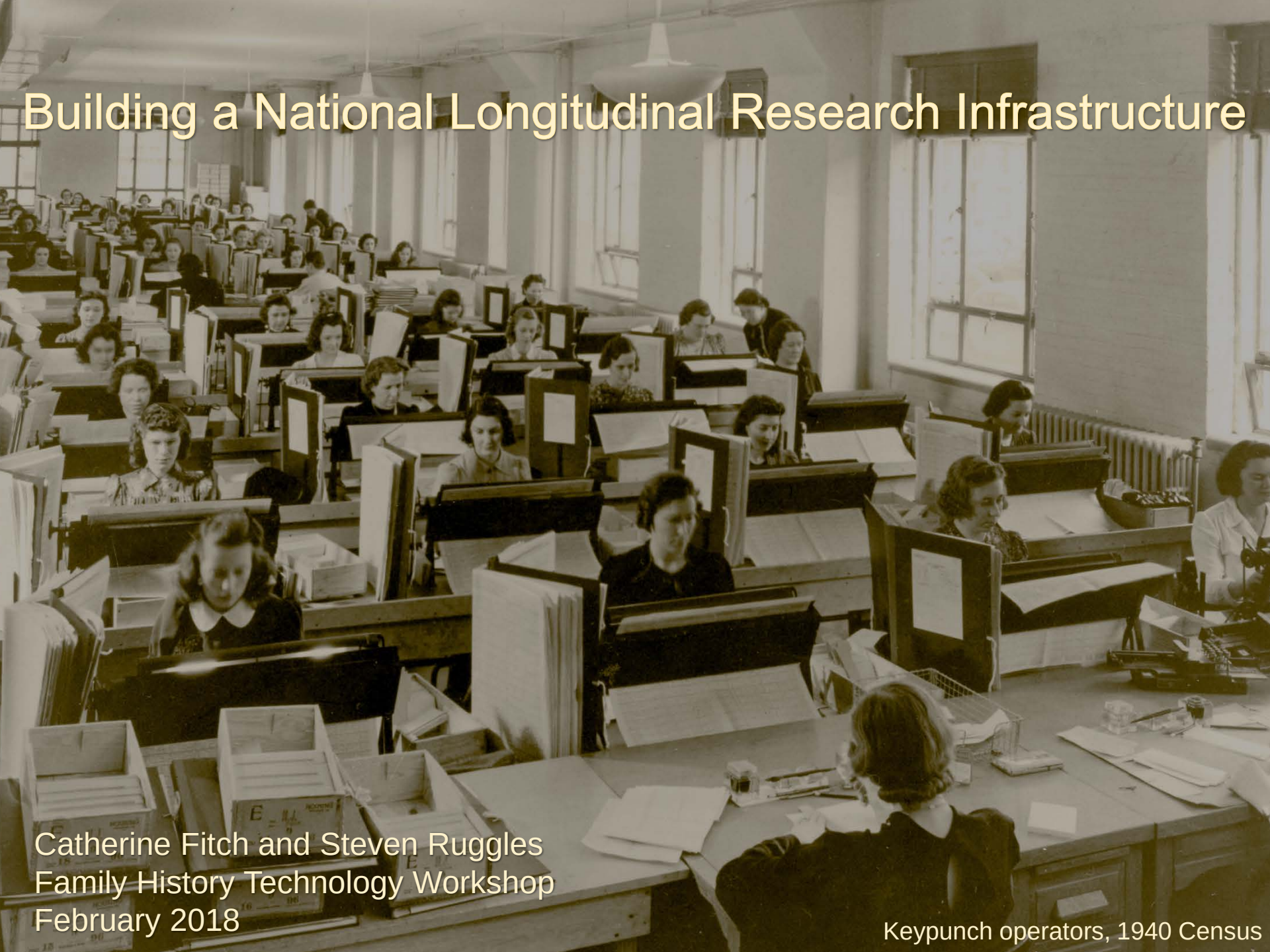
Building a National Longitudinal Research Infrastructure



Catherine Fitch and Steven Ruggles
Family History Technology Workshop
February 2018

Keypunch operators, 1940 Census

Each person in the world creates a Book of Life. This Book starts with birth and ends with death. Record linkage is the name of the process of assembling the pages of this Book into a volume.

- Halbert L. Dunn, 1946

Big Data

Transactional or "Organic" Data

- Administrative
  - Social Security
  - Medicare
  - Military
  - Taxes
- Commercial
  - Credit ratings
  - Phone records
- Social Media

Designed Data

- Censuses
- Surveys
- Remote sensing
  - satellite imagery
  - weather stations

The biggest payoff will lie in new combinations of designed data and organic data, not in one type alone

- Robert Groves, 2011

Organic/Transactional data is voluminous, but

- shallow (few variables) and
- non-representative

Both problems can be overcome by linking to Designed data

Abstract
2013-11-15

Håkan Sjöberg, Andreas Blomquist

Swedish registers - a gold mine for medical research

Sweden and other Nordic countries are unique in the sense that we hold detailed comprehensive registers with information on all the registered population. The records are of great value for Swedish research institutions and often yield competitive advantages towards researchers outside of the Nordic countries, unable to access similar data.

An independent international group of six renowned scientists, including two from the U.S. and one each from Great Britain, Holland, Denmark and Finland, have evaluated the Swedish public health research. Their positive review was summarized by the following statement:

"Sweden is one of the world leaders in PHR (public health research) – in several areas of PHR it is at the fore-front of research exemplified by publications in top international scientific journals. Among these areas are epidemiology and register based research, ... This is due to a unique combination of an excellent data infrastructure, an enlightened public sector and a productive public health research workforce." (Kamper-Jørgensen, 2005)

The prerequisite for the positive rating is a tradition of having comprehensive national data registers and in a system where personal identification number makes it possible to track individuals over long periods of time in longitudinal studies.

National Longitudinal Research Infrastructure

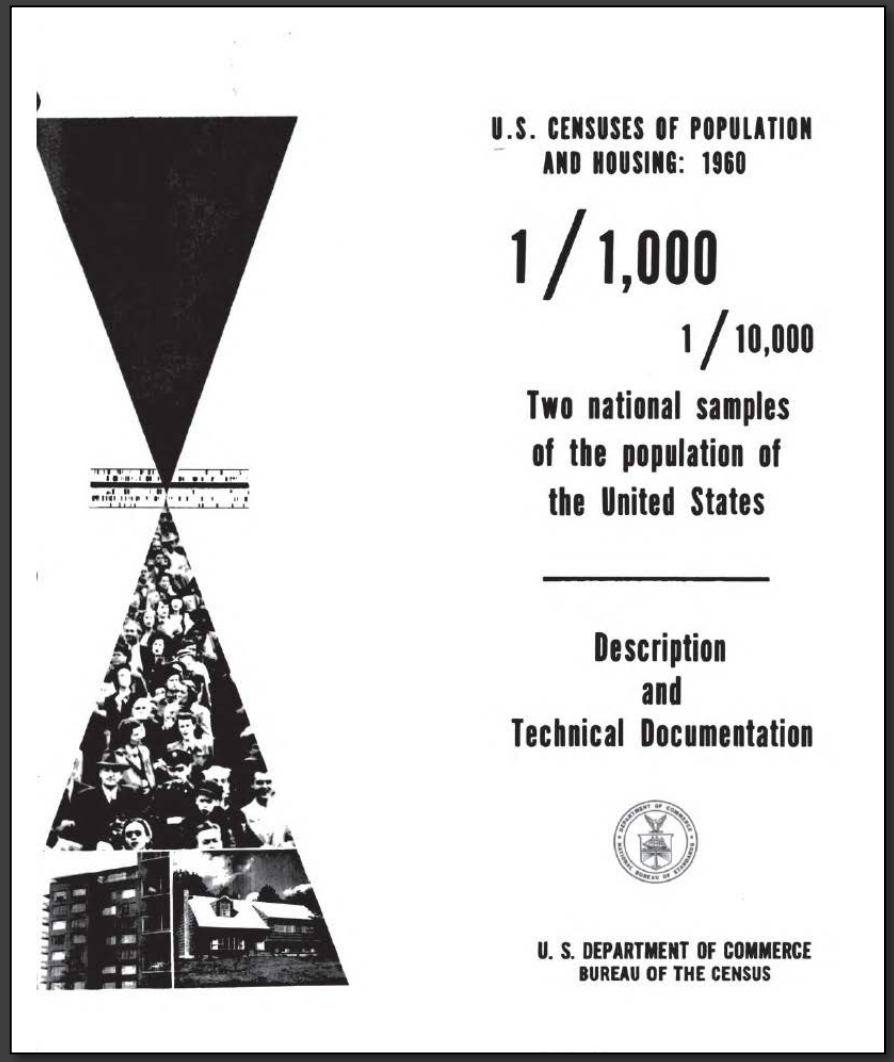
Life histories for each person

- Censuses
- Social Security
- Military records (draft, enlistment)
- Vital records (birth, death, marriage, divorce)
- Health (Medicare, Medicaid)
- Surveys

National Longitudinal Research Infrastructure

Link across 5+ generations, 1850-2020

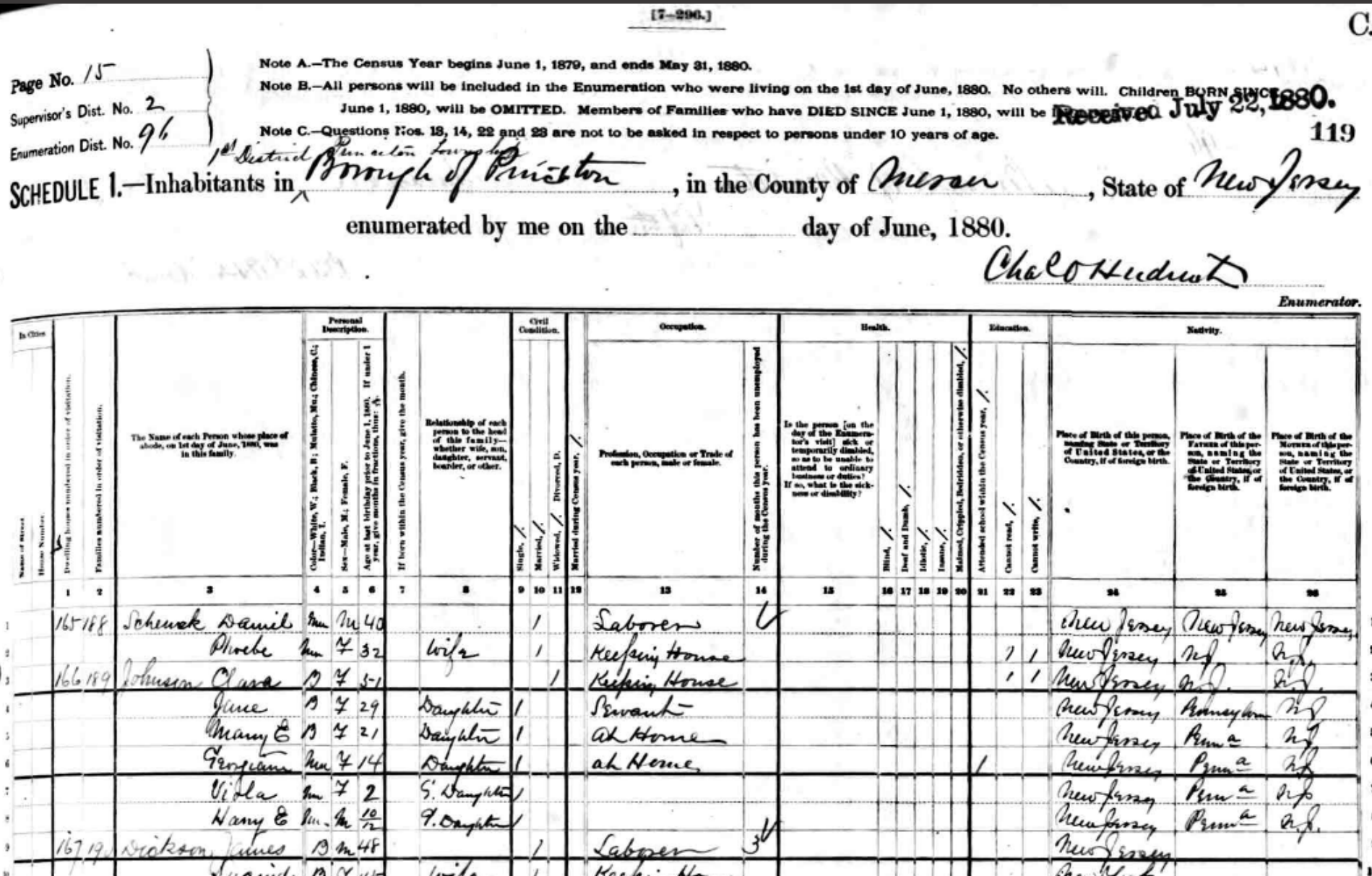
The First Microdata: The 1960 Census Samples



Cover, 1960 Census Microdata Codebook

Distributed on 13 Univac Tapes
(or 18,000 punchcards)

Historical Data



[7–296.]

C.

Page No. 15
Supervisor's Dist. No. 2
Enumeration Dist. No. 96

Note A.—The Census Year begins June 1, 1879, and ends May 31, 1880.
Note B.—All persons will be included in the Enumeration who were living on the 1st day of June, 1880. No others will. Children BORN SINCE June 1, 1880, will be OMITTED. Members of Families who have DIED SINCE June 1, 1880, will be I
Note C.—Questions Nos. 18, 14, 22 and 23 are not to be asked in respect to persons under 10 years of age.

Received July 22, 1880.

119

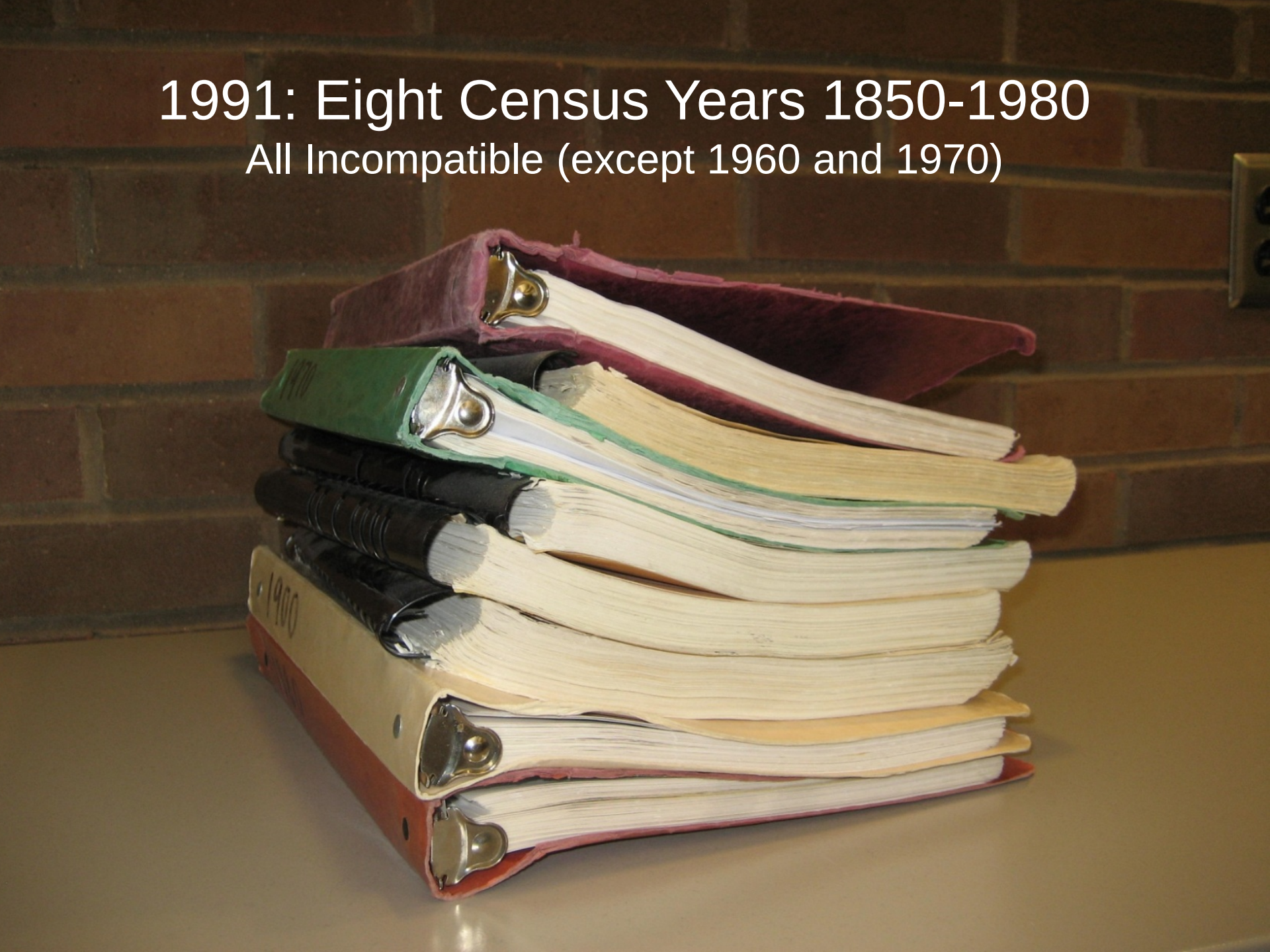
SCHEDULE 1.—Inhabitants in 1st District Princeton Township Borough of Princeton, in the County of Mercer, State of New Jersey enumerated by me on the ____ day of June, 1880.

Chas O Hedrick
Enumerator.

In Cities: Name of Street / House Number	Dwelling houses numbered in order of visitation	Families numbered in order of visitation	The Name of each Person whose place of abode, on 1st day of June, 1880, was in this family.	Personal Description: Color—White, W.; Black, B.; Mulatto, Mu.; Chinese, Ch.; Indian, I.	Sex—Male, M.; Female, F.	Age at last Birthday prior to June 1, 1880. If under 1 year, give months in fractions, thus: 6/12.	If born within the Census year, give the month.	Relationship of each person to the head of this family—whether wife, son, daughter, servant, boarder, or other.	Civil Condition: Single, /; Married, /; Widowed, /; Divorced, D.	Married during Census year, /.	Occupation: Profession, Occupation or Trade of each person, male or female.	Number of months this person has been unemployed during the Census year.	Health: Is the person [on the day of the Enumerator's visit] sick or temporarily disabled, so as to be unable to attend to ordinary business or duties? If so, what is the sickness or disability?	Blind, /.	Deaf and Dumb, /.	Idiotic, /.	Insane, /.	Maimed, Crippled, Bedridden, or otherwise disabled, /.	Education: Attended school within the Census year, /.	Cannot read, /.	Cannot write, /.	Nativity: Place of Birth of this person, naming State or Territory of United States, or the Country, if of foreign birth.	Place of Birth of the Father of this person, naming the State or Territory of United States, or the Country, if of foreign birth.	Place of Birth of the Mother of this person, naming the State or Territory of United States, or the Country, if of foreign birth.	
	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	
1	165	188	Schenck Daniel	Wm	M	40				1		Laborer		6										New Jersey	New Jersey	New Jersey	1
2			Phoebe	Wm	F	32		Wife		1		Keeping House										1	1	New Jersey	N.J.	N.J.	2
3	166	189	Johnson Clara	B	F	5-1					1	Keeping House										1	1	New Jersey	N.J.	N.J.	3
4			Jane	B	F	29		Daughter	1			Servant												New Jersey	Kentucky	N.J.	4
5			Mary E	B	F	21		Daughter	1			At Home												New Jersey	Penn^a	N.J.	5
6			Georgeanna	Wm	F	14		Daughter	1			at Home										1		New Jersey	Penn^a	N.J.	6
7			Viola	Wm	F	2		G. Daughter	/															New Jersey	Penn^a	N.J.	7
8			Mary E	Wm	F	10/12		G. Daughter	/															New Jersey	Penn^a	N.J.	8
9	167	190	Dickson James	B	M	48				1		Laborer		3										New Jersey			9
10			Amanda	B	F	45		Wife		1		Keeping House															10

1991: Eight Census Years 1850-1980

All Incompatible (except 1960 and 1970)

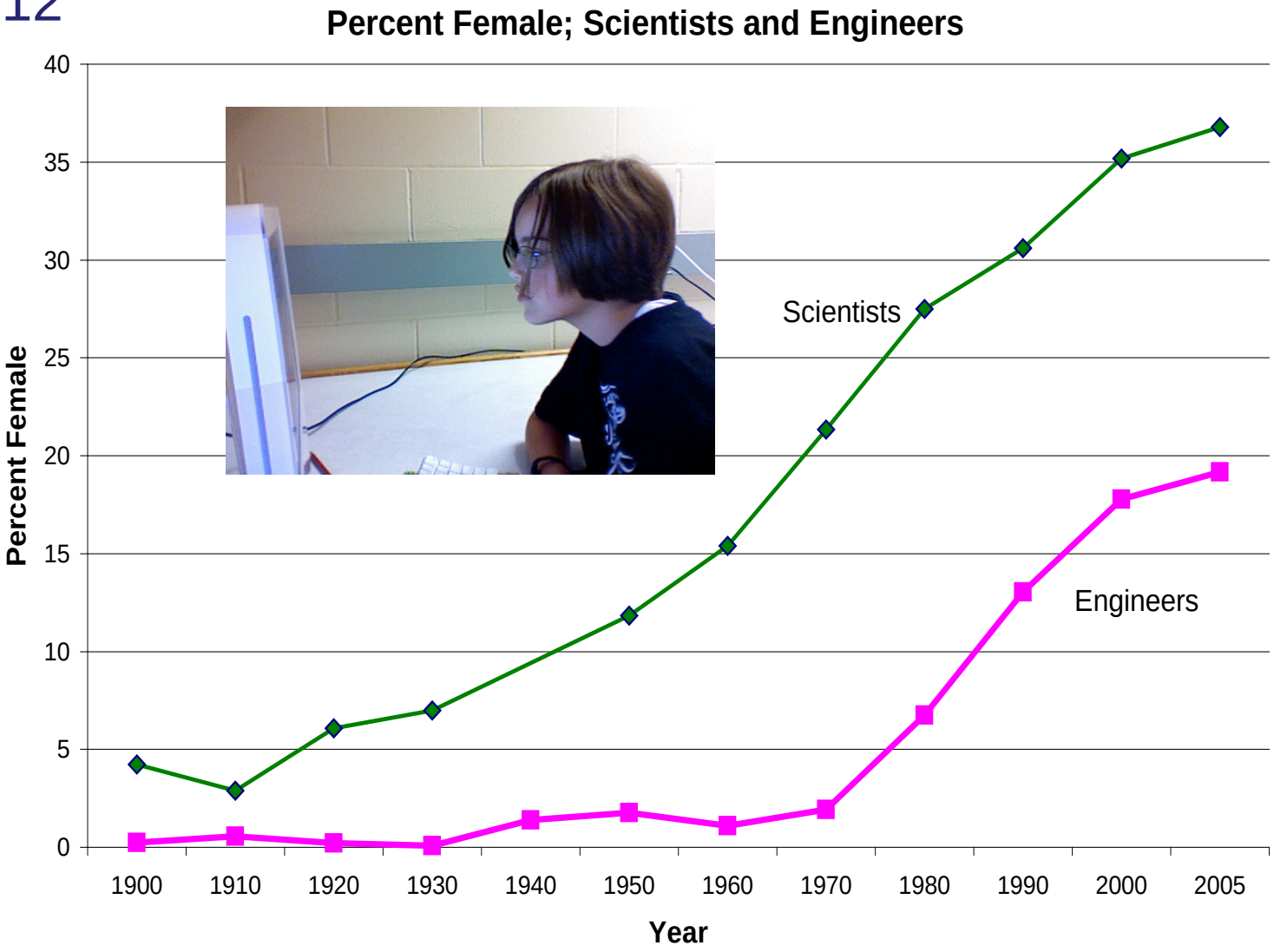


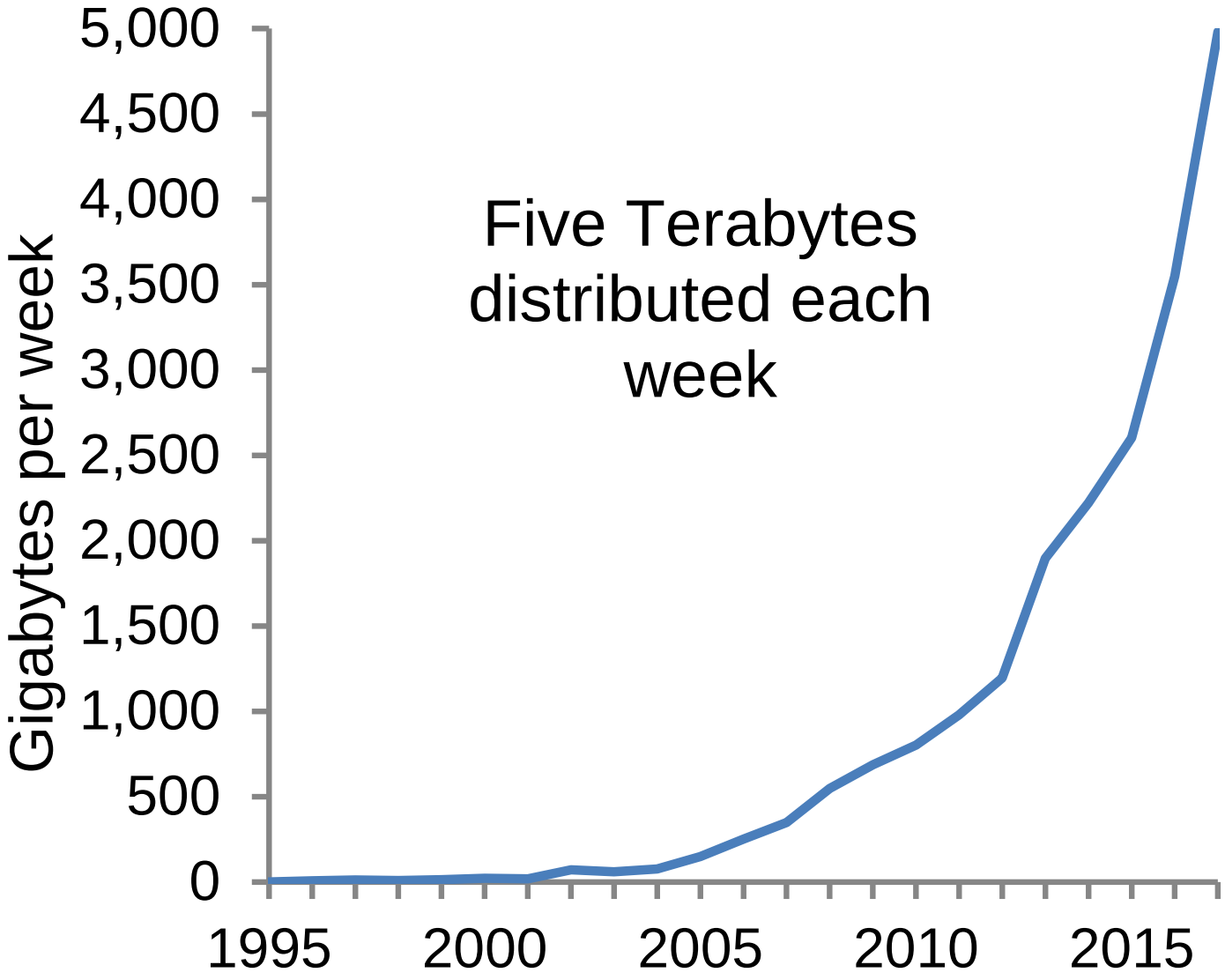
1991 IPUMS proposal: An integrated database for 1880, 1900, 1910, 1940, 1950, 1960, 1970, 1980, 1990

- ✓Harmonized codes
- ✓Consistent record layout
- ✓Integrated documentation
- ✓No loss of information

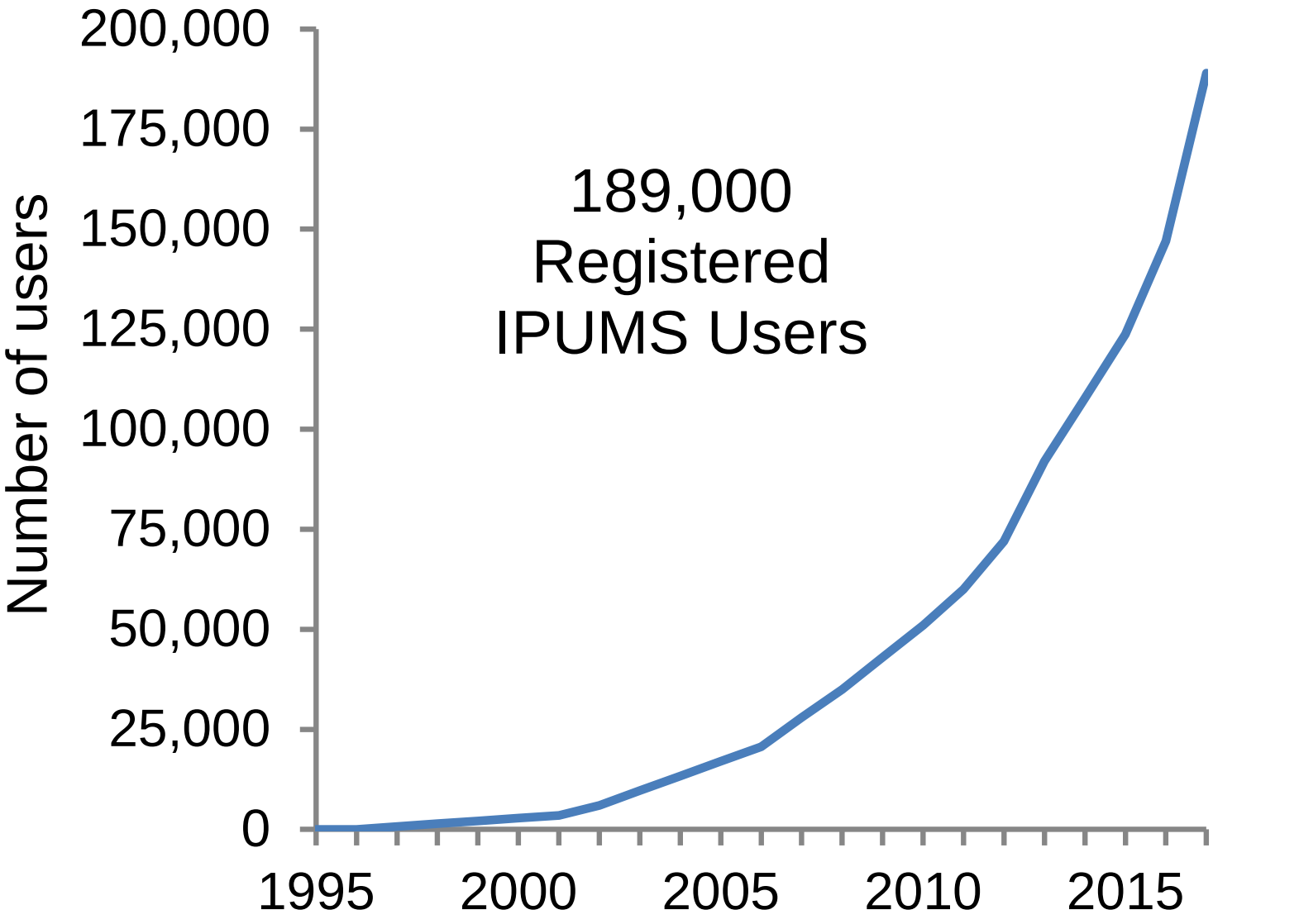
.

IPUMS Graph from “A Century of Women in Science and Engineering,” History Day project by Abby Norling-Ruggles, age 12

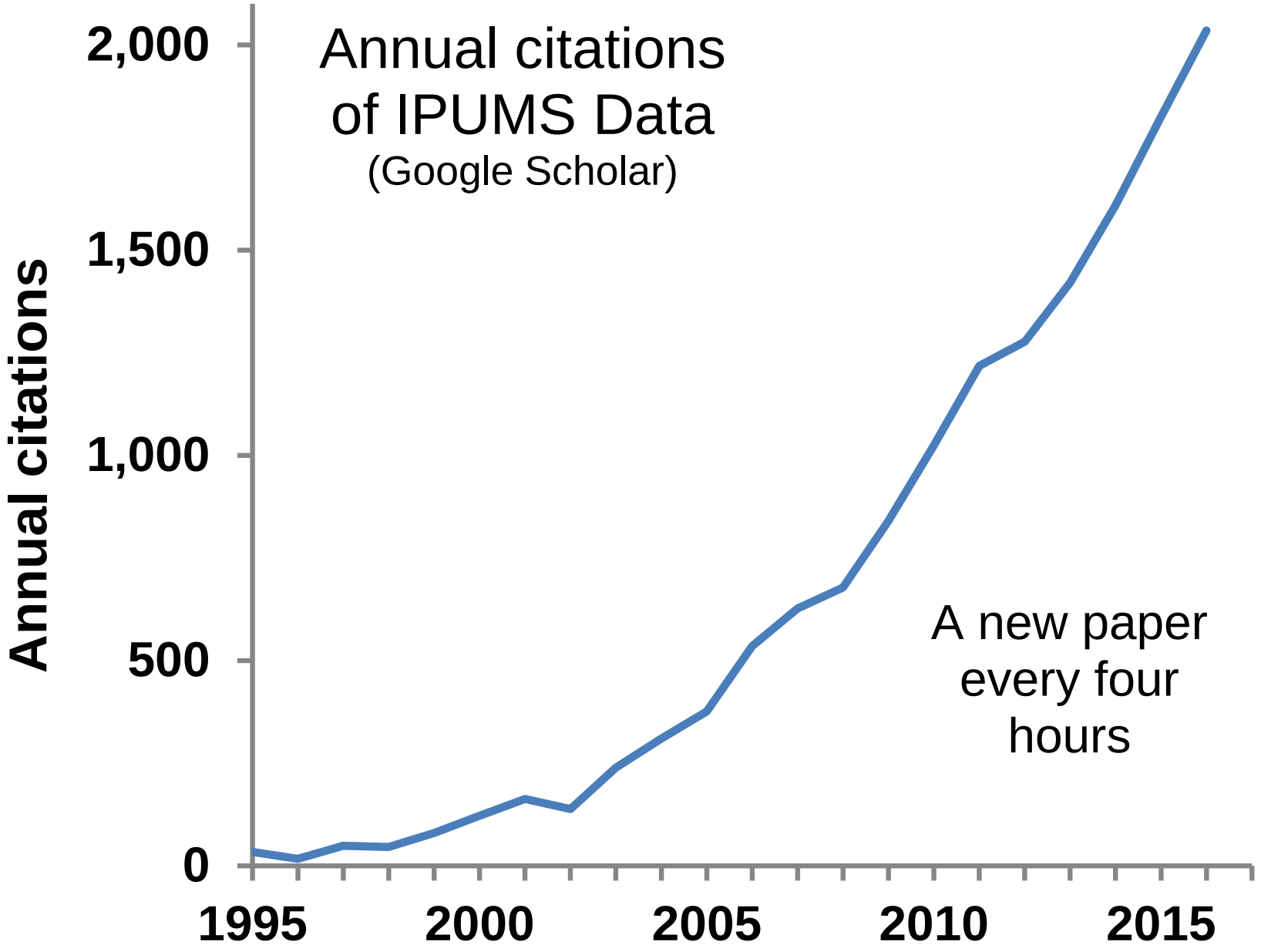




IPUMS Data Dissemination, 1995-2017



Registered IPUMS data users, 1995-2017


Annual citations of IPUMS Data
(Google Scholar)
Annual citations
2,000
1,500
1,000
500
0
1995
2000
2005
2010
2015
A new paper every four hours

IPUMS USA

usa.ipums.org

IPUMS.ORG | SELECT DATA | FAQ | HELP | LOGIN

DATA
- BROWSE AND SELECT DATA
- DOWNLOAD OR REVISE EXTRACTS
- ANALYZE DATA ONLINE
- IPUMS REGISTRATION

DOCUMENTATION
- VARIABLES
- SAMPLES
- USER'S GUIDE
- GEOGRAPHIC TOOLS
- FAQ

RESOURCES
- ENUMERATION FORMS
- PUBLISHED CENSUS VOLUMES
- ERRATA AND REVISIONS

RESEARCH
- CITATION AND USE
- BIBLIOGRAPHY
- RELATED SITES

CONTACT US
- HELP

U.S. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH

IPUMS USA collects, preserves and harmonizes U.S. census microdata and provides easy access to this data with enhanced documentation. Data includes decennial censuses from 1790 to 2010 and American Community Surveys (ACS) from 2000 to the present.

Use it for GOOD -- never for EVIL

CREATE YOUR CUSTOM DATA SET

GET DATA

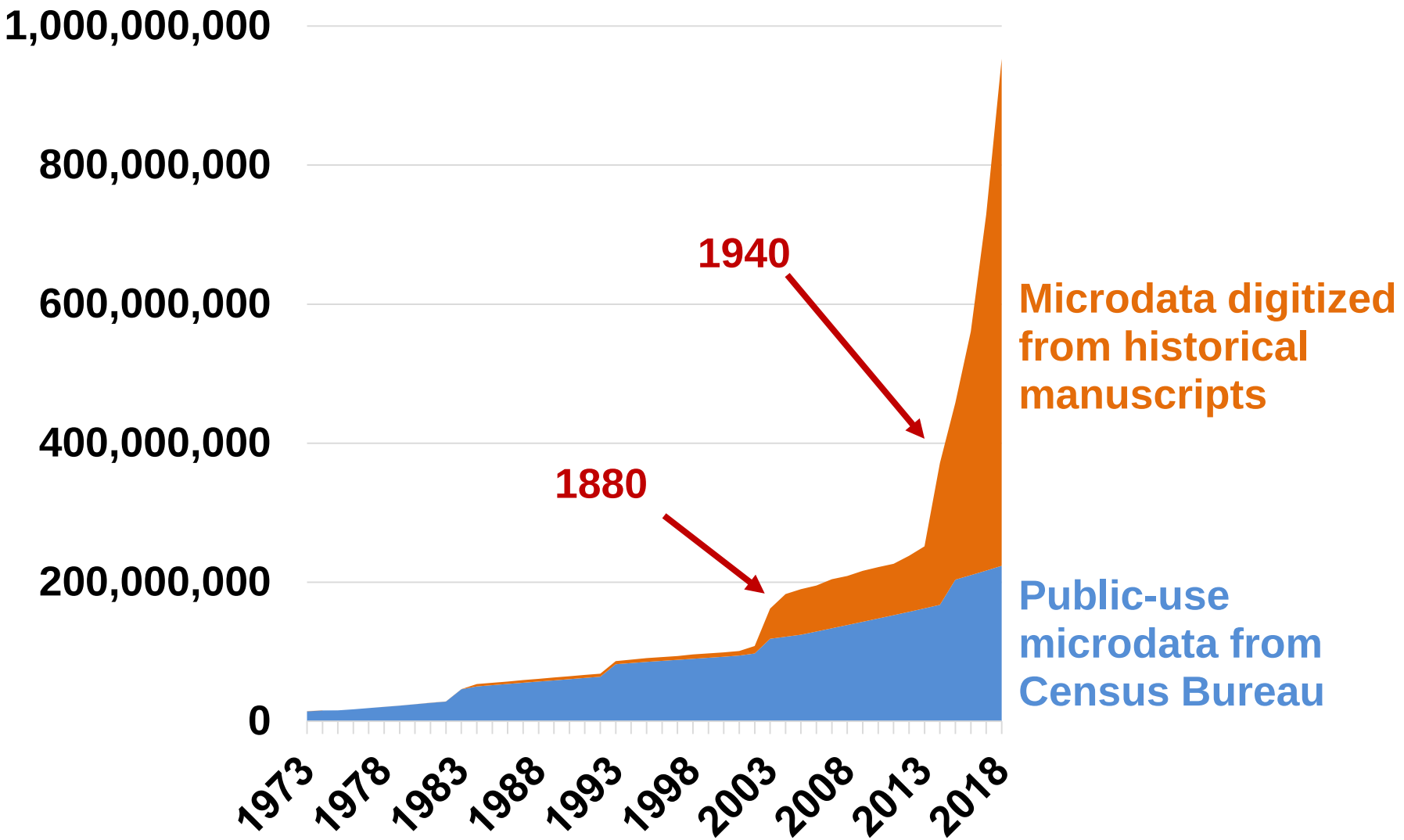
USE OUR ONLINE TOOL FOR ANALYSIS

ANALYZE DATA ONLINE

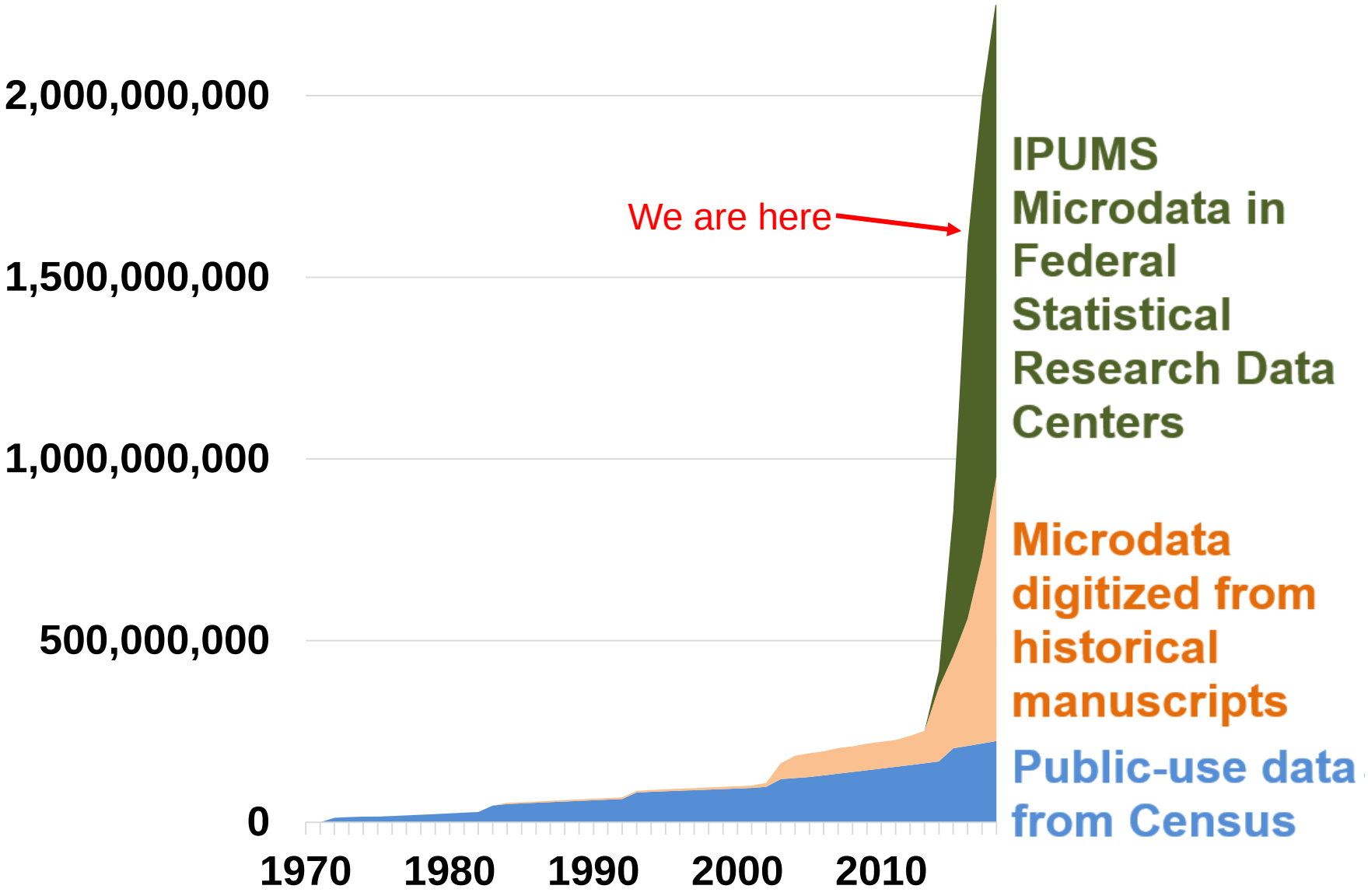
WHAT IS IPUMS?

IPUMS provides census and survey data from around the world integrated across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community context. Data and services available free of charge.

U.S. public use microdata available for research, 1973-2018
(number of person records)



Integrated U.S. microdata available for research, 1970-2018
(number of person records)



United States Census Bureau

TOPICS Population, Economy | GEOGRAPHY Maps, Products | LIBRARY Infographics, Publications | DATA Tools, Developers | SURVEYS/PROGRAMS Respond, Survey Data

Census.gov > About the Bureau > Federal Statistical Research Data Centers

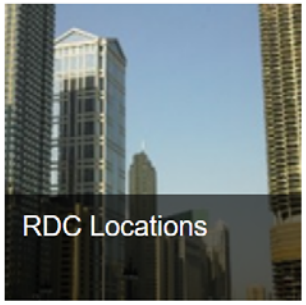
Federal Statistical Research Data Centers

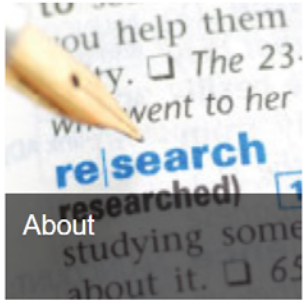
- About this Section
- Available Data
- Federal Partners
- Research Data Centers
- Events
- Updates
- Contact Information

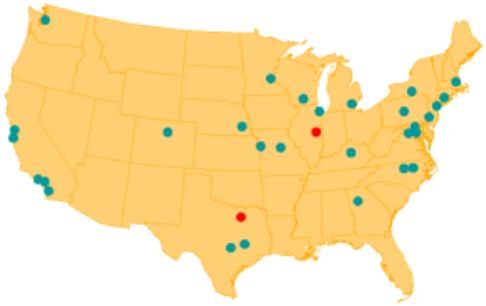
FSRDC

Federal Statistical Research Data Centers are partnerships between federal statistical agencies and leading research institutions. They are secure facilities providing authorized access to restricted-use microdata for statistical purposes only.

Read More







30 locations and growing

NLRI Collaboration

- Census Longitudinal Infrastructure Project
- IPUMS Multigenerational Longitudinal Panel

The Census Longitudinal Infrastructure Project (CLIP)

Sanders

Ferrie

O’Hara

Alexander

1940 Linking Meeting
Minneapolis, February 10-11, 2014

1940 Census

SSA
Numident

2010
uses

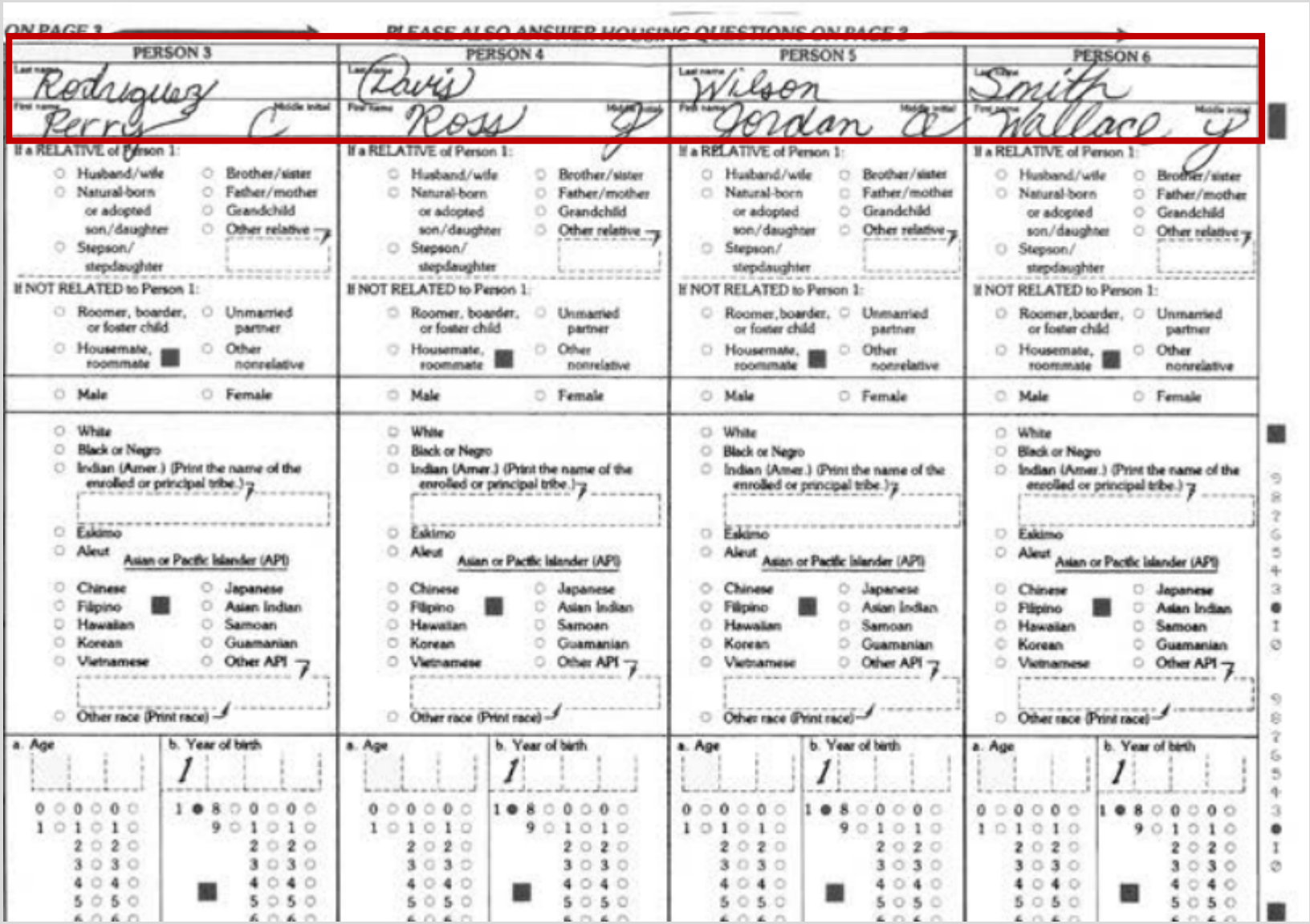
WW II
Military

HUD
Selective Service
Private Vendors

Federal
Surveys

CLIP Linking
Strategy

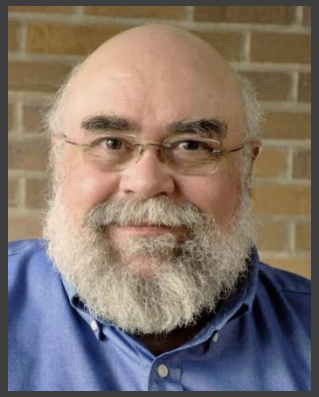
Capturing names in the 1990 census through OCR



ON PAGE 3 → PLEASE ALSO ANSWER HOUSING QUESTIONS ON PAGE 3 →

PERSON 3	PERSON 4	PERSON 5	PERSON 6
Last name: Rodriguez	Last name: Davis	Last name: Wilson	Last name: Smith
First name: Perry Middle initial: C	First name: Ross Middle initial: J	First name: Jordan Middle initial: A	First name: Wallace Middle initial: J
If a RELATIVE of Person 1: ○ Husband/wife ○ Natural-born or adopted son/daughter ○ Stepson/ stepdaughter ○ Brother/sister ○ Father/mother ○ Grandchild ○ Other relative →	If a RELATIVE of Person 1: ○ Husband/wife ○ Natural-born or adopted son/daughter ○ Stepson/ stepdaughter ○ Brother/sister ○ Father/mother ○ Grandchild ○ Other relative →	If a RELATIVE of Person 1: ○ Husband/wife ○ Natural-born or adopted son/daughter ○ Stepson/ stepdaughter ○ Brother/sister ○ Father/mother ○ Grandchild ○ Other relative →	If a RELATIVE of Person 1: ○ Husband/wife ○ Natural-born or adopted son/daughter ○ Stepson/ stepdaughter ○ Brother/sister ○ Father/mother ○ Grandchild ○ Other relative →
If NOT RELATED to Person 1: ○ Roomer, boarder, or foster child ○ Housemate, roommate ■ ○ Unmarried partner ○ Other nonrelative	If NOT RELATED to Person 1: ○ Roomer, boarder, or foster child ○ Housemate, roommate ■ ○ Unmarried partner ○ Other nonrelative	If NOT RELATED to Person 1: ○ Roomer, boarder, or foster child ○ Housemate, roommate ■ ○ Unmarried partner ○ Other nonrelative	If NOT RELATED to Person 1: ○ Roomer, boarder, or foster child ○ Housemate, roommate ■ ○ Unmarried partner ○ Other nonrelative
○ Male ○ Female	○ Male ○ Female	○ Male ○ Female	○ Male ○ Female
○ White ○ Black or Negro ○ Indian (Amer.) (Print the name of the enrolled or principal tribe.) → ○ Eskimo ○ Aleut	○ White ○ Black or Negro ○ Indian (Amer.) (Print the name of the enrolled or principal tribe.) → ○ Eskimo ○ Aleut	○ White ○ Black or Negro ○ Indian (Amer.) (Print the name of the enrolled or principal tribe.) → ○ Eskimo ○ Aleut	○ White ○ Black or Negro ○ Indian (Amer.) (Print the name of the enrolled or principal tribe.) → ○ Eskimo ○ Aleut
Asian or Pacific Islander (API): ○ Chinese ○ Filipino ■ ○ Hawaiian ○ Korean ○ Vietnamese ○ Japanese ○ Asian Indian ○ Samoan ○ Guamanian ○ Other API →	Asian or Pacific Islander (API): ○ Chinese ○ Filipino ■ ○ Hawaiian ○ Korean ○ Vietnamese ○ Japanese ○ Asian Indian ○ Samoan ○ Guamanian ○ Other API →	Asian or Pacific Islander (API): ○ Chinese ○ Filipino ■ ○ Hawaiian ○ Korean ○ Vietnamese ○ Japanese ○ Asian Indian ○ Samoan ○ Guamanian ○ Other API →	Asian or Pacific Islander (API): ○ Chinese ○ Filipino ■ ○ Hawaiian ○ Korean ○ Vietnamese ○ Japanese ○ Asian Indian ○ Samoan ○ Guamanian ○ Other API →
○ Other race (Print race) →	○ Other race (Print race) →	○ Other race (Print race) →	○ Other race (Print race) →
a. Age / b. Year of birth: 1 ■ 1 9 8 0	a. Age / b. Year of birth: 1 ■ 1 9 8 0	a. Age / b. Year of birth: 1 ■ 1 9 8 0	a. Age / b. Year of birth: 1 ■ 1 9 8 0

Multigenerational Longitudinal Panel



Ruggles

Warren

Fitch

Hacker

Sobek

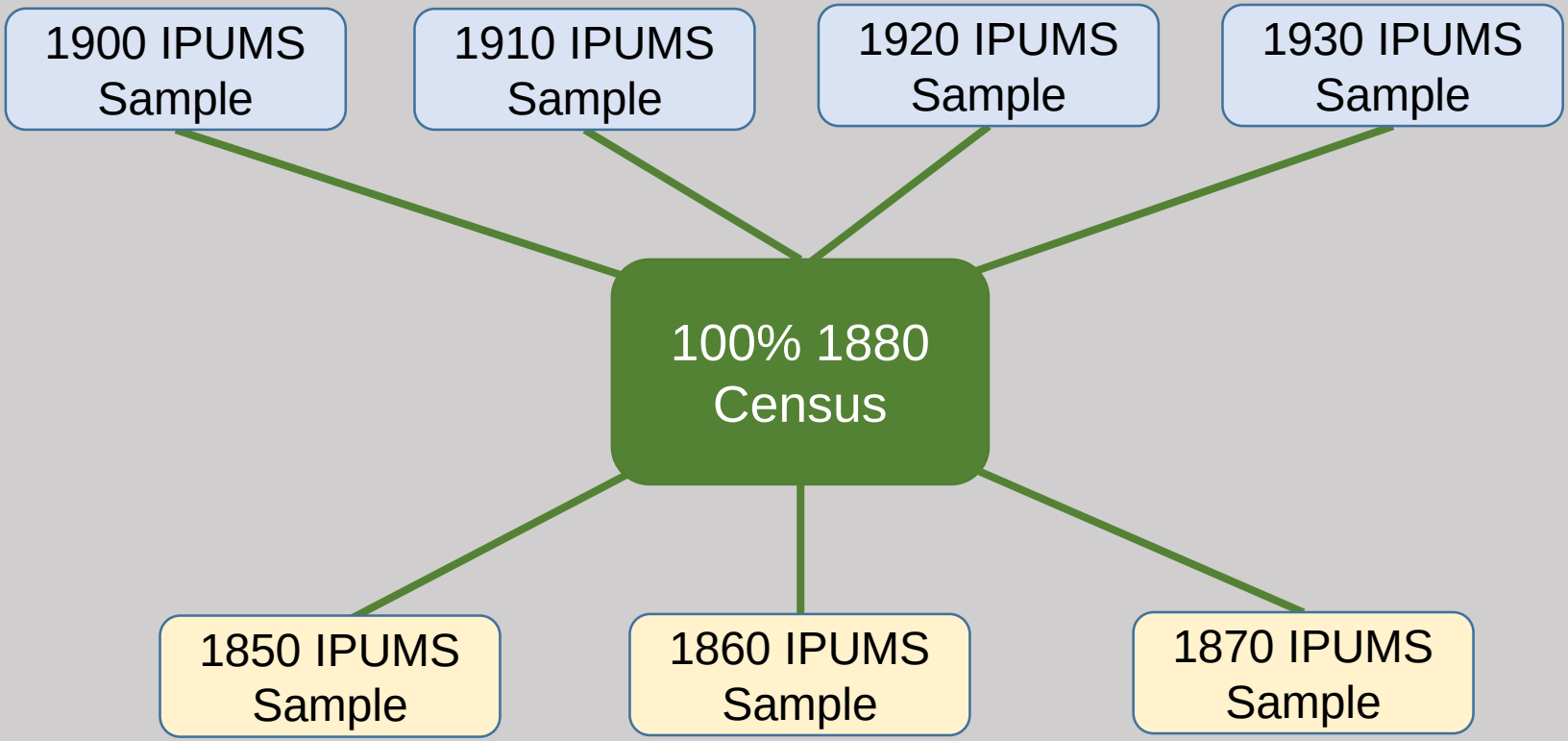
Roberts

Bailey

Goeken

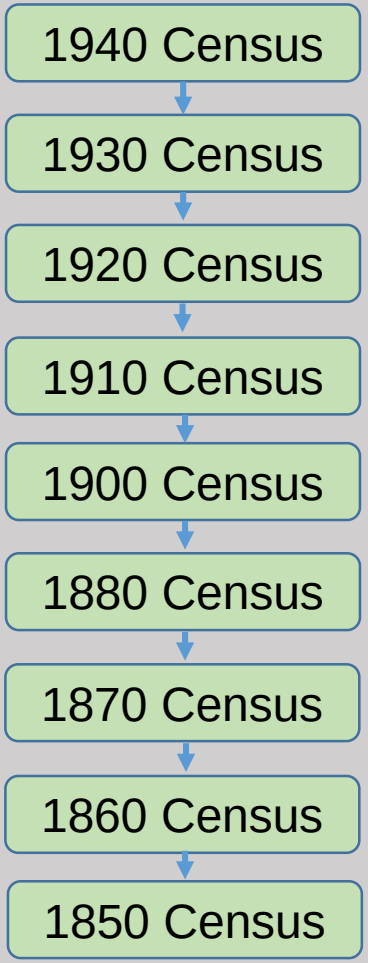
Price

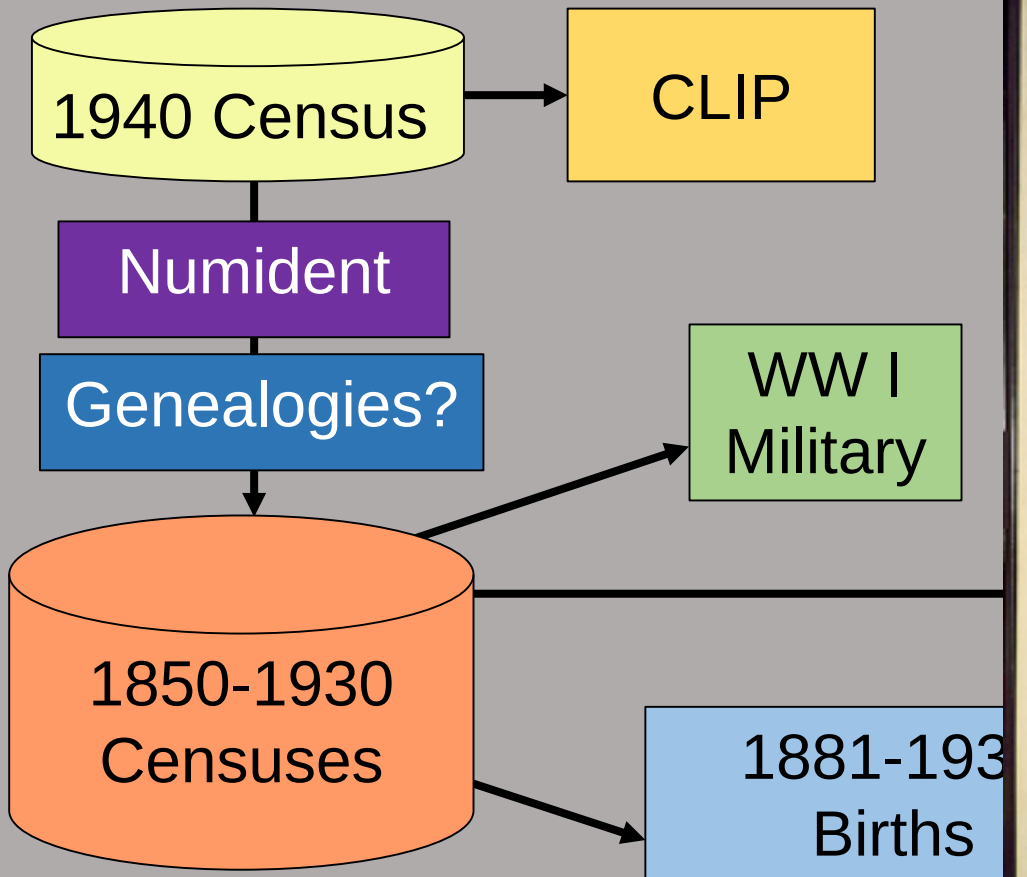
IPUMS Linked Representative Samples



Final version June 2010

Multigenerational Longitudinal Panel





MLP Linking Strategy

National Longitudinal Research Infrastructure

Life histories for each person

- Impact of early life conditions on later health and well-being
- Social, Economic, Geographic Mobility
- Life course transitions

National Longitudinal Research Infrastructure

Link across 5+ generations

- Impact of forebears on health and well-being
- Socioeconomic mobility across generations: Do we have dynasties?

National Longitudinal Research Infrastructure

Understanding the great transformations:

demographic transition, family transition, urbanization, immigration, industrialization

Economics & Human Biology

Volume 10, Issue 1, January 2012, Pages 98–111

Cognitive disparities, lead plumbing, and water chemistry: Prior exposure to water-borne lead and intelligence test scores among World War Two U.S. Army enlistees

Joseph P. Ferrie[a, b, c], Karen Rolf[d, e], Werner Troesken[f, c]

Higher prior exposure to water-borne lead among male World War Two U.S. Army enlistees was associated with lower intelligence test scores. Exposure was proxied by urban residence and the water pH levels of the cities where enlistees lived in 1930.

National Longitudinal Research Infrastructure

- Impact of lead exposure on Alzheimer's disease
- Effect of early-life cognitive capacity on later economic success
- Transmission of health and well-being over multiple generations
- Effects of early-life income support on later outcomes

Thank You.