Twenty-first Century: Why Mothers Matter

Emily Beller
U.S. Government Accountability Office

"This paper revisits intergenerational occupational mobility suspecting that previous estimates may have been too low due to the exclusion of mothers' occupations" (Hout 2018)

Issues with mobility surveys

- Small sample sizes
- Some limited geographic coverage
- Some are dated
- Measurement consistency
- Expensive to maintain
- New mobility surveys costly

Potential of administrative data

- Data already obtained
- Income data from tax information
- Relationship information from tax filings, Social Security program applications
- Near complete universe of population
- Full geographic coverage

Examples of administrative data use for social mobility

- Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review*, *104*(5), 141-47.
- Mitnik, P., Bryant, V., Weber, M., & Grusky, D. B. (2015). New estimates of intergenerational mobility using administrative data. *Statistics of Income working paper. Washington: Internal Revenue Service*.
- Larrimore, J., Mortenson, J., & Splinter, D. (2015). Income and earnings mobility in US tax data.

Potential of administrative data with linked survey data

Adds to administrative data:

- Demographic characteristics
- Socioeconomic variables
- Detailed family information

Additional outcomes:

- Occupation
- Educational attainment
- Housing

Examples of administrative data with linked survey data for mobility research

- Mazumder, B. (2014). Black–white differences in intergenerational economic mobility in the United States. *Economic Perspectives*, *38*(1).
- Akee, R., Jones, M. R., & Porter, S. R. (2019). Race matters: Income shares, income inequality, and income mobility for all US races. *Demography*, *56*(3), 999-1021.
- Song, X., Massey, C. G., Rolf, K. A., Ferrie, J. P., Rothbaum, J. L., & Xie, Y. (2020). Long-term decline in intergenerational mobility in the United States since the 1850s. *Proceedings of the National Academy of Sciences*, 117(1), 251-258.
- Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic opportunity in the United States: An intergenerational perspective. *The Quarterly Journal of Economics*, *135*(2), 711-783.

Examples of data often used

Administrative data

- IRS 1040, 1099, W2 data (1996-present)
- SSA Numident, Kidlink (Census Numident, Census Household Composition Key)

Survey data

- Decennial Census data (2000,2010)
- American Community Survey (2000-present)
- Current Population Survey (1973, 1979, 1981-1989, 1991-present)
- Survey of Income and Program Participation (1984-present)

Issues with administrative and linked survey data

- Access
- Governance, data sharing
- Information in data not collected for research
- Challenging to use data
- Coverage – representative population
- Recent generations

Coverage: availability and gaps

Easiest to link:

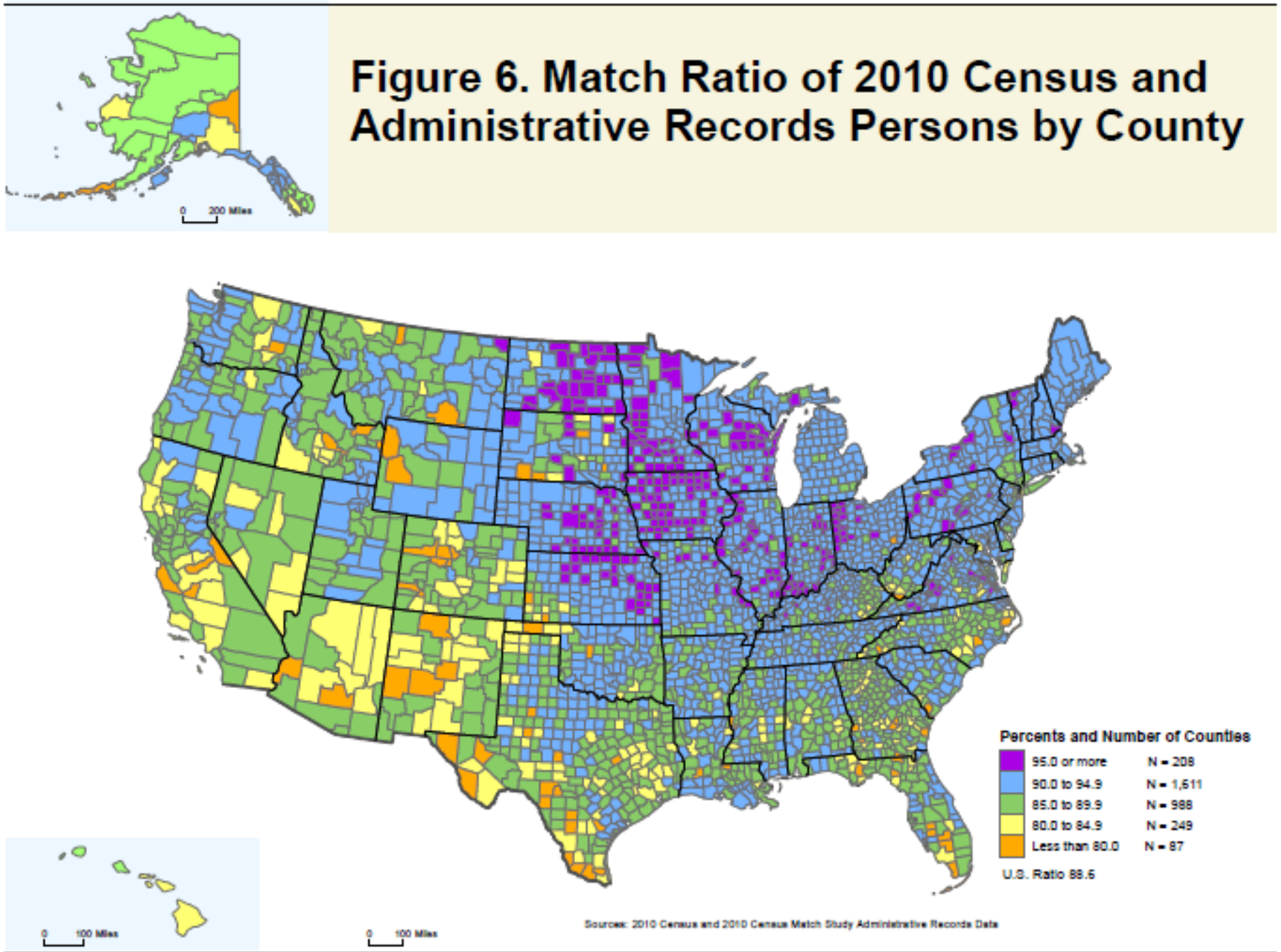
- Taxpayers
- Workers
- Persons on Medicare (most of the 65+ population)
- Persons receiving housing/medical/food assistance
- Men who ever registered for Selective Service

Harder to link:

- Mobile and transient populations
- Residents without SSN
- Children not claimed on tax forms

17

Figure 6. Match Ratio of 2010 Census and Administrative Records Persons by County



United States Census Bureau

Only recent generations

- Late 1990s, 2000 for children with parents
- Some smaller samples in 1970s and 1980s
- Links to historic public census data from 1940 and prior

- One solution: building out the Census Bureau's Data Linkage Infrastructure

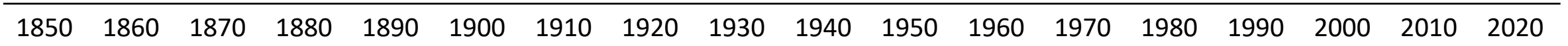
Representative samples linked to lifetime SSA earnings data (originally from IRS)

- Current Population Survey (1973, 1979, 1981-1989, 1991-present)
- Survey of Income and Program Participation (1984-present)
- American Community Survey (2005-present)

Available large-scale U.S. Census linkages

2000 census
2010 census
ACS

IPUMS Linked Samples
NBER Census Linking Project
1850-1940



Available large-scale U.S. Census

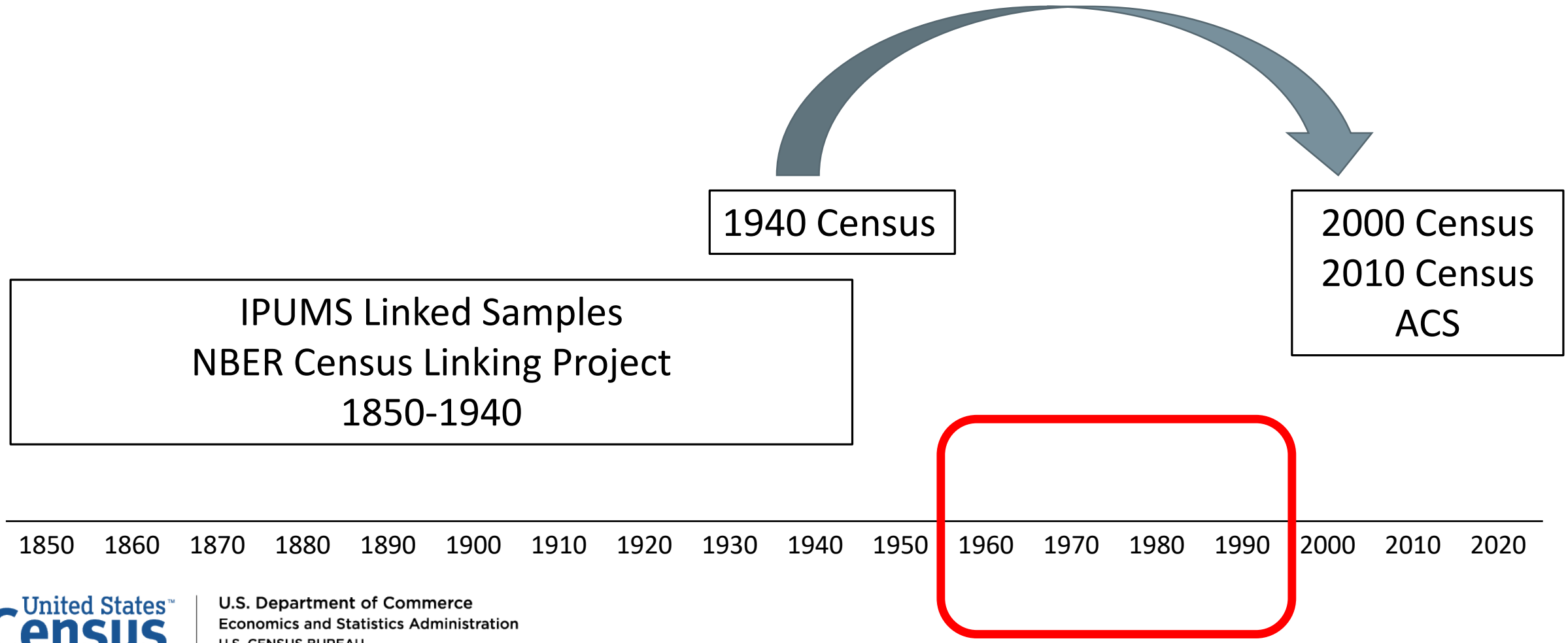
1940 Census

2000 Census
2010 Census
ACS

IPUMS Linked Samples
NBER Census Linking Project
1850-1940

1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020

Available large-scale U.S. Census linkages



Why is this difficult?

- Microdata files for 1960-1990 exist but do not include names
- Names were handwritten on census forms
- Census forms are stored on 250,000 microfilm reels
- Forms and data are highly restricted to protect respondent confidentiality

1990 Name Recovery Pilot

- Supported by NSF, National Academies, Carnegie Foundation – American Opportunity Study
- Can names be recovered from historical censuses in cost effective manor?
- Can names be recovered of high enough quality to perform linkage?

1990 Name Recovery Pilot

How often did handwriting recognition results match the truth data?

	exact match	very similar*	n
Age	95%		8,900
Last Name	82%	90%	8,850
First Name	75%	86%	8,850

*Score of >0.85 on the Jaro-Winkler string comparator

*All results were approved for release by the Census Bureau's Disclosure Review Board (CBDRB-FY21-ERD002-022).

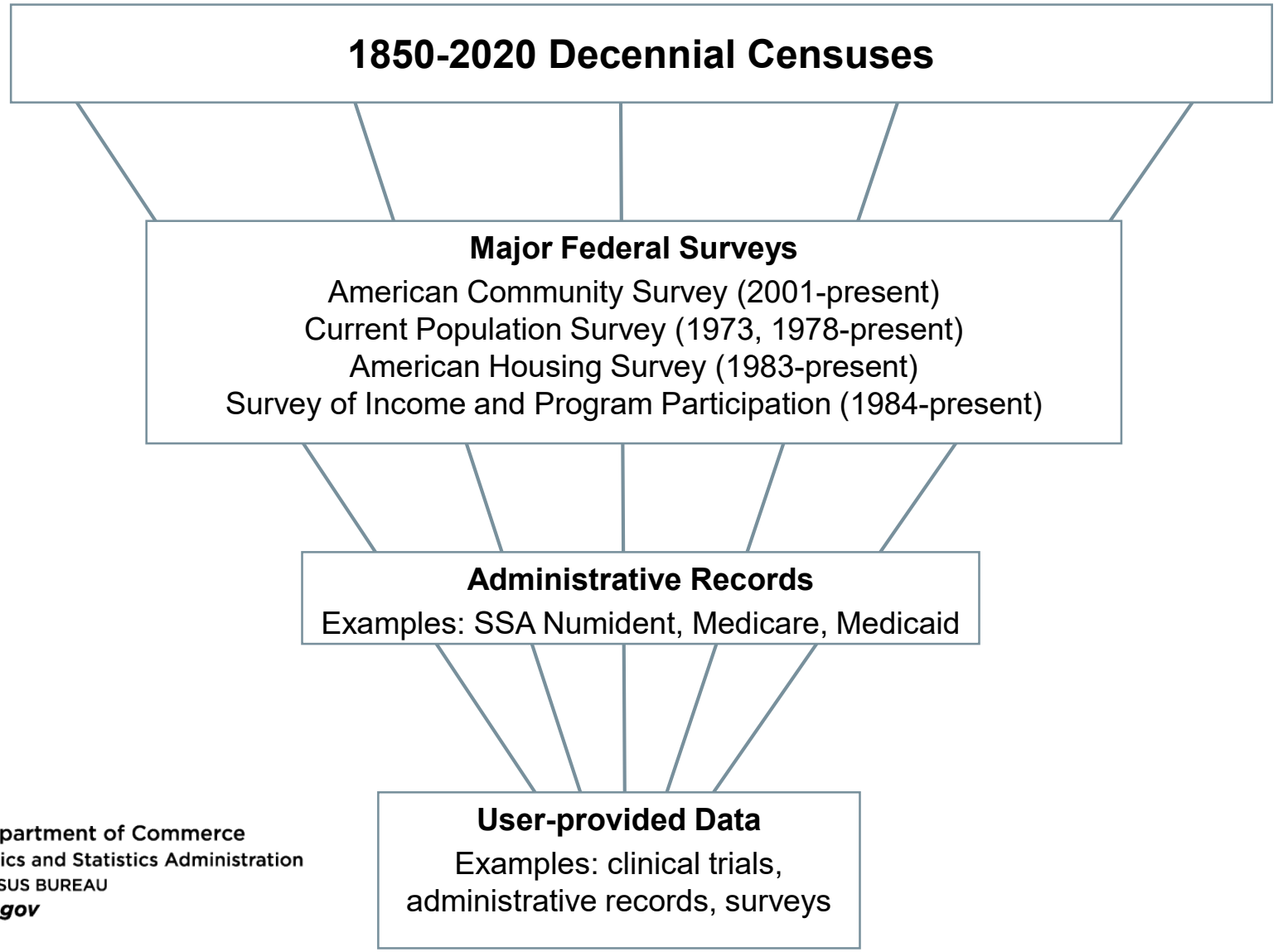
Decennial Census Digitization and Linkage Project

1. Scan the microfilm reels with images from 1960-1990 Censuses into digital images
2. Use OCR and OMR to digitize the PII and fields from the scanned images
3. Link the PII to the already digitized data
4. Create linkage key crosswalks for 1960-1990

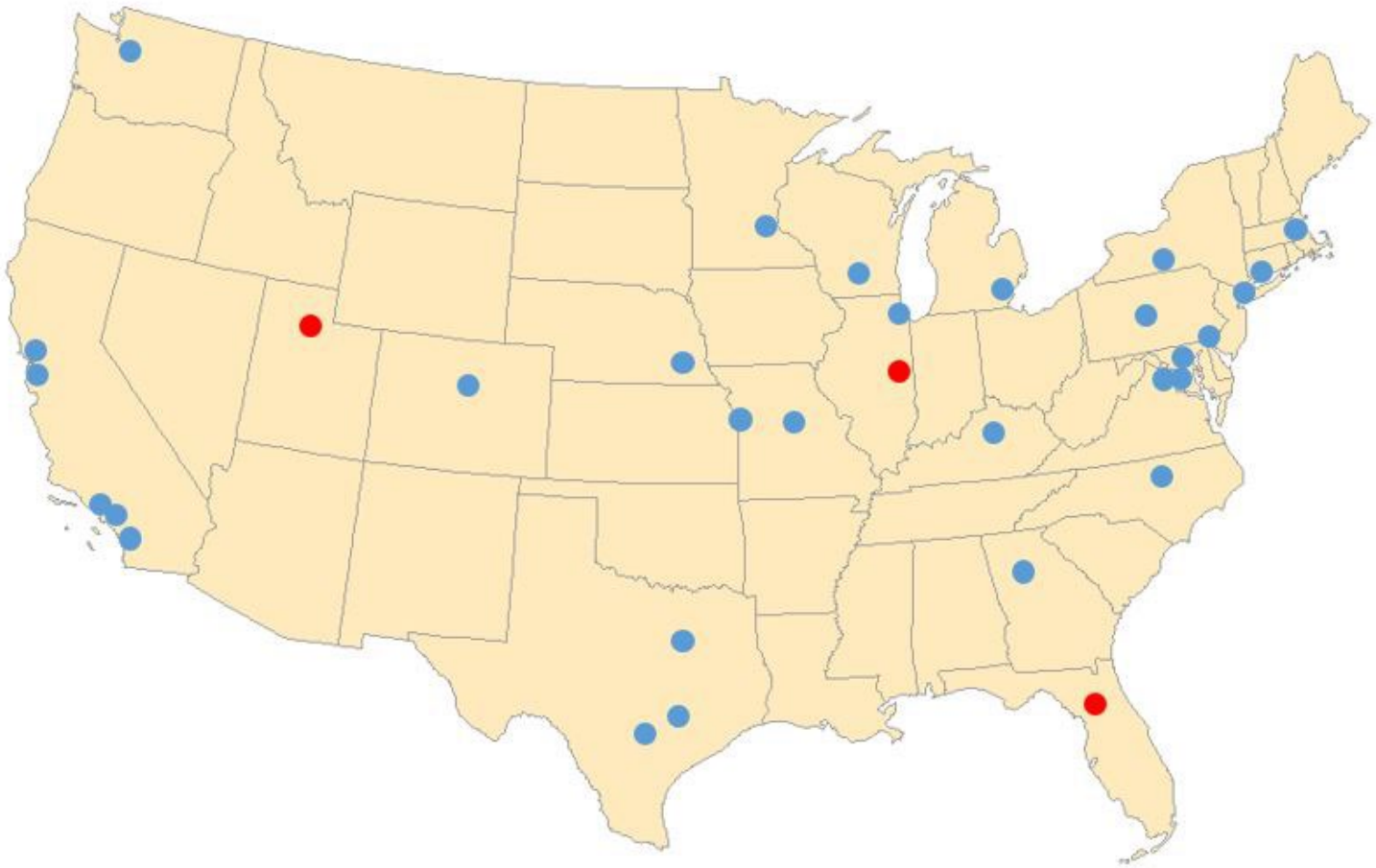
Support for this project

- Additional support of for the feasibility of the project from Stanford University and the Hewlett Foundation
- Census Bureau support and funds
- Funding from the Ballmer Group, the Overdeck Foundation, Arnold Ventures, the Gates Foundation, JPB Foundation, National Science Foundation
- Partnered with the University of Michigan (ICPSR), Brown University, Harvard University (Opportunity Insights)

Expected Longitudinal Infrastructure by 2026



Federal Statistical Research Data Centers



United States Census Bureau

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

Questions: Katie Genadek
katie.r.genadek@census.gov

Questions?

Katie Genadek

katie.r.genadek@census.gov