

# Decennial Census Digitization and Linkage Project

August 16, 2024

**Katie Genadek**

U.S. Census Bureau

University of Colorado



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

# Decennial Census Digitization and Linkage Project

- Project motivation
- Background
- Digitization
- Capture
- Linkage
- Upcoming related work
- Data access

# Motivation

# U.S. Census Bureau

Mission: The Census Bureau's mission is to serve as the nation's leading provider of quality data about its people and economy

# Title 13 U.S.C.

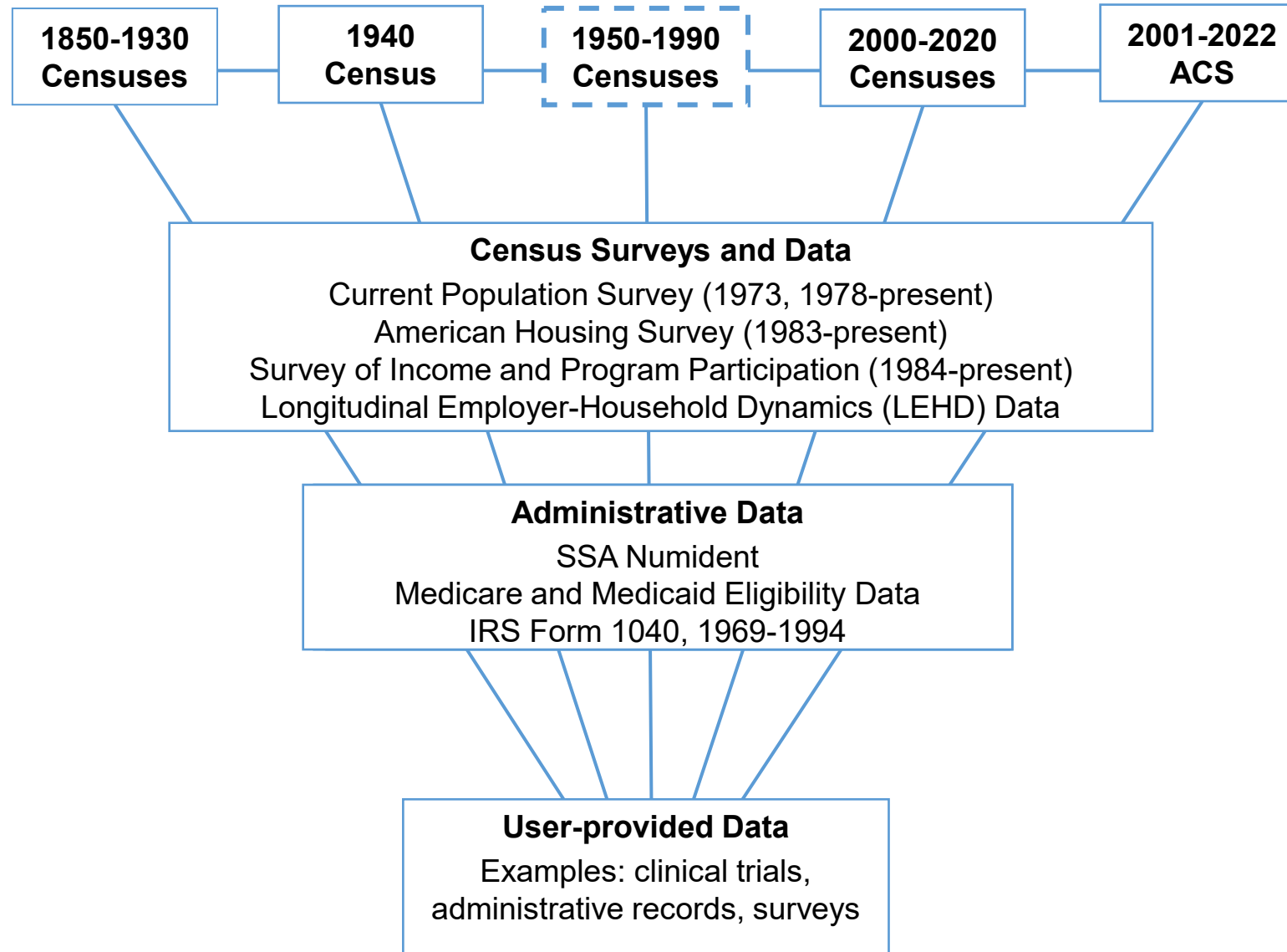
- Title 13 protects the data we collect at the Census Bureau
- It is against the law to disclose or publish any private information that identifies an individual or business, including names
- All systems, software must be approved and protected

# Research partnerships

- Federal Statistical Research Data Centers (FSRDC) are partnerships between federal statistical agencies and leading research institutions
- FSRDCs provide secure environments supporting qualified researchers using restricted-access data while protecting respondent confidentiality



# Data Linkage Infrastructure



# Digitized full-count U.S. Census data, linked over time

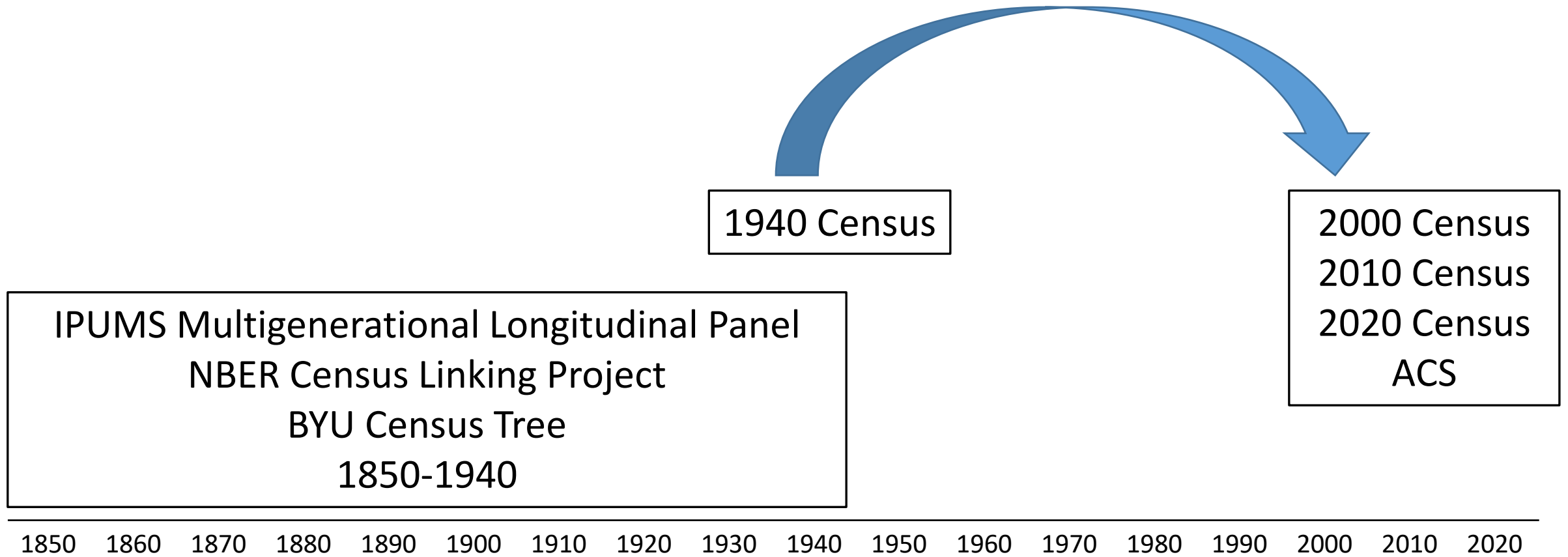
IPUMS Multigenerational Longitudinal Panel  
NBER Census Linking Project  
BYU Census Tree  
1850-1940

2000 Census  
2010 Census  
2020 Census  
ACS

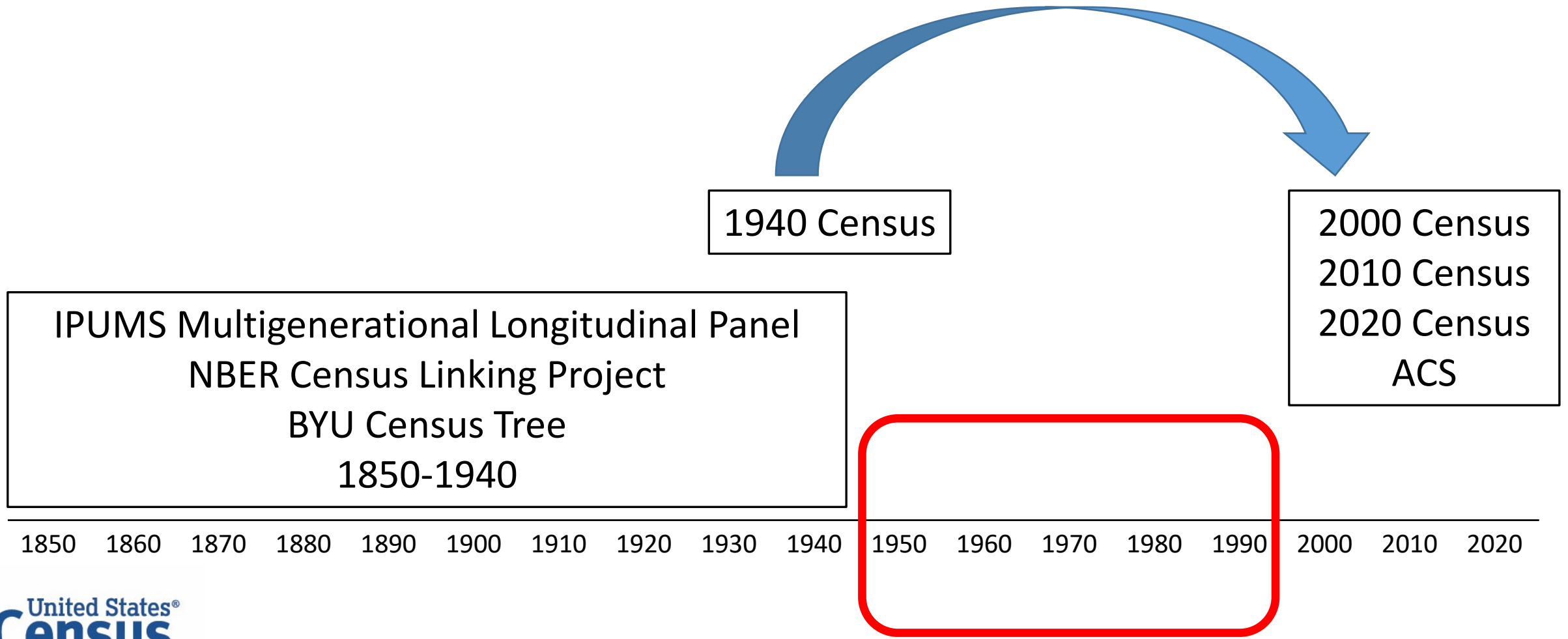
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020



# Digitized full-count U.S. Census data, linked over time



# Digitized full-count U.S. Census data, linked over time



OFFICIAL INFORMATION FOR CENSUS EMPLOYEES

# *Census Bulletin*

Vol. I, No. 3

December 22, 1950

## **Occupational Mobility 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.

# Long history of studying mobility

CURRENT POPULATION REPORTS

## Technical Studies

HA  
203

LA 14  
11-25

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

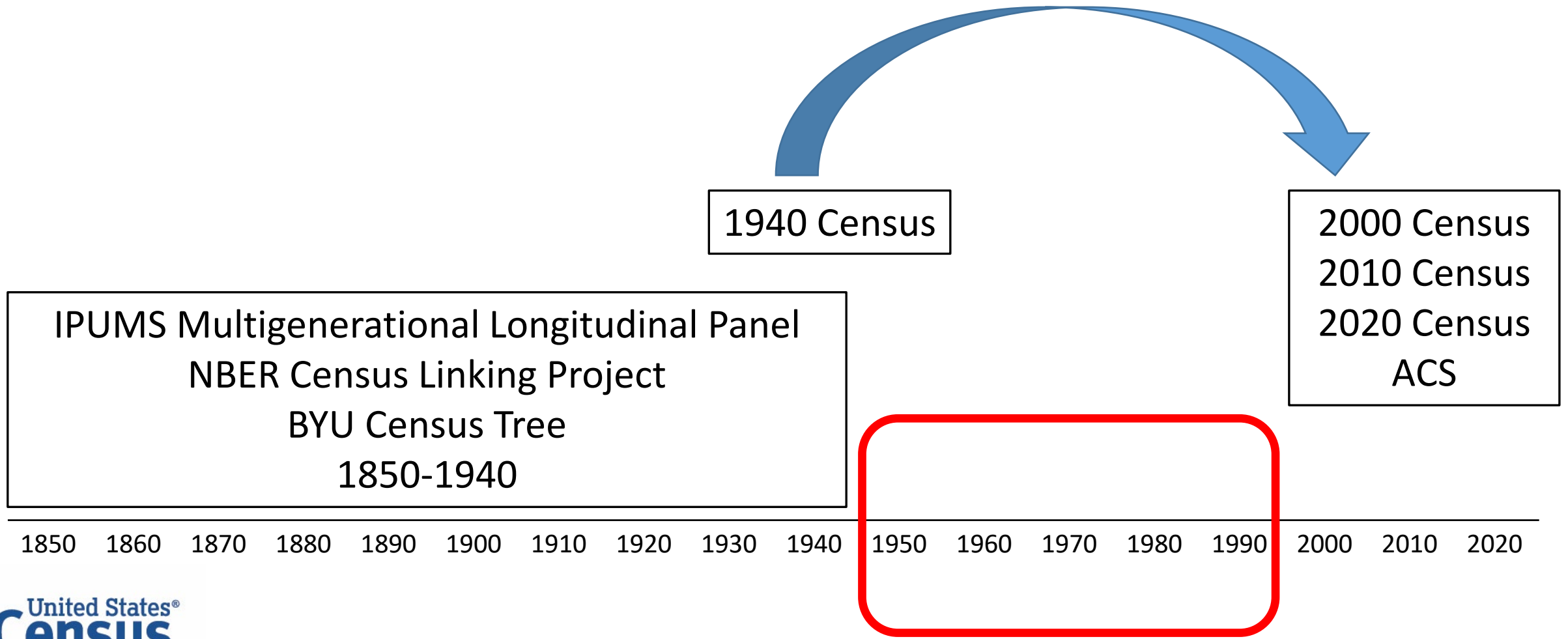
### LIFETIME OCCUPATIONAL MOBILITY OF ADULT MALES MARCH 1962

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-

# DCDL Background

# Digitized full-count U.S. Census data, linked over time



# Why are the 1950 data missing?

- Microdata for 1950 only exist for 1% of the population
- Census images were made public in April 2022
- Full-count microdata became available in 2024
- Will be linked to modern data by 2026











# CENSUS BULLETIN

U.S. Department of Commerce  
BUREAU OF THE CENSUS

**VOL. XXVIII No. 7**  
**August-September 1978**

## Q&A

### Regs. Prohibit Personal Calls

**Q.** What are Federal regulations regarding phone calls? Several of my co-workers routinely call friends or relatives long distance on office phones and this strikes me as a serious abuse.

**ASD answers.** Federal Property Management Regulation 101-35.307-1(b) states that agencies should provide only telephone "equipment necessary to carry out assigned missions." Chapter K-12, Census Administrative Manual, limits the use of government telephones to official business only—and prohibits this use for personal calls.

## Age Search 75 Years Young

The Bureau's Age Search service is 75 years old and going strong.

"The operation started when the Congress allowed the Bureau to charge a small fee to anyone who requested a search of census records for proof of age, race, or place of birth and residence," said Bureau Historian Fred Bohme (DUSD). Census records can also be used to prove citizenship, relationship, and occupation.

"People who have little or no documentation on their past need such help. Until 1920, a number of States did not have standard birth registration procedures," Bohme noted.

However, the public was slow to utilize this resource provided by the Bureau. In 1909, six years after the start of the



*FRANK Allai, Chief, Personal Census Service Branch, checks documents with secretary Kay Meyer.*

MEMBERS OF THE  
Microfilm Search-  
ing Unit examine  
census schedules.





# Microfilm Numbers

Year	Reels	Images
1960	11,745	21,141,000
1970	37,250	67,050,000
1980	92,250	166,050,000
1990	115,000	207,000,000
	256,245	461,241,000

# Why are the 1960-1990 data missing?

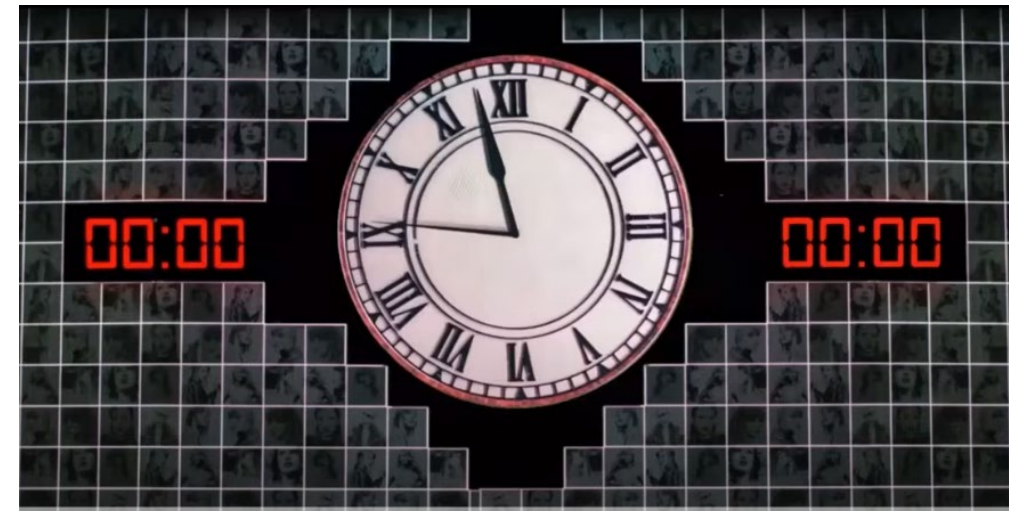
- Microdata files for 1960-1990 exist but do not include names
- Names were handwritten on census forms
- Census forms are stored on 250,000 microfilm reels
- Forms and data for 1960-1990 are protected by Title 13

# DCDL has been a long time coming...

## DCDL feasibility research (2013-2019)

- National Academies panel on intergenerational mobility
- Demonstrated methods for scanning, OCR, record linkage

## DCDL Full production work (2020-2026)





# Decennial Census Digitization and Linkage Project

1. **Digitizing** 250,000 reels of microfilm
2. **Capturing** handwritten names and some bubbled data
3. **Merging** names with original microdata files
4. **Linking** all possible records over time
5. **Sharing** linked data securely

# Decennial Census Digitization and Linkage Project

Support for feasibility research, 2013-2019



NATIONAL ACADEMIES  
Sciences  
Engineering  
Medicine



Support for full-production work, 2020-2026



BILL & MELINDA  
GATES *foundation*



ICPSR

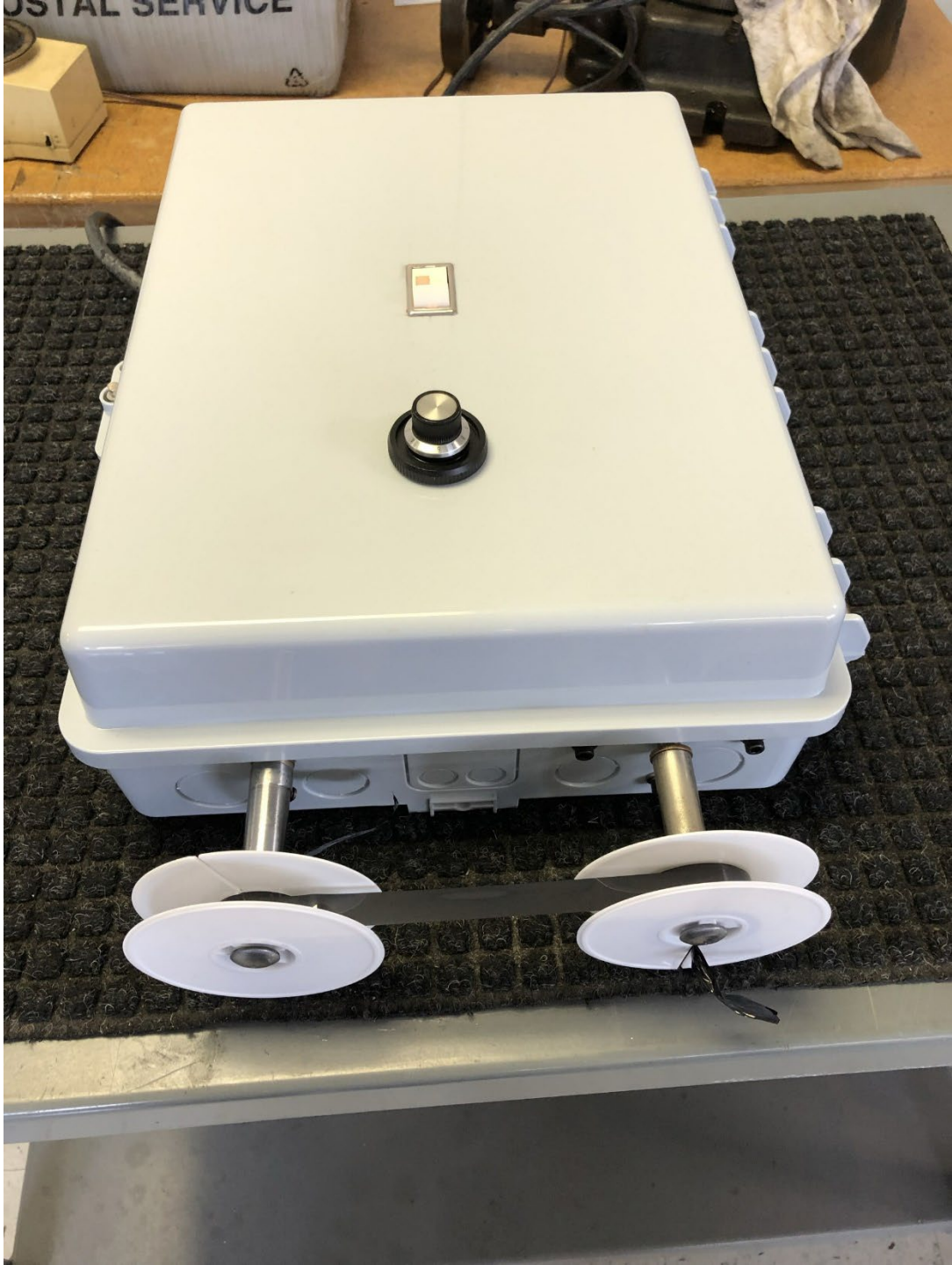


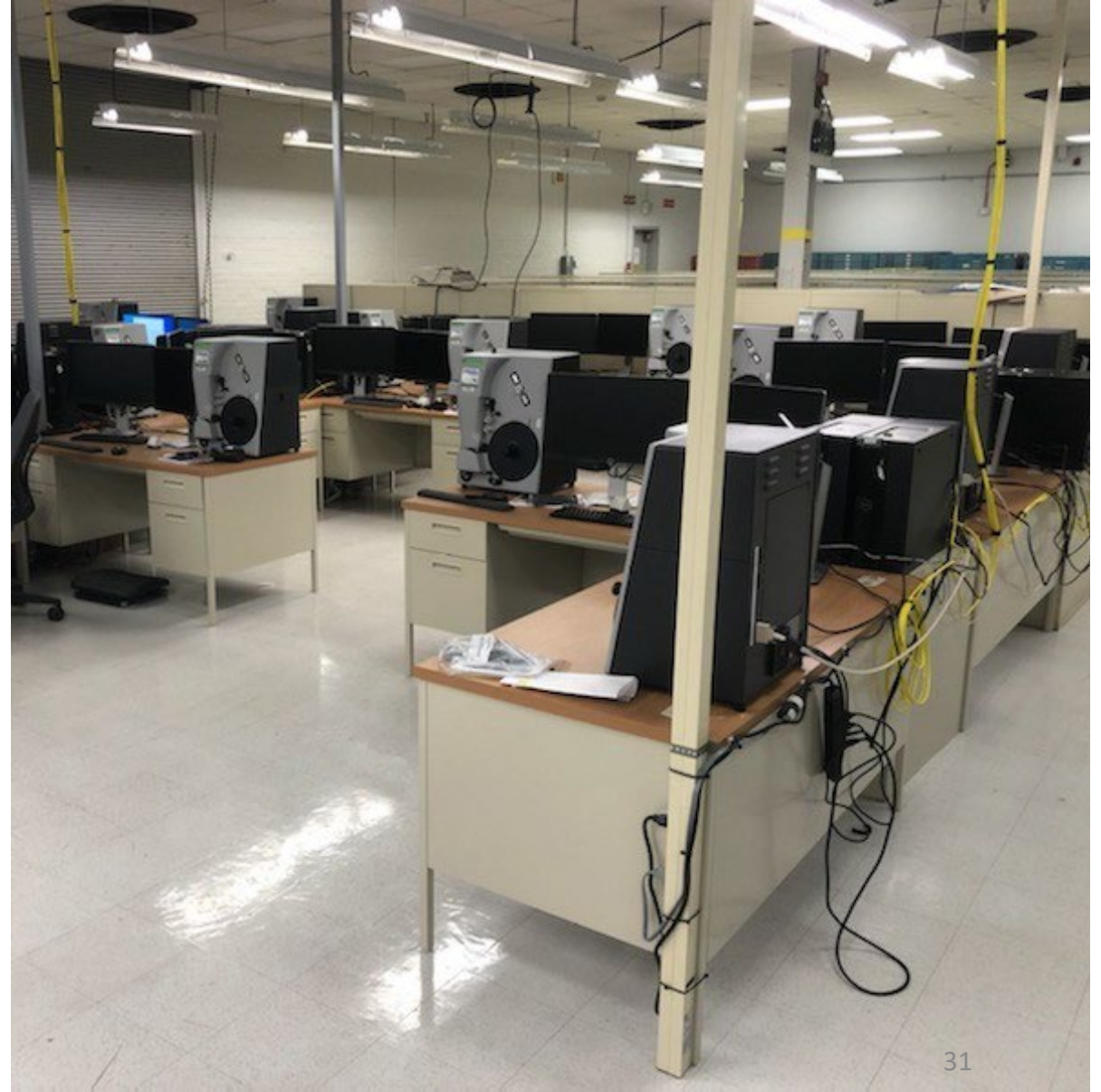
# Digitization

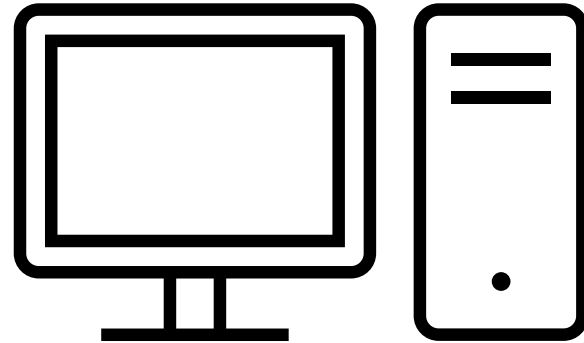
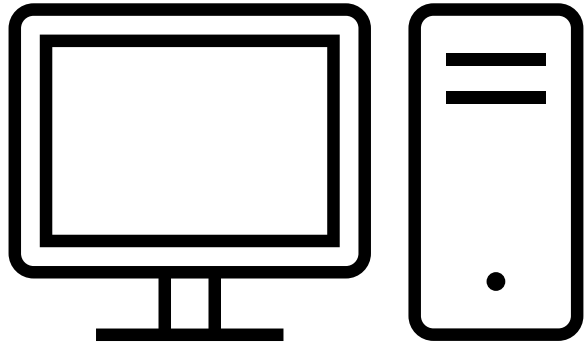
# Digitization: Microfilm Scanning

- Digital scanning at the Census Bureau's National Processing Center (NPC)
- Reel preparation includes unwinding, attaching trailers, and rewinding all of 1970 reels and various other reels











PLEASE ALSO ANSWER HOUSING QUESTIONS ON PAGE 3

PLEASE ALSO ANSWER HOUSING QUESTIONS ON PAGE 3

	PERSON 1	PERSON 2	PERSON 3	PERSON 4	PERSON 5	PERSON 6
1. Please fill one column for each person listed in Question 1a on page 1.	Last name: _____ First name: _____ Middle initial: _____					
2. How is this person related to PERSON 1? Fill ONE circle for each person. If Other relative of person in column 1, fill circle and print exact relationship, such as mother-in-law, grandparent, son-in-law, niece, cousin, and so on.	<input type="radio"/> Husband/wife <input type="radio"/> Brother/sister <input type="radio"/> Natural-born or adopted son/daughter <input type="radio"/> Father/mother or Grandchild <input type="radio"/> Other relative <input type="checkbox"/> <input type="radio"/> Stepson/stepdaughter <input type="radio"/> Roomer, boarder, or foster child <input type="radio"/> Unmarried partner <input type="radio"/> Housemate, roommate <input type="radio"/> Other nonrelative					
3. Sex Fill ONE circle for each person.	<input type="radio"/> Male <input type="radio"/> Female					
4. Race Fill ONE circle for the race that the person considers himself/herself to be. If Indian (Amer.), print the name of the enrolled or principal tribe. If Other Asian or Pacific Islander (API), print one group, for example: Fijian, Laotian, Thai, Tongan, Pakistani, Cambodian, and so on. If Other race, print race.	<input type="radio"/> White <input type="radio"/> Black or Negro <input type="radio"/> Indian (Amer.) (Print the name of the enrolled or principal tribe) _____ <input type="radio"/> Eskimo <input type="radio"/> Aleut <input type="radio"/> Asian or Pacific Islander (API) _____ <input type="radio"/> Chinese <input type="radio"/> Japanese <input type="radio"/> Filipino <input type="radio"/> Asian Indian <input type="radio"/> Hawaiian <input type="radio"/> Samoan <input type="radio"/> Korean <input type="radio"/> Guamanian <input type="radio"/> Vietnamese <input type="radio"/> Other API _____ <input type="radio"/> Other race (Print race) _____					
5. Age and year of birth a. Print each person's age at last birthday. Fill in the matching circle below each box. b. Print each person's year of birth and fill the matching circle below each box.	a. Age: 0 0 0 0 0 1 8 0 0 0 0 0 1 9 0 1 0 1 0 b. Year of birth: 1 8 0 0 0 0 0 0 1 9 0 1 0 1 0					
6. Marital status Fill ONE circle for each person.	<input type="radio"/> Now married <input type="radio"/> Separated <input type="radio"/> Widowed <input type="radio"/> Never married <input type="radio"/> Divorced					
7. Is this person Spanish/Hispanic origin? Fill ONE circle for each person. If Yes, other Spanish/Hispanic, print one group.	<input type="radio"/> No (not Spanish/Hispanic) <input type="radio"/> Yes, Mexican, Mexican-Am., Chicano <input type="radio"/> Yes, Puerto Rican <input type="radio"/> Yes, Cuban <input type="radio"/> Yes, other Spanish/Hispanic (Print one group, for example: Argentinian, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.) _____					
FOR CENSUS USE	2958 500 2271 00,647,312					

NOW PLEASE ANSWER QUESTIONS H1a-H7b FOR YOUR HOUSEHOLD

<p>PERSON 7</p> <p>Last name: _____ First name: _____ Middle initial: _____</p> <p>If a RELATIVE of Person 1:</p> <input type="radio"/> Husband/wife <input type="radio"/> Brother/sister <input type="radio"/> Natural-born or adopted son/daughter <input type="radio"/> Father/mother or Grandchild <input type="radio"/> Other relative <input type="checkbox"/> <input type="radio"/> Stepson/stepdaughter <input type="radio"/> Roomer, boarder, or foster child <input type="radio"/> Unmarried partner <input type="radio"/> Housemate, roommate <input type="radio"/> Other nonrelative	<p>H1a. Did you leave anyone out of your list of persons for Question 1a on page 1 because you were not sure if the person should be listed — for example, someone temporarily away on a business trip or vacation, a newborn baby still in the hospital, or a person who stays here once in a while and has no other home?</p> <input type="checkbox"/> Yes, please print the name(s) _____ and reason(s): _____ <input type="checkbox"/> No
<p>If a RELATIVE of Person 1:</p> <input type="radio"/> Husband/wife <input type="radio"/> Brother/sister <input type="radio"/> Natural-born or adopted son/daughter <input type="radio"/> Father/mother or Grandchild <input type="radio"/> Other relative <input type="checkbox"/> <input type="radio"/> Stepson/stepdaughter <input type="radio"/> Roomer, boarder, or foster child <input type="radio"/> Unmarried partner <input type="radio"/> Housemate, roommate <input type="radio"/> Other nonrelative	<p>H1b. Did you include anyone in your list of persons for Question 1a on page 1 even though you were not sure that the person should be listed — for example, a visitor who is staying here temporarily or a person who usually lives somewhere else?</p> <input type="checkbox"/> Yes, please print the name(s) _____ and reason(s): _____ <input type="checkbox"/> No
<p>If a RELATIVE of Person 1:</p> <input type="radio"/> Husband/wife <input type="radio"/> Brother/sister <input type="radio"/> Natural-born or adopted son/daughter <input type="radio"/> Father/mother or Grandchild <input type="radio"/> Other relative <input type="checkbox"/> <input type="radio"/> Stepson/stepdaughter <input type="radio"/> Roomer, boarder, or foster child <input type="radio"/> Unmarried partner <input type="radio"/> Housemate, roommate <input type="radio"/> Other nonrelative	<p>H2. Which best describes this building? Include all apartments, flats, etc., even if vacant.</p> <input type="checkbox"/> A mobile home or trailer <input type="checkbox"/> A one-family house detached from any other house <input type="checkbox"/> A one-family house attached to one or more houses <input type="checkbox"/> A building with 2 apartments <input type="checkbox"/> A building with 3 or 4 apartments <input type="checkbox"/> A building with 5 to 9 apartments <input type="checkbox"/> A building with 10 to 19 apartments <input type="checkbox"/> A building with 20 to 49 apartments <input type="checkbox"/> A building with 50 or more apartments <input type="checkbox"/> Other _____
<p>a. Age: 0 0 0 0 0 1 8 0 0 0 0 0 1 9 0 1 0 1 0            b. Year of birth: 1 8 0 0 0 0 0 0 1 9 0 1 0 1 0</p>	<p>H3. How many rooms do you have in this house or apartment? Do NOT count bathrooms, porches, balconies, foyers, halls, or half-rooms.</p> <input type="checkbox"/> 1 room <input type="checkbox"/> 2 rooms <input type="checkbox"/> 3 rooms <input type="checkbox"/> 4 rooms <input type="checkbox"/> 5 rooms <input type="checkbox"/> 6 rooms <input type="checkbox"/> 7 rooms <input type="checkbox"/> 8 rooms <input type="checkbox"/> 9 or more rooms
<p>FOR CENSUS USE</p>	<p>H4. Is this house or apartment —</p> <input type="checkbox"/> Owned by you or someone in this household with a mortgage or loan? <input type="checkbox"/> Owned by you or someone in this household free and clear (without a mortgage)? <input type="checkbox"/> Rented for cash rent? <input type="checkbox"/> Occupied without payment of cash rent?
<p>A. Total persons: _____</p> <p>B. Type of unit: <input type="radio"/> First form <input type="radio"/> Regular <input type="radio"/> Cont'n <input type="radio"/> Usual home elsewhere</p> <p>C1. Vacancy status: <input type="radio"/> For rent only <input type="radio"/> For sale only <input type="radio"/> Rented or sold, not occupied <input type="radio"/> Other vacant</p> <p>C2. Is this unit boarded up? <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>D. Months vacant: <input type="checkbox"/> Less than 1 <input type="checkbox"/> 1 up to 2 <input type="checkbox"/> 2 up to 6 <input type="checkbox"/> 6 up to 12 <input type="checkbox"/> 12 up to 24 <input type="checkbox"/> 24 or more</p> <p>E. Complete after: <input type="checkbox"/> LR <input type="checkbox"/> TC <input type="checkbox"/> QA <input type="checkbox"/> P/F <input type="checkbox"/> RE <input type="checkbox"/> I/T <input type="checkbox"/> MV <input type="checkbox"/> ED <input type="checkbox"/> EN <input type="checkbox"/> P0 <input type="checkbox"/> P3 <input type="checkbox"/> P6 <input type="checkbox"/> P1 <input type="checkbox"/> P4 <input type="checkbox"/> IA <input type="checkbox"/> P2 <input type="checkbox"/> P5 <input type="checkbox"/> SM</p> <p>F. Cov. <input type="checkbox"/> 1b <input type="checkbox"/> 1a <input type="checkbox"/> 7 <input type="checkbox"/> H1</p>
<p>H5a. Is this house on ten or more acres? <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>H5b. Is there a business (such as a store or barber shop) or a medical office on this property? <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>H6. What is the value of this property? that is, how much do you think this house and lot or condominium unit would sell for if it were for sale?</p> <p>Answer only if you own or someone in this household OWNS OR IS BUYING this house or apartment —</p> <p>Less than \$10,000 <input type="checkbox"/> \$10,000 to \$14,999 <input type="checkbox"/> \$15,000 to \$19,999 <input type="checkbox"/> \$20,000 to \$24,999 <input type="checkbox"/> \$25,000 to \$29,999 <input type="checkbox"/> \$30,000 to \$34,999 <input type="checkbox"/> \$35,000 to \$39,999 <input type="checkbox"/> \$40,000 to \$44,999 <input type="checkbox"/> \$45,000 to \$49,999 <input type="checkbox"/> \$50,000 to \$54,999 <input type="checkbox"/> \$55,000 to \$59,999 <input type="checkbox"/> \$60,000 to \$64,999 <input type="checkbox"/> \$65,000 to \$69,999 <input type="checkbox"/> \$70,000 to \$74,999 <input type="checkbox"/> \$75,000 to \$79,999 <input type="checkbox"/> \$80,000 to \$89,999 <input type="checkbox"/> \$90,000 to \$99,999 <input type="checkbox"/> \$100,000 to \$124,999 <input type="checkbox"/> \$125,000 to \$149,999 <input type="checkbox"/> \$150,000 to \$174,999 <input type="checkbox"/> \$175,000 to \$199,999 <input type="checkbox"/> \$200,000 to \$249,999 <input type="checkbox"/> \$250,000 to \$299,999 <input type="checkbox"/> \$300,000 to \$399,999 <input type="checkbox"/> \$400,000 to \$499,999 <input type="checkbox"/> \$500,000 or more</p> <p>H7a. What is the monthly rent? Answer only if you PAY RENT for this house or apartment —</p> <p>Less than \$80 <input type="checkbox"/> \$80 to \$99 <input type="checkbox"/> \$100 to \$124 <input type="checkbox"/> \$125 to \$149 <input type="checkbox"/> \$150 to \$174 <input type="checkbox"/> \$175 to \$199 <input type="checkbox"/> \$200 to \$224 <input type="checkbox"/> \$225 to \$249 <input type="checkbox"/> \$250 to \$274 <input type="checkbox"/> \$275 to \$299 <input type="checkbox"/> \$300 to \$324 <input type="checkbox"/> \$325 to \$349 <input type="checkbox"/> \$350 to \$374 <input type="checkbox"/> \$375 to \$399 <input type="checkbox"/> \$400 to \$424 <input type="checkbox"/> \$425 to \$449 <input type="checkbox"/> \$450 to \$474 <input type="checkbox"/> \$475 to \$499 <input type="checkbox"/> \$500 to \$524 <input type="checkbox"/> \$525 to \$549 <input type="checkbox"/> \$550 to \$599 <input type="checkbox"/> \$600 to \$649 <input type="checkbox"/> \$650 to \$699 <input type="checkbox"/> \$700 to \$749 <input type="checkbox"/> \$750 to \$999 <input type="checkbox"/> \$1,000 or more</p> <p>H7b. Does the monthly rent include any meals? <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	

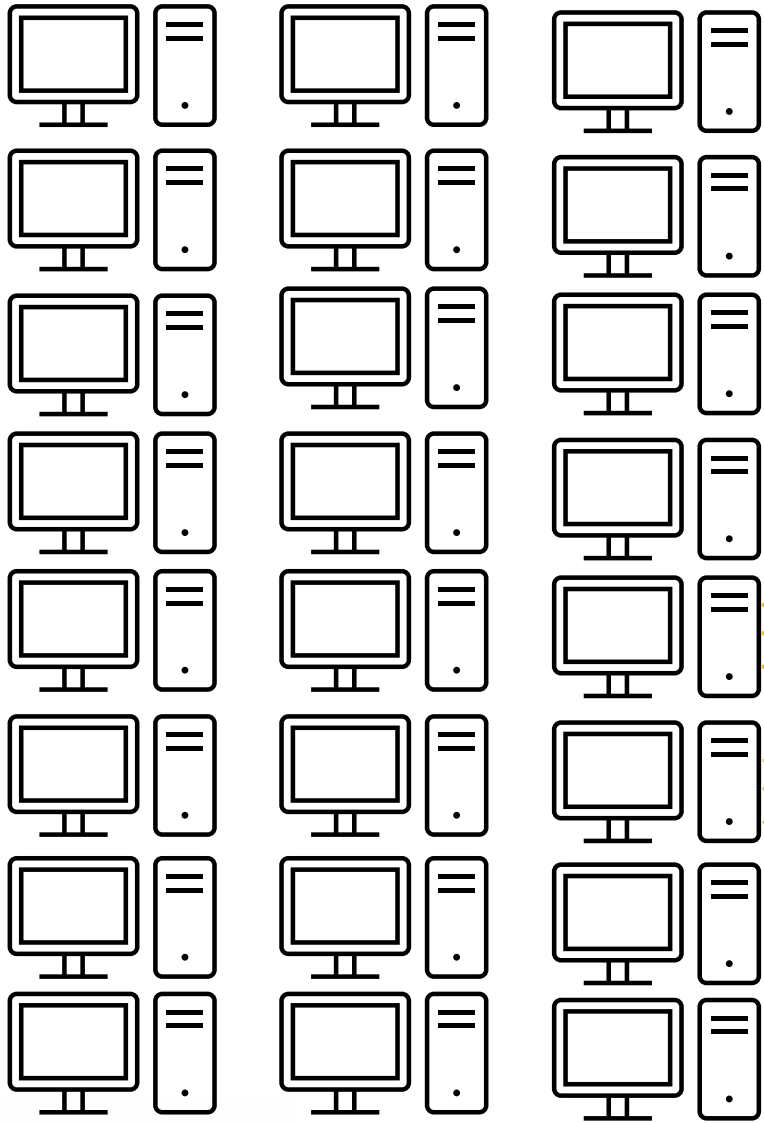


# Microfilm reel scanning specifics

- Each reel has **~1800 images**
- Each scanning machine can scan about **15-20 reels** per day
- Creates one gray scale .jpg and one black and white .tif of each image
- Gray scale reel: ~12 Gb
- Black and white reel: ~.16 Gb
- Producing **5-6 Tb digital images per day**
- ~250,000 reels, 460 million images, 700 million names
- Estimating nearly **4 Pb of images total**

# Image Storage

- Move scans to 4 Pb Dell Isilon Network Attached Storage (NAS)
- Back up to digital tape in facility in Maryland
- Created storage organization for live scanning and back ups
- Perform technical quality assurance and scanning of images



United States<sup>®</sup>  
**Census**  
Bureau



# Project Status

- Follow our progress on the DCDL website: [www.census.gov/dcdl](http://www.census.gov/dcdl)

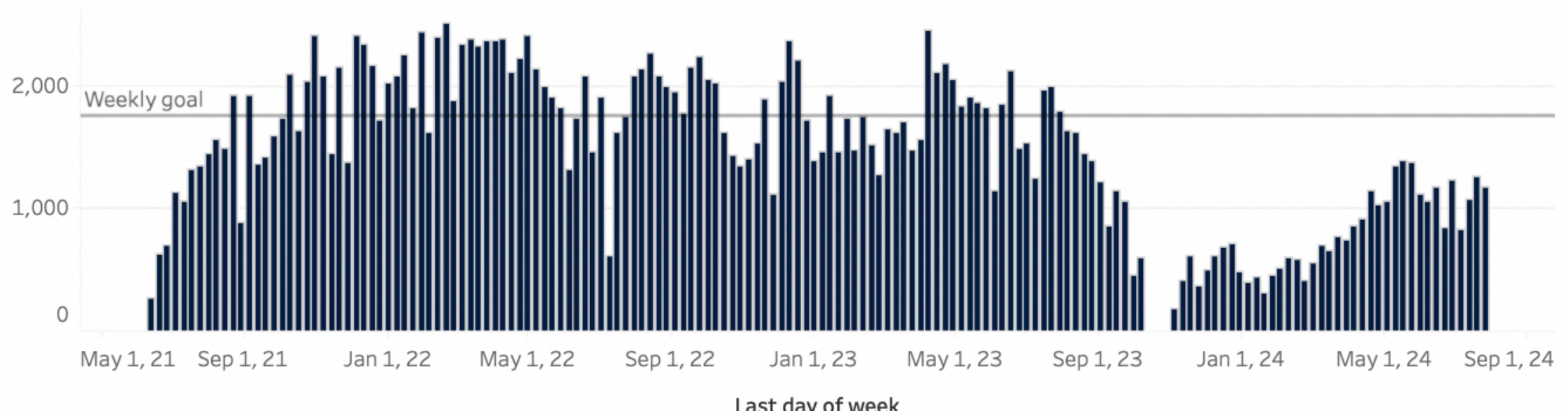
Percent Completion of Reel Preparation and Scanning



# Decennial Census Microfilm Scanning Progress by Week

These figures represent the number of microfilm reels scanned into digital images for a given week. The original weekly goal and current projected timeline for microfilm scanning are also displayed.

## Weekly Progress



# Capture

# Form Capture – Questionnaire

80, 15, and 5 percent (100 percent)

Page 2

**DO NOT MARK THIS COLUMN**

Line	1. WHAT IS THE NAME OF EACH PERSON who was living here on Wednesday, April 1, 1970 or who was staying or visiting here and had no other home?	2. HOW IS EACH PERSON RELATED TO THE HEAD OF THIS HOUSEHOLD?	3. SEX	4. COLOR OR RACE	DATE OF BIRTH			8. WHAT IS EACH PERSON'S MARITAL STATUS?	
					5. Month and year of birth and age last birthday	6. Month of birth	7. Year of birth		
1	<p>Print names in this order</p> <p>Head of the household Wife of head Unmarried children, oldest first Married children and their families Other relatives of the head Persons not related to the head</p> <p>Last name _____ First name _____ Middle initial _____</p>	<p>Fill one circle.</p> <p>1) "Other relative of head," also give exact relationship, for example, mother-in-law, brother, niece, grandson, etc. 2) "Other not related to head," also give exact relationship, for example, partner, maid, etc.</p> <p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Fill one circle</p> <p>Male Female</p>	<p>Fill one circle.</p> <p>If "Indian (American)," also give tribe. If "Other," also give race.</p> <p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Print</p> <p>Month _____ Year _____ Age _____</p>	<p>Fill one circle</p> <p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>Fill one circle for first three numbers</p> <p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>Fill one circle for last number</p> <p>0-9</p>	<p>Fill one circle</p> <p>Now married Widowed Divorced Separated Never married</p>
2	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
3	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
4	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
5	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
6	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
7	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
8	<p>Last name _____ First name _____ Middle initial _____</p>	<p>Head of household Wife of head Son or daughter of head Other relative of head—Print exact relationship</p> <p>Roomer, boarder, lodger Patient or inmate Other not related to head—Print exact relationship</p>	<p>Male Female</p>	<p>White Negro or Black Indian (Amer.) Japanese Chinese Filipino Hawaiian Korean Other—Print race</p>	<p>Month _____ Year _____ Age _____</p>	<p>Jan.-Mar. Apr.-June July-Sept. Oct.-Dec.</p>	<p>186-192 187-193 188-194 189-195 190-196 191-197</p>	<p>0-9</p>	<p>Now married Widowed Divorced Separated Never married</p>
<p>9. If you need all lines—Are there any other persons in this household?</p> <p>Yes <input type="checkbox"/> No <input type="checkbox"/></p> <p>Do not list the others; we will call to get the information.</p>		<p>10. Did you leave anyone out of Question 1 because you were not sure if he should be listed—for example, a new baby still in the hospital, or a lodger who also has another home?</p> <p>Yes <input type="checkbox"/> No <input type="checkbox"/></p> <p>On back page, give name(s) and reason left out.</p>		<p>11. Did you list anyone in Question 1 who is away from home now—for example, on a vacation or in a hospital?</p> <p>Yes <input type="checkbox"/> No <input type="checkbox"/></p> <p>On back page, give name(s) and reason person is away.</p>		<p>12. Did anyone stay here on Tuesday, March 31, who is not already listed?</p> <p>Yes <input type="checkbox"/> No <input type="checkbox"/></p> <p>On back page, give name of each visitor for whom there is no one at his home address to report him to a census taker.</p>			

26-1



# Form Capture – Newly Digitized Data

80, 15, and 5 percent (100 percent)

Page 2

**DO NOT MARK THIS COLUMN**

1. WHAT IS THE NAME OF EACH PERSON who was living here on Wednesday, April 1, 1970 or who was staying or visiting here and had no other home?	2. HOW IS EACH PERSON RELATED TO THE HEAD OF THIS HOUSEHOLD? <i>Fill one circle.</i> If "Other relative of head," give exact relationship, for example, mother-in-law, brother, niece, grandson, etc. If "Other not related to head," give exact relationship, for example, partner, maid, etc.	3. SEX  • <input type="checkbox"/> Male  • <input type="checkbox"/> Female <i>Fill one circle.</i>	4. COLOR OR RACE  <i>Fill one circle.</i> If "Indian (American)," give tribe. If "Other," give race.	DATE OF BIRTH			8. WHAT IS EACH PERSON'S MARITAL STATUS?  <i>Fill one circle.</i>
				5. Month and year of birth and age last birthday  <i>Fill one circle.</i>	6. Month of birth  <i>Fill one circle.</i>	7. Year of birth  <i>Fill one circle for first three numbers.</i> <i>Fill one circle for last number.</i>	
<p>Head of the household Wife of head Disjointed children, sister first Married children and their families Other relatives of the head Persons not related to the head</p> <p>Print names in this order</p> <p>1. Last name</p> <p>First name</p> <p>Middle initial</p>	<p><input type="radio"/> Head of household     <input type="radio"/> Roomer, boarder, lodger</p> <p><input type="radio"/> Wife of head     <input type="radio"/> Patient or inmate</p> <p><input type="radio"/> Son or daughter of head     <input type="radio"/> Other not related to head—<i>Print exact relationship</i></p> <p><input type="radio"/> Other relative of head—<i>Print exact relationship</i></p>	<p>Male <input type="radio"/></p> <p>Female <input type="radio"/></p>	<p><input type="radio"/> White     <input type="radio"/> Japanese     <input type="radio"/> Hawaiian</p> <p><input type="radio"/> Negro or Black     <input type="radio"/> Chinese     <input type="radio"/> Korean</p> <p><input type="radio"/> Indian (Amer.)     <input type="radio"/> Filipino     <input type="radio"/> Other—<i>Print race</i></p> <p><i>Print color</i></p>	<p>Month</p> <p>Year</p> <p>Age</p>	<p>Jan-Mar <input type="radio"/></p> <p>Apr-June <input type="radio"/></p> <p>July-Sept <input type="radio"/></p> <p>Oct-Dec <input type="radio"/></p>	<p>196-     <input type="radio"/> 197-     <input type="radio"/> 198-     <input type="radio"/> 199-     <input type="radio"/> 200-</p> <p>0     1     2     3     4     5</p>	<p><input type="radio"/> Now married</p> <p><input type="radio"/> Widowed</p> <p><input type="radio"/> Divorced</p> <p><input type="radio"/> Separated</p> <p><input type="radio"/> Never married</p>
<p>2. Last name</p> <p>First name</p> <p>Middle initial</p>	<p><input type="radio"/> Head of household     <input type="radio"/> Roomer, boarder, lodger</p> <p><input type="radio"/> Wife of head     <input type="radio"/> Patient or inmate</p> <p><input type="radio"/> Son or daughter of head     <input type="radio"/> Other not related to head—<i>Print exact relationship</i></p> <p><input type="radio"/> Other relative of head—<i>Print exact relationship</i></p>	<p>Male <input type="radio"/></p> <p>Female <input type="radio"/></p>	<p><input type="radio"/> White     <input type="radio"/> Japanese     <input type="radio"/> Hawaiian</p> <p><input type="radio"/> Negro or Black     <input type="radio"/> Chinese     <input type="radio"/> Korean</p> <p><input type="radio"/> Indian (Amer.)     <input type="radio"/> Filipino     <input type="radio"/> Other—<i>Print race</i></p> <p><i>Print color</i></p>	<p>Month</p> <p>Year</p> <p>Age</p>	<p>Jan-Mar <input type="radio"/></p> <p>Apr-June <input type="radio"/></p> <p>July-Sept <input type="radio"/></p> <p>Oct-Dec <input type="radio"/></p>	<p>196-     <input type="radio"/> 197-     <input type="radio"/> 198-     <input type="radio"/> 199-     <input type="radio"/> 200-</p> <p>0     1     2     3     4     5</p>	<p><input type="radio"/> Now married</p> <p><input type="radio"/> Widowed</p> <p><input type="radio"/> Divorced</p> <p><input type="radio"/> Separated</p> <p><input type="radio"/> Never married</p>

*Mark for punch in this margin*

# Capture: Keyed training and truth data


The screenshot shows a web browser window with the address bar displaying 'localhost:7000/EntryApp/'. The page title is 'Decennial Census Digitization and Data Linkage'. The main content area includes a description of the DCDL project, a status message for user 'jbid789', and a section for 'Recent images' with a table of sample record entry pages for the years 1960, 1970, and 1980.

## Decennial Census Digitization and Data Linkage

The Decennial Census Digitization and Linkage project (DCDL) is an initiative to produce linked restricted microdata files from the decennial censuses of 1960 through 1990. This tool enables hand-entry of data from a subset of scanned images of the historical files, an essential part of developing a scalable tool to capture respondent information so that individuals can be studied over time.

**User jbid789 has completed data entry for 0 / 2 images.**

The current reel is from 1990. Next image (2 remaining):



[Code Image](#)

### Recent images

image	year	type	complete?	updated	problem?	EDIT
-------	------	------	-----------	---------	----------	------

Click the buttons below to view sample record entry pages for each year.

[View 1960](#)

[View 1970](#)

[View 1980](#)



master ▾

1 Branch 0 Tags

Go to file

<> Code ▾

	<b>murray-cecile and murray-cecile</b> final production code	8b56fe4 · 2 years ago	🕒 1 Commit
📁	EntryApp	final production code	2 years ago
📁	dcdl	final production code	2 years ago
📁	load_test	final production code	2 years ago
📁	requirements	final production code	2 years ago
📁	templates/registration	final production code	2 years ago
📄	README.md	final production code	2 years ago

### About

Django app to enable hand entry of records from the 1960-1990 decennial censuses to construct training/test dataset for custom OCR/OMR algorithms

📖 Readme

📈 Activity

📄 Custom properties

☆ 1 star

👁️ 1 watching

🍴 1 fork

Report repository

# Optical Character Recognition (OCR) for handwritten names

- Team from Information Sciences Institute at University of Southern California for custom OCR software
  - Performed best on 1990 feasibility test in **2015**
  - Alexander, J. T., Fisher, J. D., & Genadek, K. R. (2022). Digitizing hand-written data with automated methods: A pilot project using the 1990 US Census. *Journal of Economic and Social Measurement*, 46(2), 95-108
- OCR supported by Census Bureau staff

# OCR Progress

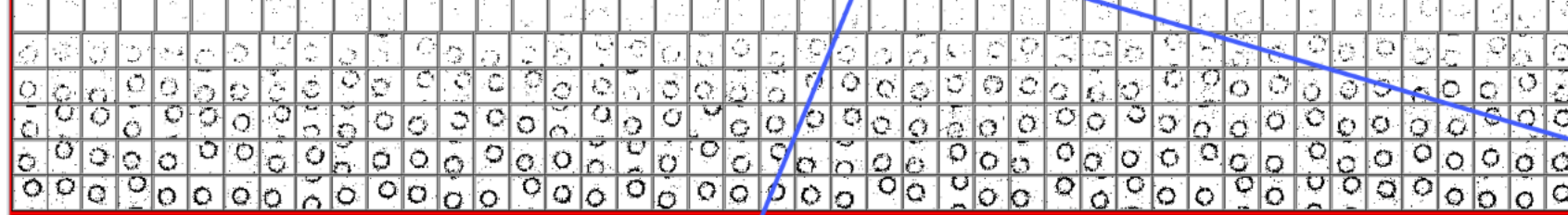
- Handwriting Capture
  - 1990 software developed: 87-95% exact name capture
  - 1980 software in development: 86-93% exact name match



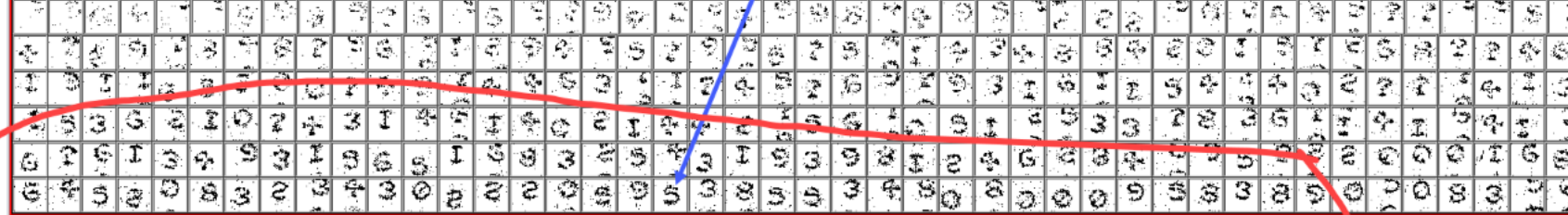
# Optical Mark Recognition (OMR)

- Census Bureau Information Technology Office developed and implementing custom OMR
- Team at NPC supporting OMR with Quality Assurance checks

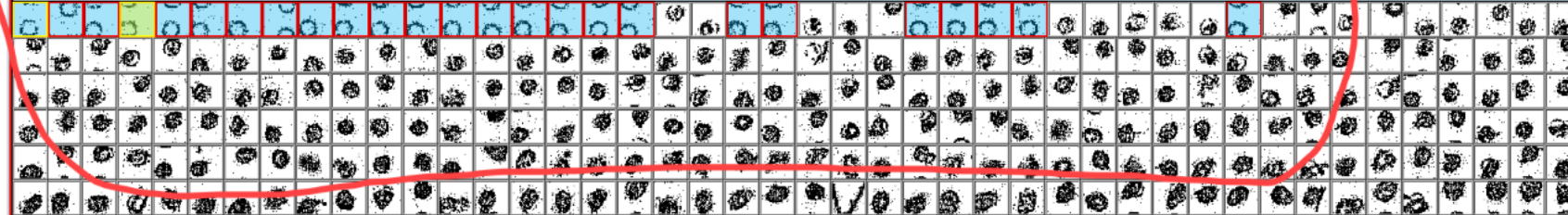
Type 1 unmarked: Click any items that are marked accuracy=386/386=100.0% dirt=found chargeable=0 non-chargeable=0



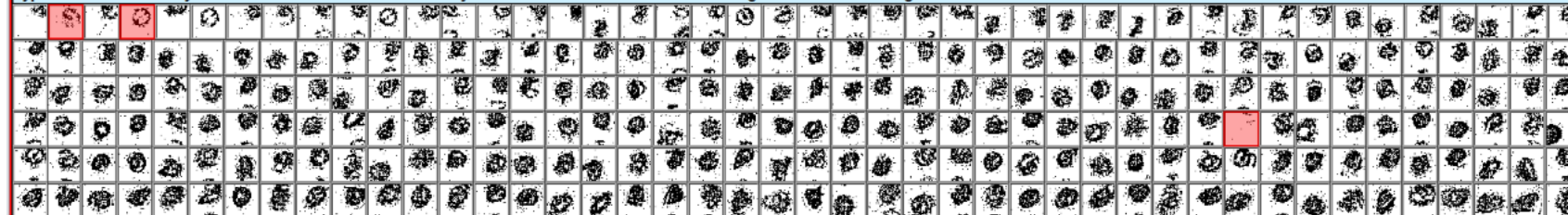
Type 2 unmarked: Click any items that are marked accuracy=386/386=100.0% dirt=found chargeable=0 non-chargeable=0



Type 1 Marked: Click any items that are unmarked accuracy=361/386=93.5% dirt=found chargeable=0 non-chargeable=25



Type 2 Marked: Click any items that are unmarked accuracy=384/386=99.5% dirt=found chargeable=2 non-chargeable=0



QC Page Viewer page 6b1990\_46011508\_0555.tif

FPV Menu

ITEMS ON PAGE 2

PERSON

RELATIVE OF Person 1

RELATIVE OF Person 2

NOT RELATED TO Person 1

NOT RELATED TO Person 2

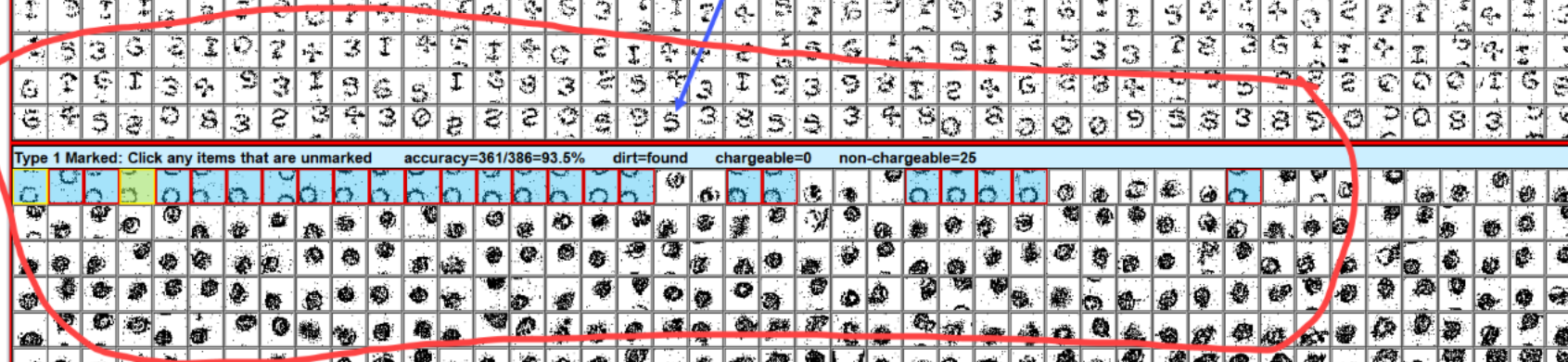
Other use (Specify)

FOR CENSUS USE

What is the monthly rent?

FOR CENSUS USE

Pixel And PT Counter Size 12 Original Size Fit Screen Center Zoom 0.19





Type 1 unmarked: Click any items that are marked accuracy=396/396=100.0% dirt=not found chargeable=0 non-chargeable=0

Type 2 unmarked: Click any items that are marked accuracy=396/396=100.0% dirt=not found chargeable=0 non-chargeable=0

Type 1 Marked: Click any items that are unmarked accuracy=396/396=100.0% dirt=not found chargeable=0 non-chargeable=0

Type 2 Marked: Click any items that are unmarked accuracy=395/395=100.0% dirt=not found chargeable=0 non-chargeable=0

# OMR Progress

- OMR – Bubble Capture
  - 1990 OMR software developed: 99% accuracy rates
  - In production, QA ongoing
  - 1980 software in production
  - 1970 software in development

7. Since February 1, 1980, has . . . attended regular school or college at any time?  
Fill one circle. Count Head Start, pre-kindergarten, kindergarten, elementary school, and schooling which leads to a high school diploma or college degree.

8. What is the highest grade (or year) of regular school . . . has ever attended?  
Fill one circle.  
If now attending school, mark grade person is in.  
If high school was graduated, mark grade (20-24).

9. How many of these persons in a family . . .  
a. have any babies born alive since April 1, 1979?  Yes  No  
b. How many of these children born alive since April 1, 1979 are still living?

10. Has . . . had any babies born alive since April 1, 1979?  Yes  No

11. Did . . . work at any time last week?  
Count part-time work, or looking without pay in a family business or farm. Also count active duty in the U.S. Armed Forces. Substantive activity includes fishing, handicrafts, etc. not for commercial purposes.

12. How many hours did . . . work last week?  
a. How many hours did . . . work last week? (at all jobs, including substantive activity) Subtract any time off, sick leave or extra hours worked.

13. How many hours did . . . work last week? (at all jobs, including substantive activity) Subtract any time off, sick leave or extra hours worked.

14. FOR CENSUS USE ONLY

14a.	15a.	16a.	17a.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

# 1970 CENSUS -- BREAKER SHEET

State <b>MICHIGAN</b>	County <b>94 MACOMB</b>	ED <b>000 0775</b>	Block <b>YES YES</b>	ACG <b>YES YES</b>	SNRA <b>2100</b>
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Form 70 - 51  
6 - 2  
2

- - -	=	0
-   - -	=	1
- -   -	=	2
- - -	=	3
-     - -	=	4
-   -   -	=	5
-   - -	=	6
- -     -	=	7
- -   -	=	8
- - -	=	9

# Additional Capture

- 1960, 1970 Enumeration District Books
- 1970 Census Addresses
  
- 1980 Address files
- 1990 Group Quarters

# Digitizing and parsing semi-structured historical administrative documents from the G.I. Bill mortgage guarantee program

Sara Lafia, David A. Bleckley, J. Trent Alexander ▼

Journal of Documentation

ISSN: 0022-0418

📅 Article publication date: 31 July 2023  
Issue publication date: 18 December 2023

DOWNLOADS



787

Aaron, Samuel H. and Pearl C.  
Wichita, Kansas

LHG-10124  
\$9,000.  
CC-64

Kansas City Agency

The Reserve Building & Loan Association  
Wichita, Kansas

(VA)  
FIRST MORTGAGE LOAN GUARANTEED

The image shows a document with several text blocks highlighted by colored boxes. Each box has a small icon and a label. A legend on the right side of the image lists these labels with corresponding numbers and colors. The legend items are: Name 1 (pink), Location 2 (red), Amount 3 (orange), ID 4 (yellow), Status 5 (light green), Agency 6 (green), and Other 7 (light pink). The document text includes: 'Aar... Samuel H. and Pearl C.', 'Wichita, Kansas', 'Kansas City Agency', 'The Reserve Building & Loan Association', 'Wichita, Kansas', '\$9,000.', 'CC-64', '(VA)', and 'FIRST MORTGAGE LOAN GUARANTEED'. The legend also includes 'Amount 0124' and 'ID' which are not explicitly labeled on the document but correspond to the highlighted areas.

Name 1

Location 2

Amount 3

ID 4

Status 5

Agency 6

Other 7



Character accuracy and flex character accuracy for each OCR method

---

<b>OCR method</b>	<b>Character accuracy</b>	<b>Flex character accuracy</b>
ABBYY FineReader	0.84	0.98
Adobe Acrobat	0.80	0.84
Tesseract in LayoutParser	0.83	0.87
OmniPage	0.90	0.95
Tesseract in pytesseract	0.95	0.99

# 1960 ED Listings

Original

County	Prefix	E. D.	Roll Number
Walker	64	1-16	119
	64	17-31	126
	64	32N-49	127
Washington	65	1N	60
	65	1P-14	37
Milcox	66	1-7P	59
	66	8-21P	105
Kinston	67	1N-5P	94
	67	6-15	5
Bessemer City	401	1A-11	47
	401	12N-26	128
Birmingham City	402	1-17	129
	402	18-32	130
	402	33-51	131
	402	52N-73	132
	402	74-90	133
	402	91-110	134
	402	111-127	135
	402	128-146	136
	402	147-164	137
	402	165-181	138
	402	182-194	139
	402	195-218	140
	402	219-235B	141
	402	236-251	142
	402	252-269C	143
	402	270N-290	144
	402	291A-305	145
	402	306-320	146
	402	321-339N	147
	402	339P-358	148
	402	359-366	67
Cullman City	403	1N-9P	41
Cadtsden City	405	1N-11S	149
	405	12-26	150
	405	27N-41	151
	405	42A-51P	30

Crop, align, clean

Pre-processed

64	1-16	119
64	17-31	126
64	32N-49	127
65	1N	60
65	1P-14	37
66	1-7P	59
66	8-21P	105
67	1N-5P	94
67	6-15	5
401	1A-11	47
401	12N-26	128
402	1-17	129
402	18-32	130
402	33-51	131
402	52N-73	132
402	74-90	133
402	91-110	134
402	111-127	135
402	128-146	136
402	147-164	137
402	165-181	138
402	182-194	139
402	195-218	140
402	219-235B	141
402	236-251	142
402	252-269C	143
402	270N-290	144
402	291A-305	145
402	306-320	146
402	321-339N	147
402	339P-358	148
402	359-366	67
403	1N-9P	41
405	1N-11S	149
405	12-26	150
405	27N-41	151
405	42A-51P	30

Tesseract OCR

OCR Output

64	1-16	119
64	17-31	126
64	32N-49	127
65	1N	60
65	1P-14	37
66	1-7P	59
66	8-21P	105
67	1N-5P	94
67	6-15	5
401	1A-11	47
401	12N-26	128
402	1-17	129
402	18-32	130
402	33-51	131
402	52N-73	132
402	74-90	133
402	91-110	134
402	111-127	135
402	128-146	136
402	147-164	137
402	165-181	138
402	182-194	139
402	195-218	140
402	219-235B	141
402	236-251	142
402	252-269C	143
402	270N-290	144
402	291A-305	145
402	306-320	146
402	321-339N	147
402	339P-358	148
402	359-366	67
403	1N-9P	41
405	1N-11S	149
405	12-26	150
405	27N-41	151
405	42A-51P	30

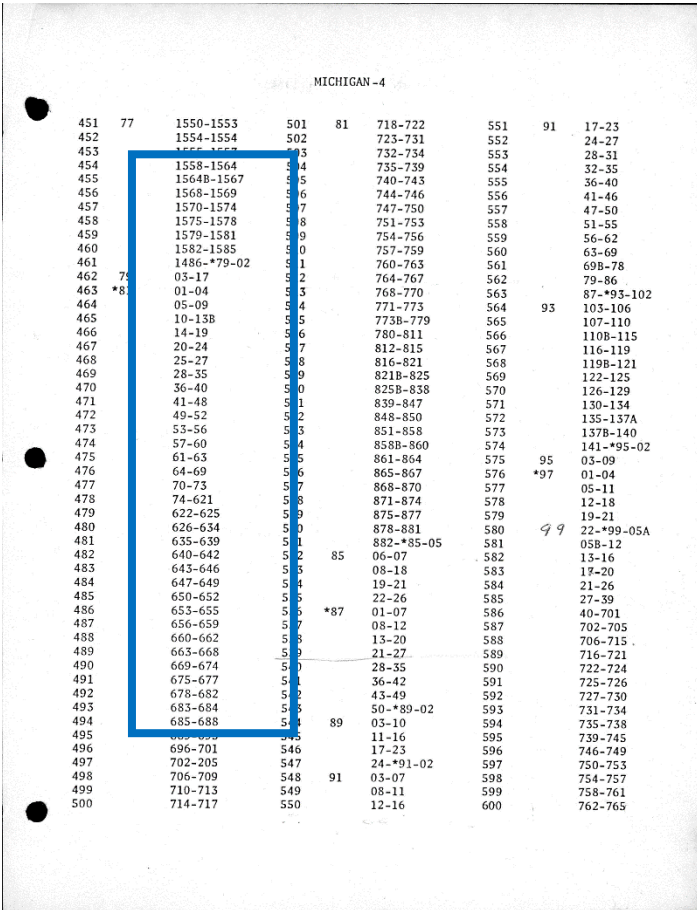
Edit, parse, csv

Project-ready data

state	prefix	min_ed	max_ed	reel
AL	64	1	16	119
AL	64	17	31	126
AL	64	32N	49	127
...	...	...	...	...
AL	402	147	164	137
AL	402	165	181	138
AL	402	182	194	139
AL	402	195	218	140

# 1970 ED Listings

Original



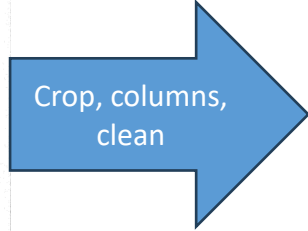
Pre-processed

```

451 77 1550-1553
452 1554-1554
453 1555-1557
454 1558-1564
455 1564B-1567
456 1568-1569
457 1570-1574
458 1575-1578
459 1579-1581
460 1582-1585
461 1486-*79-02
462 01-17
463 *8 01-04
464 05-09
465 10-13B
466 14-19
467 20-24
468 25-27
469 28-35
470 36-40
471 41-48
472 49-52
473 53-56
474 57-60
475 61-63
476 64-69
477 70-73
478 74-621
479 622-625
480 626-634
481 635-639
482 640-642
483 643-646
484 647-649
485 650-652
486 653-655
487 656-659
488 660-662
489 663-668
490 669-674
491 675-677
492 678-682
493 683-684
494 685-688
495 689-695
496 696-701
497 702-205
498 706-709
499 710-713
500 714-717

501 81 718-722
502 723-731
503 732-734
504 735-739
505 740-743
506 744-746
507 747-750
508 751-753
509 754-756
510 757-759
511 760-765
512 764-767
513 768-770
514 771-773
515 773B-779
516 780-811
517 812-815
518 816-821
519 821B-825
520 825B-838
521 839-847
522 848-850
523 851-858
524 858B-860
525 861-864
526 865-867
527 868-870
528 871-874
529 875-877
530 878-881
531 882-*85-05
532 06-07
533 08-18
534 19-21
535 22-26
536 01-07
537 08-12
538 13-20
539 21-27
540 28-35
541 36-42
542 43-49
543 50-*89-02
544 03-10
545 11-16
546 17-23
547 24-*91-02
548 03-07
549 08-11
550 12-16

551 91 17-23
552 24-27
553 28-31
554 32-35
555 36-40
556 41-46
557 47-50
558 51-55
559 56-62
560 63-69
561 69B-78
562 79-86
563 87-*93-102
564 103-106
565 107-110
566 110B-115
567 116-119
568 119B-121
569 122-125
570 126-129
571 130-134
572 135-137A
573 137B-140
574 141-*95-02
575 03-09
576 01-04
577 05-11
578 12-18
579 19-21
580 22-*99-05A
581 05B-12
582 13-16
583 17-20
584 21-26
585 27-39
586 40-701
587 702-705
588 706-715
589 716-721
590 722-724
591 725-726
592 727-730
593 731-734
594 735-738
595 739-745
596 746-749
597 750-753
598 754-757
599 758-761
600 762-765
    
```



OCR Output

```

451 77 1550-1553
452 1554-1554
453 1555-1557
454 1558-1564
455 1564B-1567
456 1568-1569
457 1570-1574
458 1575-1578
459 1579-1581
460 1582-1585
461 1486-*79-02
462 03-17
463 *81 01-04
464 05-09
465 10-13B
466 14-19
467 20-24
468 25-27
469 28-35
470 36-40
471 41-48
472 49-52
473 53-56
474 57-60
475 61-63
476 64-69
477 70-73
478 74-621
479 622-625
480 626-634
481 635-639
482 640-642
483 643-646
484 647-649
485 650-652
486 653-655
487 656-659
488 660-662
489 663-668
490 669-674
491 675-677
492 678-682
493 683-684
494 685-688
495 689-695
496 696-701
497 702-205
498 706-709
499 710-713
500 714-717
    
```



Project-ready data

state	reel	prefix	ed_min	ed_max
MI	451	77	1550	1553
MI	452	77	1554	1554
MI	453	77	1555	1557
...	...	...	...	...
MI	464	81	5	9
MI	465	81	10	13B
MI	466	81	14	19

# Preprocessing: Clean Images

Methods: Binarization → noise reduction → de-skewing → line removal → binder hole removal → increasing DPI

ZIP	STREET NAME	HOUSE NO. RANGE		SERIAL RANGE		L.D.	TRACT	BLOCK
COUL		FIRST	LAST	FIRST	LAST	NO.	NUMBER	NUMBER
22030	BRADDOCK RD	6591	6599			0968	4035	403
22030	BRADDOCK RD	6601	6609			0968	4035	404
22030	BRADDOCK RD	6611	6619			0968	4035	405
22030	BRADDOCK RD	6901	6999			0968	4035	412
22030	BRADDOCK RD	7233	7299			0971	4034	103
22030	BRADDOCK RD	7301	7499			0972	4034	408
22030	BRADDOCK RD	7501	7699			0972	4034	409
22030	BRADDOCK RD	7700	7998			0827	4063	216
22030	BRADDOCK RD	7701	7999			0972	4034	501
22030	BRADDOCK RD	7706	7698			0829	4064	209
22030	BRADDOCK RD	8000	8398			0834	4061	904
22030	BRADDOCK RD	8001	8099			0836	4065	101
22030	BRADDOCK RD	8101	8199			0836	4065	101
22030	BRADDOCK RD	8201	8299			0836	4065	102
22030	BRADDOCK RD	8301	8599			0836	4065	901
22030	BRADDOCK RD	8400	8498			0834	4061	905
22030	BRADDOCK RD	8500	8598			0834	4061	903
22030	BRADDOCK RD	8600	8698			0834	4061	903
22030	BRADDOCK RD	8601	8799			0836	4065	202
22030	BRADDOCK RD	8700	8814			0834	4061	903
22030	BRADDOCK RD	8801	8899			0836	4065	203
22030	BRADDOCK RD	8816	9098			0834	4061	903
22030	BRADDOCK RD	8901	8999			0836	4065	204
22030	BRADDOCK RD	9001	9199			0967	4032	999
22030	BRADDOCK RD	9100	9198			0835	4061	909
22030	BRADDOCK RD	9200	9298			0834	4061	903



ZIP	STREET NAME	HOUSE NO. RANGE		SERIAL RANGE		L.D.	TRACT	BLOCK
COUL		FIRST	LAST	FIRST	LAST	NO.	NUMBER	NUMBER
22030	BRADDOCK RD	6591	6599			0968	4035	403
22030	BRADDOCK RD	6601	6609			0968	4035	404
22030	BRADDOCK RD	6611	6619			0968	4035	405
22030	BRADDOCK RD	6901	6999			0968	4035	412
22030	BRADDOCK RD	7233	7299			0971	4034	103
22030	BRADDOCK RD	7301	7499			0972	4034	408
22030	BRADDOCK RD	7501	7699			0972	4034	409
22030	BRADDOCK RD	7700	7998			0827	4063	216
22030	BRADDOCK RD	7701	7999			0972	4034	501
22030	BRADDOCK RD	7706	7698			0829	4064	209
22030	BRADDOCK RD	8000	8398			0834	4061	904
22030	BRADDOCK RD	8001	8099			0836	4065	101
22030	BRADDOCK RD	8101	8199			0836	4065	101
22030	BRADDOCK RD	8201	8299			0836	4065	102
22030	BRADDOCK RD	8301	8599			0836	4065	901
22030	BRADDOCK RD	8400	8498			0834	4061	905
22030	BRADDOCK RD	8500	8598			0834	4061	903
22030	BRADDOCK RD	8600	8698			0834	4061	903
22030	BRADDOCK RD	8601	8799			0836	4065	202
22030	BRADDOCK RD	8700	8814			0834	4061	903
22030	BRADDOCK RD	8801	8899			0836	4065	203
22030	BRADDOCK RD	8816	9098			0834	4061	903
22030	BRADDOCK RD	8901	8999			0836	4065	204
22030	BRADDOCK RD	9001	9199			0967	4032	999
22030	BRADDOCK RD	9100	9198			0835	4061	909
22030	BRADDOCK RD	9200	9298			0834	4061	903



Original Image

Preprocessed Image

# Document Image Analysis: Detect image layout

**Methods:** Annotating 50 scans with Label Studio, splitting up train-test data, training layout detection model.

ZIP CODE	STREET NAME	HOUSE NO. RANGE		SERIAL RANGE		E.D. NO.	TRACT NUMBER	BLOCK NUMBER
		FIRST	LAST	FIRST	LAST			
22041	ARNETT ST	5400	5422			0897	4053	104
22041	ARNETT ST	5401	5423			0897	4053	104
22041	ARNETT ST	5424	5498			0898	4053	104
22041	ARNETT ST	5425	5499			0898	4053	106
22041	ASHWOOD PLACE	3900	3996			0906	4052	999
22041	ASHWOOD PLACE	3901	3999			0906	4052	999
22041	AURA COURT	4100	4198			0902	4051	102
22041	AURA COURT	4101	4199			0902	4051	102
22041	BANCROFTVY TER	3600	3698			0899	4053	999
22041	BANCROFTVY TER	3601	3699			0899	4053	999
22041	BAY TREE CT	4600	4698			0905	4052	205
22041	BAY TREE CT	4601	4699			0905	4052	205
22041	BAY TREE LANE	4600	4698			0905	4052	205
22041	BAY TREE LANE	4601	4699			0905	4052	210
22041	BEACHWAY DR	4100	4124			0904	4051	405
22041	BEACHWAY DR	4101	4125			0904	4051	403
22041	BEACHWAY DR	4127	4141			0904	4051	404
22041	BEACHWAY DR	4128	4198			0902	4051	106
22041	BLANCHWAY DR	4193	4207			0904	4051	405
22041	BLANCHWAY DR	4200	4210			0902	4051	105
22041	BLANCHWAY DR	4209	4213			0904	4051	405
22041	BLANCHWAY DR	4212	4298			0887	4047	24
22041	BLANCHWAY DR	4215	4299			0904	4051	407
22041	BELLEVIEW DR	4000	4050			0903	4051	207
22041	BELLEVIEW DR	4001	4099			0903	4051	207
22041	BELLEVIEW DR	4051	4099			0903	4051	206

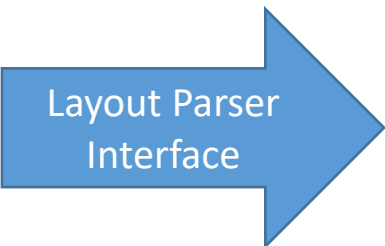


STREET NAME	HOUSE NO. RANGE		SERIAL RANGE		E.D. NO.	TRACT NUMBER	BLOCK NUMBER
	FIRST	LAST	FIRST	LAST			
1746 W ALBANY ST	200	218			5735	11200	311
1746 W ALBANY ST	201	219			5735	11200	316
1746 ALBERMARLE AVE	2	18			5720	1114	105
1746 ALBERMARLE AVE	2	18			5720	1114	104
1746 ALBERMARLE AVE	2	20			5721	1114	301
1746 ALBERMARLE AVE	3	7			5720	1114	105
1746 ALBERMARLE AVE	9	27			5720	1114	106
1746 ALBERMARLE AVE	11	33			5721	1114	210
1746 ALBERMARLE AVE	15	215			5720	1114	101
1746 ALBERMARLE AVE	19	71			5720	1114	101
1746 ALBERMARLE AVE	20	48			5720	1114	106
1746 ALBERMARLE AVE	26	40			5721	1114	303
1746 ALBERMARLE AVE	46	46			5721	1114	208
1746 ALBERMARLE AVE	48	48			5721	1114	302
1746 ALBERMARLE AVE	49	55			5721	1114	209
1746 ALBERMARLE AVE	50	70			5720	1114	107
1746 ALBERMARLE AVE	60	64			5721	1114	302
1746 ALBERMARLE AVE	66	76			5721	1114	303
1746 ALBERMARLE AVE	71	79			5721	1114	304
1746 ALBERMARLE AVE	81	93			5721	1114	304
1746 ALBERMARLE AVE	161	177			5720	1114	107
1746 ALBERMARLE AVE	162	214			5720	1114	101
1746 ALBERMARLE AVE	254	258			5721	1114	208
1746 ALDERFIELD LANE	1	5			5727	11220	214
1746 ALDERFIELD LANE	2	22			5727	11220	215
1746 ALDERFIELD LANE	2	37			5727	11220	215

# Text Extraction: Tesseract OCR Engine

**Methods:** For each column, predict most likely text and transform raw output into readable data frame.

NAME	HOUSE NO. RANGE		SERIAL RANGE		E.D. NO.	TRACT NUMBER	BLOCK NUMBER
	FIRST	LAST	FIRST	LAST			
1746 W ALBANY ST	200	218			5735	11200	311
1746 W ALBANY ST	201	219			5735	11200	316
1746 ALBERMARLE AVE	2	18			5720	1114	105
1746 ALBERMARLE AVE	2	18			5720	1114	104
1746 ALBERMARLE AVE	2	20			5721	1114	301
1746 ALBERMARLE AVE	3	7			5720	1114	105
1746 ALBERMARLE AVE	9	27			5720	1114	106
1746 ALBERMARLE AVE	11	55			5721	1114	210
1746 ALBERMARLE AVE	15	215			5720	1114	101
1746 ALBERMARLE AVE	19	71			5720	1114	101
1746 ALBERMARLE AVE	20	48			5720	1114	106
1746 ALBERMARLE AVE	26	40			5721	1114	303
1746 ALBERMARLE AVE	46	46			5721	1114	208
1746 ALBERMARLE AVE	48	48			5721	1114	302
1746 ALBERMARLE AVE	49	55			5721	1114	209
1746 ALBERMARLE AVE	50	70			5720	1114	107
1746 ALBERMARLE AVE	60	64			5721	1114	302
1746 ALBERMARLE AVE	66	76			5721	1114	303
1746 ALBERMARLE AVE	71	79			5721	1114	304
1746 ALBERMARLE AVE	81	93			5721	1114	304
1746 ALBERMARLE AVE	161	177			5720	1114	107
1746 ALBERMARLE AVE	162	214			5720	1114	101
1746 ALBERMARLE AVE	234	238			5721	1114	208
1746 ALDERFIELD LANE	1	5			5727	11220	214
1746 ALDERFIELD LANE	2	22			5727	11220	215
1746 ALDERFIELD LANE	2	57			5727	11220	215



	zip	street_name	hn_first	hn_last	serial_first	serial_last	ed	tract	block
0	1746	ILE	200	218			5735	11200	311
1	1744	ALBANY ST	201	219			5735	112001	316
2	1746	ALBANY ST	11				53720	1114	109
3	1746	ALBERMARLE ZVE	13	NaN			5720	1114	104
4	1746	ALBERMARLE AVE	19	NaN			57231	1114	301
5	1746	_ALBERMARLE AVE	20	NaN			5720	1114	108
6	1746	ALBERMARLE AVE	26	NaN			5720	1114	106
7	1746	ALBERMARLE AVE	46	NaN			5721	1114	210
8	1746	ALBERMARLE AVE	48	NaN			5720	1114	10
9	1746	ALBERMARLE AVE	49	NaN			5720	1114	101
10	1746	ALBERMARLE AVE	50	NaN			53720	1114	106
11	1746	_ALBERMARLE AVE	60	NaN			5721	1114	303
12	1766	ALBERMARLE AVE	66	NaN			5721	1114	208
13	1746	ALBERMARLE AVE	71	NaN			5721	1114	302
14	1766	ALBERMARLE AVE	81	NaN			5721	1114	209
15	1744	ALBERMARLE AVE	161	NaN			8720	1114	7
16	1746	ALBERMARLE AVE	162	NaN			5721	1114	302
17	17646	ALBERMARLE AVE	34	NaN			5721	1114	303
18	1746	ALBERMARLE AVE		NaN			3721	1114	304
19	1746	ALBERMARLE AVE	NaN	NaN			8721	1114	107
20	746	ALBERMARLE AVE	NaN	NaN			5720	1114	101
21	1746	ALBERMARLE AVE	NaN	NaN			3720	1114	208
22	1746	ALBERMARLE AVE	NaN	NaN			5721	1114	214
23	1746	ALBERMARLE AVE	NaN	NaN			5727	112201	2153
24	1746	ALDERFIELD LANE	NaN	NaN			5727	112207	218
25	7406	ALDERFIELD LANE	NaN	NaN			5727	112201	



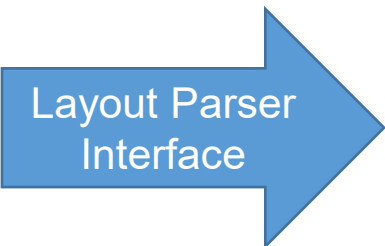
Univer  
Ce  
Detectron2 Model Detection

OCR Output

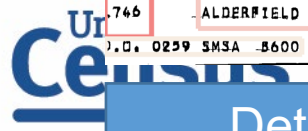
# 1970 Block Header Records

**Methods:** Document Image Analysis, OCR and transform raw output into readable data frame.

NAME	HOUSE NO. RANGE		SERIAL RANGE		E.D. ND.	TRACT NUMBER	BLOCK NUMBER
	FIRST	LAST	FIRST	LAST			
1746 W ALBANY ST	200	218			5735	11200	311
1746 W ALBANY ST	201	219			5735	11200	316
1746 ALBERMARLE AVE	2	18			5720	1114	105
1746 ALBERMARLE AVE	2	18			5720	1114	104
1746 ALBERMARLE AVE	2	20			5721	1114	301
1746 ALBERMARLE AVE	3	7			5720	1114	105
1746 ALBERMARLE AVE	9	27			5720	1114	106
1746 ALBERMARLE AVE	11	55			5721	1114	210
1746 ALBERMARLE AVE	15	215			5720	1114	101
1746 ALBERMARLE AVE	19	71			5720	1114	101
1746 ALBERMARLE AVE	20	48			5720	1114	106
1746 ALBERMARLE AVE	26	40			5721	1114	303
1746 ALBERMARLE AVE	46	46			5721	1114	208
1746 ALBERMARLE AVE	48	48			5721	1114	302
1746 ALBERMARLE AVE	49	55			5721	1114	209
1746 ALBERMARLE AVE	50	70			5720	1114	107
1746 ALBERMARLE AVE	60	64			5721	1114	302
1746 ALBERMARLE AVE	66	76			5721	1114	303
1746 ALBERMARLE AVE	71	79			5721	1114	304
1746 ALBERMARLE AVE	81	93			5721	1114	304
1746 ALBERMARLE AVE	161	177			5720	1114	107
1746 ALBERMARLE AVE	162	214			5720	1114	101
1746 ALBERMARLE AVE	254	258			5721	1114	208
1746 ALDERFIELD LANE	1	5			5727	11220	214
1746 ALDERFIELD LANE	2	22			5727	11220	215
1746 ALDERFIELD LANE	2	57			5727	11220	215



	zip	street_name	hn_first	hn_last	serial_first	serial_last	ed	tract	block
0	1746	ILE	200	218			5735	11200	311
1	1744	ALBANY ST	201	219			5735	112001	316
2	1746	ALBANY ST	11				53720	1114	109
3	1746	ALBERMARLE ZVE	13	NaN			5720	1114	104
4	1746	ALBERMARLE AVE	19	NaN			57231	1114	301
5	1746	_ALBERMARLE AVE	20	NaN			5720	1114	108
6	1746	ALBERMARLE AVE	26	NaN			5720	1114	106
7	1746	ALBERMARLE AVE	46	NaN			5721	1114	210
8	1746	ALBERMARLE AVE	48	NaN			5720	1114	10
9	1746	ALBERMARLE AVE	49	NaN			5720	1114	101
10	1746	ALBERMARLE AVE	50	NaN			53720	1114	106
11	1746	_ALBERMARLE AVE	60	NaN			5721	1114	303
12	1766	ALBERMARLE AVE	66	NaN			5721	1114	208
13	1746	ALBERMARLE AVE	71	NaN			5721	1114	302
14	1766	ALBERMARLE AVE	81	NaN			5721	1114	209
15	1744	ALBERMARLE AVE	161	NaN			8720	1114	7
16	1746	ALBERMARLE AVE	162	NaN			5721	1114	302
17	17646	ALBERMARLE AVE	34	NaN			5721	1114	303
18	1746	ALBERMARLE AVE		NaN			3721	1114	304
19	1746	ALBERMARLE AVE	NaN	NaN			8721	1114	107
20	746	ALBERMARLE AVE	NaN	NaN			5720	1114	101
21	1746	ALBERMARLE AVE	NaN	NaN			3720	1114	208
22	1746	ALBERMARLE AVE	NaN	NaN			5721	1114	214
23	1746	ALBERMARLE AVE	NaN	NaN			5727	112201	2153
24	1746	ALDERFIELD LANE	NaN	NaN			5727	112207	218
25	7406	ALDERFIELD LANE	NaN	NaN			5727	112201	



Detectron2 Layout Detection

Tesseract OCR Output

# Linkage



# Census Bureau Record Linkage

All cases linked to a common composite of administrative records

- Linkage with name, date of birth, sex, birthplace, address
- Standard techniques of data cleaning and probabilistic matching
- Matched cases are assigned a “Protected Identification Key” (PIK)
- Names removed before data are made available for research

# Linkage rates for decennial censuses

	Records with PIK assigned	% of all records with PIK
2010 Census	283,000,000	91%
2000 Census	249,000,000	89%
1940 Census	54,000,000	41%

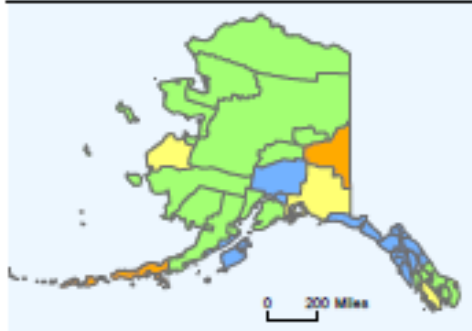
# Administrative data coverage: availability and gaps

Easiest to assign PIKs:

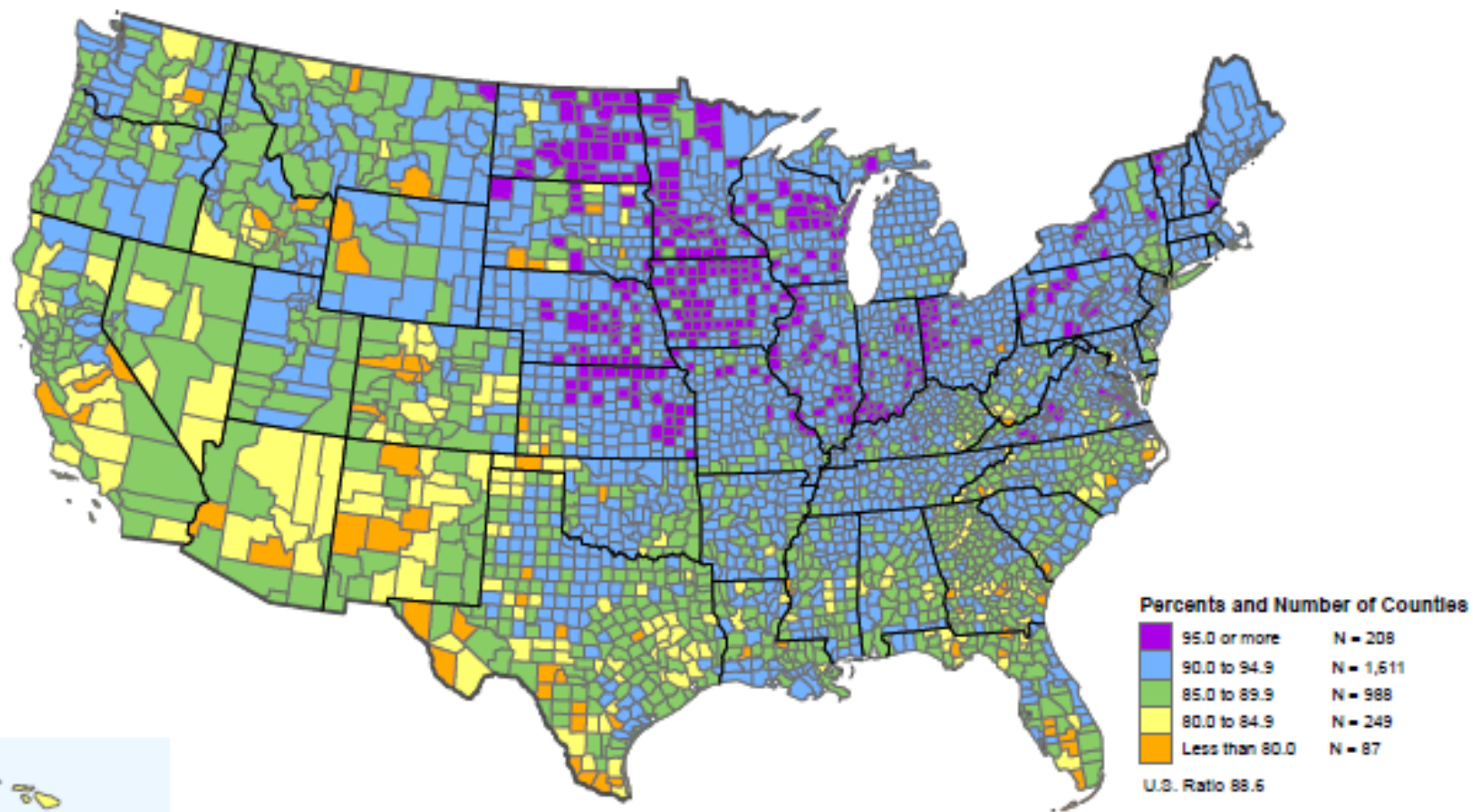
- Taxpayers
- Workers
- People on Medicare (most of the 65+ population)
- People receiving housing/medical/food assistance
- Men who ever registered for Selective Service

Harder to assign PIKs:

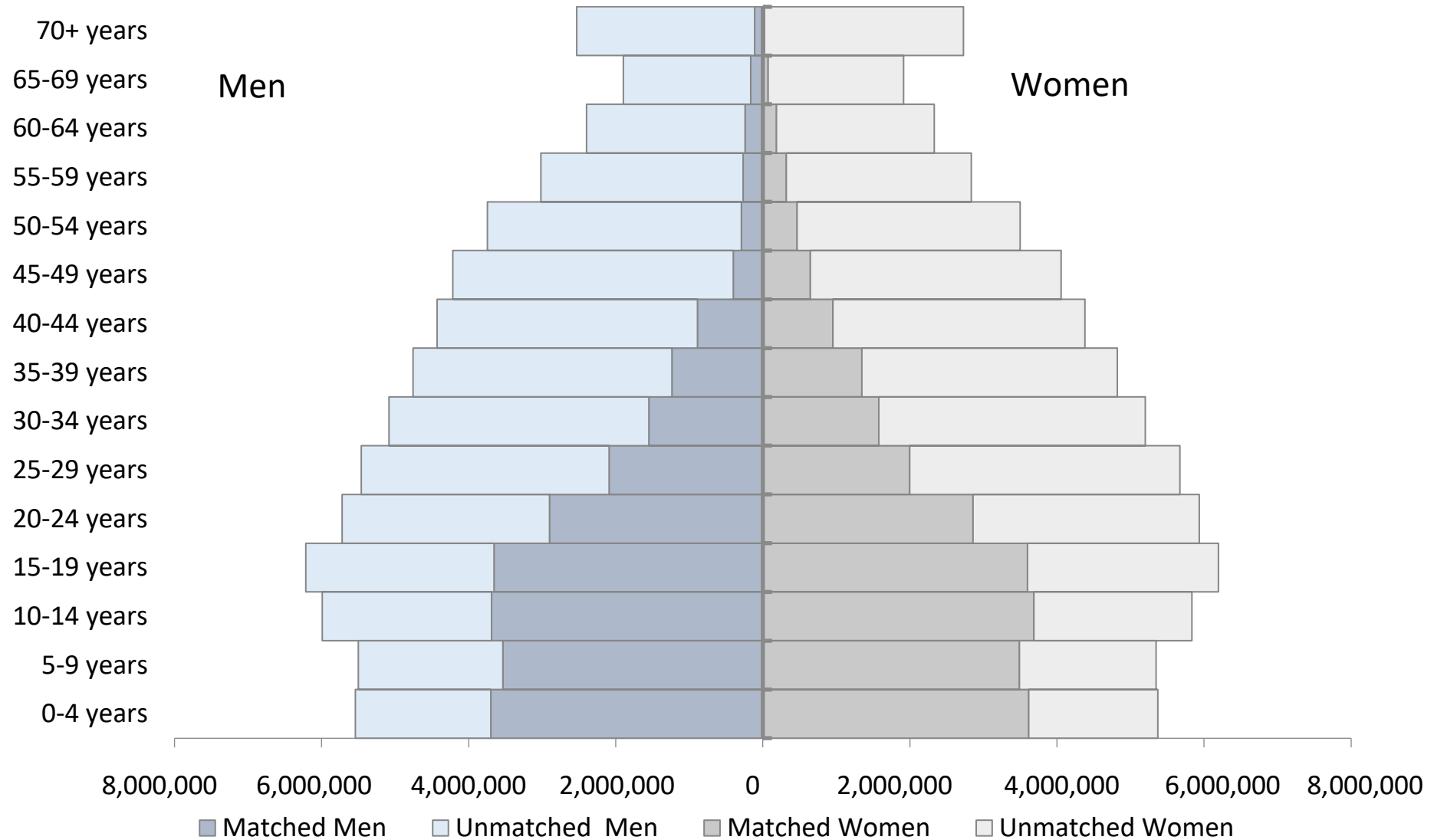
- Mobile and transient populations
- Residents without SSN
- Children not claimed on tax forms



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



# PIKs in the 1940 Census by Age and Sex



# DCDL Linkage

Techniques being developed with available samples

Using standard Census Bureau linkage techniques

- Making historical composite with 1969-1989 tax files
- Using recovered street addresses from 1970/80/90 censuses

# Preliminary Results

National Survey of College Graduates – 150,000 cases

1990 Census names entered for a sample of college grads

PIK rate: 93%

1990 Census Content Reinterview Survey – 28,000 cases

Names entered in follow-up interviews

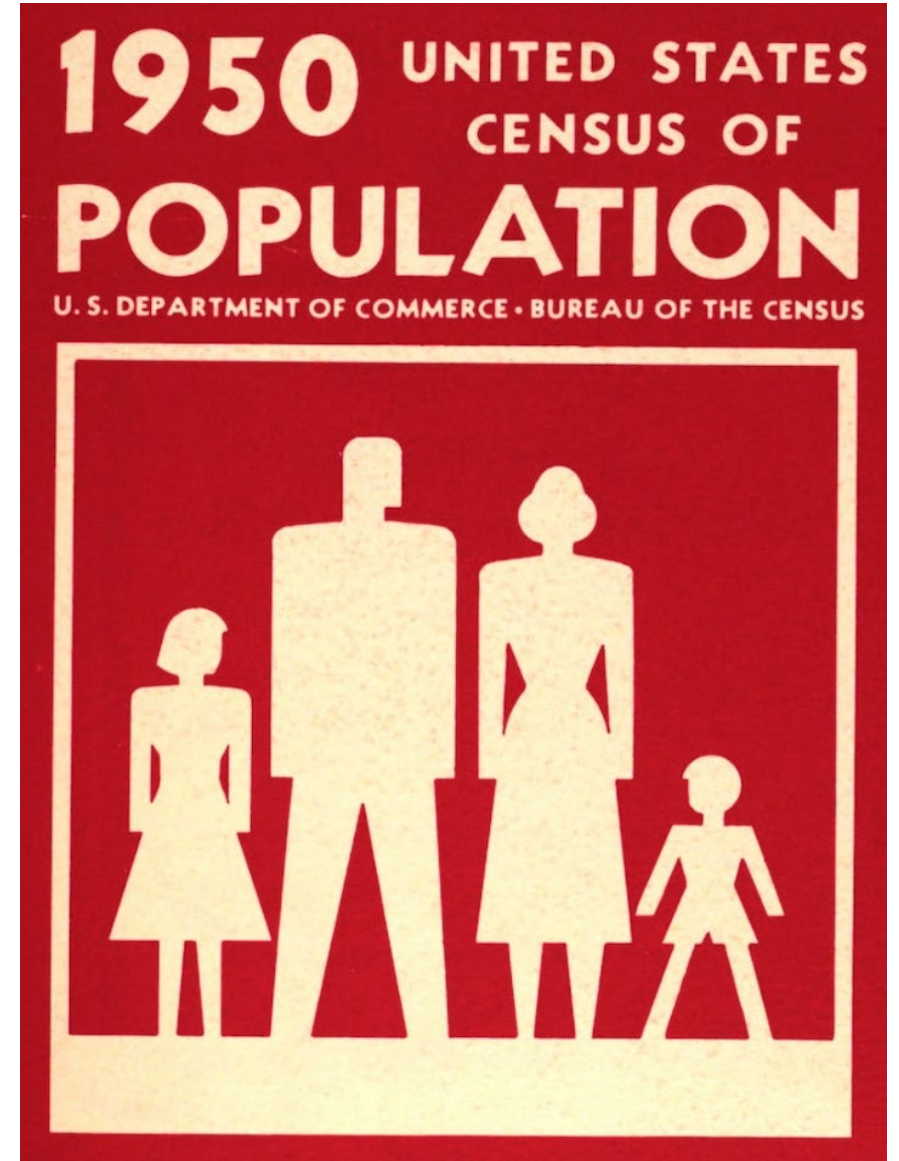
PIK rate: 82%

# Upcoming work



# 1950 Census

- University of Minnesota partnered with Ancestry.com to obtain full count microdata
- Census Bureau agreement with IPUMS for use of restricted full count data
- We will link 1950 into Census Bureau infrastructure
- Funding through NIH-NIA R21
- Expect linkages in 2025



Notes

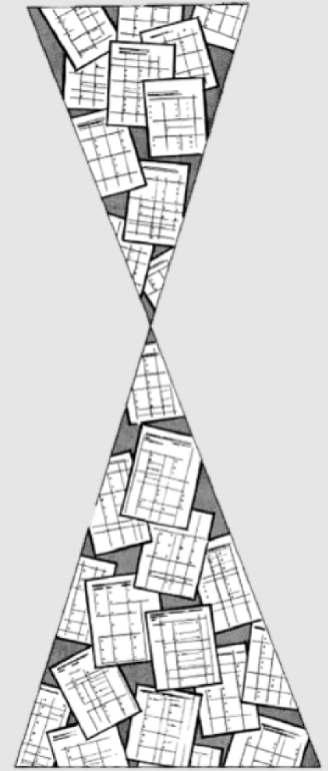
LINE NUMBER	FOR HEAD OF HOUSEHOLD						FOR ALL PERSONS										FOR PERSONS 14 YEARS OF AGE AND OVER									
	House (and apartment) number	Is this house on a place of three or more acres?	Is this house on a farm (or ranch)?	Agriculture Quantitative Number	NAME	RELATIONSHIP	RACE	SEX	How old was he on his last birthday?	Is he now married, widowed, divorced, separated, or never married?	What State (or foreign country) was he born in?	Foreign born?	What was this person doing last week—working, keeping house, or something else?	1. If employed (Wk in item 16, or Yes in item 16 or item 18), describe job or business held last week.	2. If looking for work (Yes in item 19), describe last job or business.	3. For all other persons, leave blank.	What kind of work was he doing?	What kind of business or industry was he working in?	Class of worker							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20a	20b	20c	20d				

4004	153	no	no		Hilwood, Lillie M.	Head	W	F	61	Mar	Minn		21	no	no	no					
					Hazel C.	Daughter	W	F	33	Mar	Minn		21	no	no	no	40	Book keeper	Wholesale Hardware	P	210 6161
					Childs, Dale	Lodger	W	M	21	Mar	Utah	037	21	no	no	no	70	Missionary Work	Santa Church	P	278 7961
					Caruth, David J.	Lodger	W	M	21	Mar	Utah	057	21	no	no	no	70	Missionary Work	Santa Church	P	078 8261
4018	154	no	no		O'Reilly, Eugene J.	Head	W	M	64	Mar	Minn		21	no	no	no	60	Labor Agency	Union Business	C	490 6593
					Grace E.	Wife	W	F	53	Mar	Wisconsin	035	21	no	no	no					
					Mary C.	Daughter	W	F	33	Mar	North Dakota	044	21	no	no	no					
					William O.	Son	W	M	25	Mar	Minn		21	no	no	no					
4022	155	no	no		O'Reilly, Eugene J.	Head	W	M	30	Mar	Minn		21	no	no	no	40	Manufactures Representative	Labor Agency	P	490 6591
					Marjorie L.	Wife	W	F	28	Mar	Minn		21	no	no	no					
					Katherine L.	Daughter	W	F	1	Mar	Minn		21	no	no	no					
					Margaret	Daughter	W	F	10	Mar	Minn		21	no	no	no					
4026	156	no	no		Cull, Catherine M.	Head	W	F	42	Mar	Minn		21	no	no	no	40	Telephone Operator	Telephone Co	P	370 5781
					Cordell, Arthur	Lodger	W	M	46	Mar	Minn		21	no	no	no	40	Stockroom Clerk	Retail Store	P	390 76991
					Dorance	Lodger	W	F	46	Mar	Minn		21	no	no	no	40	Power Machine Operator	Overall Grocery	P	690 4481
4034	157	no	no		Genadek, Thomas A.	Head	W	M	26	Mar	Minn		21	no	no	no	40	Switchman	Telephone Co	P	681 5781
					Janet C.	Wife	W	F	26	Mar	Minn		21	no	no	no					
					Theresa A.	Daughter	W	F	4	Mar	Minn		21	no	no	no					
					Paul L.	Son	W	M	2	Mar	Minn		21	no	no	no					
					John A.	Son	W	M	2 1/2	Mar	Minn		21	no	no	no					
4034	158	no	no		Genadek, Rose A.	Head	W	F	57	Mar	Minn		21	no	no	no	40	Power Machine Operator	Spice Factory	P	490 4481
					Joseph A.	Son	W	M	21	Mar	Minn		21	no	no	no	40	Accounting Clerk	Brain Co	P	310 6191
4038	159	no	no		Valley, Mabel M.	Head	W	F	47	Mar	Iowa	042	21	no	no	yes	40	Power Machine Operator	Lingerie Factory	P	490 4481
					Helen A.	Daughter	W	F	16	Mar	Minn		21	no	no	no					
					Harold J.	Son	W	M	14	Mar	Minn		21	no	no	no					

# 1960 Census

- 1960 Census microdata is only from long form
- 1960 Census microfilm at the Census Bureau from short form – 100% file
- We will create microdata file for the full 1960 Census
- Funding from NIH-NIA R01
- Work will take place 2024-2029

## UNITED STATES CENSUSES OF POPULATION AND HOUSING 1960



**So many  
people have  
worked on  
this project  
or are  
working on it  
at this  
moment.**

**It could not  
be done  
without  
them.**

Aaron Glover  
Abigail Hiller  
Adam Galemore  
Alessandro Ferrucci  
Allison D Fox  
Amanda Pait  
Ana Cherry  
Andrew Hennesy  
Andrew Larson  
Angeline K Brock  
Anna Simpson  
Aretha Wilkerson  
Arrington Smith  
Babatunde Adewuyi  
Barbara J Maupin  
Becky J Francis  
Benjamin Peterson  
Benny Alvey  
Betty J Jones  
Bijou Kisula  
Bill Peil  
Bittany Dixon  
Brenda Figueroa  
Cam Dang  
Carla Medalia  
Carlos Bacerra  
Carlos Mena  
Carrie A Ridge  
Catherine L Beauchamp  
Cathy Sharp  
Cecile Murray  
Chad Adkins  
Charles Andrew Stahl  
Cheryl Grim  
Cheryl Smeltekop  
Christian Moscardi  
Christine Mosely  
Christine Parkhill  
Christophe Saliboko  
Christopher Baker  
Christopher Ison  
Christy E Stewart  
Christy S Scoggins  
Cindy Johnson  
Claudia Shacklette  
Connie Hong

Consolee Mulindahabi  
Corey Silvia  
Cory Reed  
Damon M Lewis  
Damon Smith  
Dana Rowland  
Darcie Oubezza  
Darlene Poisson  
David Bleckley  
David Grusky  
David Middleton  
Dean Michalak  
Debbie Johnson  
Deborah Kay Dietz  
Debra A Pound  
Denise Henderson  
Denise L Nein  
Diana Magnuson  
Don Badrack  
Epa Uwimana  
Erick Wingfield  
Erin Meyer  
Felix Hayes  
Felix Ramirez  
Francesca Marini  
Fred Krenke  
Freddie Branham  
Gina Erickson  
Girsie D Lee  
Halla Joumaa-Jouney  
Helen Keeton  
Holly Neilson  
Huaigu Cao  
Ian Dodd  
Jackie Alexander  
Jacquese Dubose  
James Christy  
James Richardson  
Jan Dickenson  
Jarred M Roberts  
Jason Hall  
Jeff Behler  
Jeffrey A Evans  
Jeffrey Monroe  
Jennifer L Violette  
Jennifer Seitz

Jerry Collins  
Jim Csoka  
Joey Morales  
John Abowd  
John Elting  
John Friedman  
John Haskill  
John Sullivan  
John T Smith  
John Trambaugh  
Jon Morgan  
Jonathan Cropper  
Jonathan Fisher  
Juan Carlos Humad  
Judy M Brading  
Jumanne Young  
Kanese Gardner  
Karen McGuffin  
Karen Walker  
Katie Genadek  
Keith Finlay  
Kelsey Drotning  
Kent Patterson  
Kevin Deardorff  
Krista Chan  
Lara Karacasu  
Laveda Hall  
Lesley Wallis  
Linda G Robinson  
Linda Vernooy  
Lisa Carroll  
Lisa Donaldson  
Lisa Kelly  
London Johnson  
Lucia Foster  
Lukas Althoff  
Marcus Whitesides  
Margaret Levenstein  
Marguerite Molner  
Mark Weng  
Martia L Riley  
Marvin Credle  
Mary A O Bryan  
Mary Buchanan  
Mary Holbrook  
Mary O'Rourke

Mary Reynolds  
Matthew Bouch  
Matthew Lenfert  
Melissa Banzhaf  
Michael Baker  
Michael K Hunter  
Michael Winggerd  
Michele K Middleton  
Michelle Karlsson  
Michelle Snow  
Mindi Hanna  
Morgan Miller  
Nathaniel Kly  
Nevada Basdeao  
Nhut K Dao  
Nick Orsini  
Nicole Lane  
Nilofer Rajput  
Norm Johnson  
Norma B Ettling  
Patricia De la Cruz  
Paul Friday  
Paula Moore  
Peggy Robinson  
Phyllis M Smitson  
Prem Natarajan  
Raj Chetty  
Rana W Alwazani  
Randy Becker  
Rhonda L Tobey  
Richard Flagg  
Robert Karlen  
Robin Lawrence  
Robin Roberts  
Rodney Gaddis  
Rolanda R Scherer  
Ron Seymour  
Roy Ashley  
Ryann Brown  
Sabdi Taylor  
Sallie Keller  
Sam Jones  
Samantha Gognat  
Sandra E Rice  
Sandra L Post  
Sara Roseman

Sarah Herman  
Sarah Oppenhiemer  
Shanna M Benson  
Shannon Felton Spence  
Shannon G Koenig  
Sharon Moitoso  
Sherry L Nichols  
Silvia N Nava  
Sohial Kamdar  
Sumit Khaneja  
Susan Heitz-Carpenter  
Tahir Zeb  
Tammy Ann Shelton  
Tammy Samuels  
Tara Jones  
Taylor Hanson  
Teresa Holmes Scott  
Terrah A Bruner  
Terri Kennerly  
Terri Sheckles  
Terrie L Henderson  
Tetyana Volkova  
Theresa Toole  
Thomas Sherman  
Timothy Morris  
Todd Gardner  
Tonya L Roberts  
Toshica Harvey  
Tracy Coulter  
Trent Alexander  
Troy Erickson  
Tynell Johnson  
Vincent Biller  
Wendy R Montgomery  
Winona Miniard  
Xujun Peng  
Yukari Nakata

# Quick Data Access

# Large-scale linked data infrastructure in the modern U.S.

<b>Decennial censuses</b>	1940				2000 2010 2020
<b>IRS Form 1040 returns</b>					

# Large-scale linked data infrastructure in the modern U.S.

## Available NOW!

<b>Decennial censuses</b>	1940				2000 2010 2020
<b>IRS Form 1040 returns</b>		1969 1974	1979 1984	1989 1994	

# Large-scale linked data infrastructure in the modern U.S. Coming Soon (hopefully by the end 2026)

<b>Decennial censuses</b>	1940 1950 1960	1970	1980	1990	2000 2010 2020
<b>IRS Form 1040 returns</b>		1969 1974	1979 1984	1989 1994 1995	



# Standard Application Portal

## IRS Form 1040 External Commingled File

Dataset

### Description

This dataset contains information from the IRS Form 1040 tax data that has been approved for use by ...

[show more](#)

Request access

Begin an application for these data by adding them to your request.

#### Metadata

#### Identification and Summary

Scope and Coverage

Detailed Methodology

Data Access

Application-related

### Metadata: Identification and Summary

#### Title ?

IRS Form 1040 External Commingled File

#### Description ?

This dataset contains information from the IRS Form 1040 tax data that has been approved for use by external researchers. Taxpaying households report their income data on the Form 1040 Individual Income Tax Return. These data files were derived from the IRS's Individual Master File (IMF) from 1969, 1974, 1979, 1984, 1989, and 1994, and they were processed by the Census Bureau for use by the General Revenue Sharing program and for generating inter-censal migration estimates. The Census Bureau has also added Protected Identification Keys and Master Address File

# Federal Statistical Research Data Centers (FSRDC)





Contact information:

Katie Genadek – [katie.r.genadek@census.gov](mailto:katie.r.genadek@census.gov)

Trent Alexander – [jtalex@umich.edu](mailto:jtalex@umich.edu)

[www.censusdigitization.org](http://www.censusdigitization.org)

[www.census.gov/dcdl](http://www.census.gov/dcdl)

**1960**  
census

**1970**  
CENSUS OF  
POPULATION  
AND HOUSING

**1980**  
Census of  
Population

CENSUS '90

