Social Science History Association Annual Meeting
Building and Interpreting Censuses III | Session 193 | November 21, 2026

Building Social Science Data Infrastructure: The 1900, 1910, 1940 and 1950 Public Use Census Samples

Diana L. Magnuson and Steven Ruggles
University of Minnesota
September 2026

Abstract

This paper documents creation of the 1900, 1910, 1940 and 1950 U.S. census microdata samples. In 1963, the Census Bureau produced the first public use microdata sample for the 1960 census, stimulating a surge of social science research. In the 1970s, two separate research teams independently embarked on projects to create historical public use samples for earlier censuses. Halliman H. Winsborough, in collaboration with the U.S. Census Bureau, oversaw construction of microdata samples of the 1940 and 1950 censuses. Samuel H. Preston directed the 1900 and 1910 public use sample projects. The paper draws on evidence from the National Archives and Records Administration (including Census Bureau internal memoranda, technical census advisory committee minutes, and correspondence). We also use evidence from the Halliman H. Winsborough Collection in the IPUMS Archive, oral histories for both Preston and Winsbrough, and a variety of other sources.

Introduction

In 1963, the Census Bureau created the world's first public use microdata sample: a 1-in-1,000 sample of individuals and households from the 1960 census. The samples included virtually all the information collected in the 1960 census, except that identifying information such as name and address was redacted to preserve respondent confidentiality. Rising public policy demands, expanding computing capacity at universities, and active advocacy from social scientists motivated the Bureau to make the data widely accessible (Magnuson and Ruggles 2026; U.S. Census Bureau 1963).

The success of the 1960 sample stimulated efforts to create microdata samples for earlier censuses. This paper documents the creation of the 1900, 1910, 1940 and 1950 microdata samples. In the 1970s, two separate research teams independently embarked on projects to create historical microdata samples for earlier decennial censuses. Drawing on historical evidence from the National Archives and Records Administration (NARA), including U.S. Census Bureau internal memoranda, technical census advisory committee minutes, and correspondence, as well as the Halliman H. Winsborough Collection (IPUMS Archive at the University of Minnesota), oral histories, and a variety of secondary and primary source materials

gathered from private and public collections, our research aims to uncover the history of the 1900, 1910, 1940 and 1950 microdata samples.

Creating the 1940 and 1950 PUMS (1970-1984)

The creation, availability, and researcher enthusiasm for the 1960 public use sample prompted social scientists and "professional groups" to express to the Census Bureau "the need and value" of additional public use microdata records from historic U.S. censuses (U.S. Census Bureau 1970, September 17; U.S. Census Bureau 1970, October 23). In 1969 sociologist Otis Dudley Duncan recommended that "The idea of creating 1-in-1,000 samples for census years prior to 1960 should be carefully explored," thus enabling researchers to carry out studies of "complexity and sophistication than were possible hitherto" (Duncan 1969, pp. 35 and 12).

A proposal to create historical microdata samples was raised at the February 1970 meeting of the Census Advisory Committee on Population Statistics:

> "One method of removing a gap in the availability of statistics on social change would be the provision of public use samples tapes for 1940, 1950, and 1960 on a basis comparable with the code structure used in the public use sample for 1970. Consideration is being given to doing this for these earlier censuses on the basis of a 1/100 sample, although it would be quite expensive and at present no funds are available for doing this for the earlier censuses … The 1/100 sample for 1970 would permit analysis for areas below the national level. The Committee, in general, agreed that these public use sample tapes should be prepared with priority being given first to the 1970 and 1960 tapes and, as resources permit, continuing with the 1950 and 1940 tapes. A 1/100 sample would be ideal, but if cost precluded doing this for each of the censuses, then a 1/1000 sample tape would be preferable to missing either the 1940 or 1950 censuses" (U.S. Census Bureau 1970, February 20).

The proposal for a new 1/100 sample for 1960 compatible with the samples planned for the 1970 census went forward. That project was comparatively inexpensive, because the necessary data were readily available in machine-readable form and there was no need for new data capture. Moreover, computer programs being developed for processing the new the 1970 census samples could be re-used for the new version of the 1960 data, since the data format would be virtually identical.

The proposed new samples for the 1940 and 1950 censuses were considerably more complicated. The punch cards used to process the 1940 and 1950 decennial censuses were destroyed after their use because of limited storage capacity and deterioration of the cards. The original census questionnaires were preserved on microfilm, but a sample would have to be digitized from the handwritten forms (Turner 1970).

The earliest Census Bureau proposal in February 1970 to create samples of the 1940 and 1950 censuses contemplated manual numeric coding of a systematic sample of records, followed by keypunching onto punch cards (Turner 1970). In addition to the clerical work, the project would involve significant software development for editing the data and allocating missing and inconsistent responses. The proposal estimated a total cost of "about $500,000 or more to create one-percent samples for 1940 and 1950." The report also noted that "the preparation of a more precise cost estimate would require two or three months to develop" (Turner 1970).

A more detailed plan, dated seven months later, proposed using the Census Bureau's high speed optical mark recognition system, the Film Optical Sensing Device for Input to Computers (FOSDIC), a system used for data capture in the 1960 and 1970 censuses (U.S. Census Bureau 1970, September 17). To digitize the sample, clerical staff transcribed census information onto specially designed FOSDIC "bubble sheets" like the forms used for standardized testing. The bubble sheets were then microfilmed, and the FOSDIC machine converted the microfilm into digital data stored on magnetic tape (Ruggles and Magnuson 2020). Table 1 provides the September 1970 cost estimates for the proposed project, then called Public Use Records Samples (PURS).

Table 1. Preliminary cost estimate for preparing public use records samples (PURS) (U.S. Census Bureau 1970, September 17).

Census Year and Sample Size	Approximate Cost
1/100 PURS for 1940 and 1950	$2,000,000
1/100 PURS for 1940 only	$900,000
1/100 PURS for 1950 only	$1,100,000
1/1000 PURS for 1940 and 1950	$500,000
1/1000 PURS for 1940 only	$200,000
1/1000 PURS for 1950 only	$300,000

The cost estimate concluded that "Without a funded project, the developmental work now underway can probably be continued into the spring of next year. After that time, funds will be required to begin writing the necessary computer editing programs and printing the necessary instructions and forms for the clerical coding and transcription work. If this schedule could be maintained, we estimate that the 1940 and 1950 PURS could be available in the last half of calendar year 1972" (U.S. Census Bureau 1970, September 17; U.S. Census Bureau 1970, October 23).

Funding for the project did not materialize and it was not feasible for the Bureau to solely undertake the project within its own budget constraints (Mason et al 1977, p. 5). The Committee on Population Statistics did, however, establish a subcommittee "to study and encourage the

development of 1940 and 1950 samples." The subcommittee was chaired by Karl Taeuber, a demographer affiliated with the Center for Demography and Ecology at the University of Wisconsin (U.S. Census Bureau 1970, September 17).

The search for funding was underway. In 1973 the Interagency Committee on Population Research, which included representatives from many federal agencies, noted the "substantial support that would be needed form one or more sources for the creation and tabulation of (one percent or one in a thousand) public use samples from the 1950 and 1940 cenuses [sic], similar to corresponding samples from the 1970 and 1960 censuses that are available for tabulation by any person or agency with funds to purchase the sample computer tapes" (Interagency Committee on Population Research 1973, pp. 22-23). According to Mason et al, "tentative discussions about such a project were held with officials from the Office of Management and Budget, the Bureau of the Census, the National Science Foundation, the Center for Population Research of the National Institute of Child Health and Human Development, and other agencies" (Mason et al 1977, pp. 5-6).

The research community outside the Census Bureau did not let the idea of creating public use samples from population censuses preceding 1960 drop from its ambitions. Greater numbers of researchers were echoing Otis Dudley Duncan's urging to create samples prior to 1960 and were persuaded that access to large data files held "profound significance to social research" (Mason et al 1977, p. 3).

In June of 1976, a National Science Foundation (NSF) funded conference, attended by three dozen scholars, was held in Madison, Wisconsin, at the Center for Demography and Ecology. The attendees were tasked with "writ[ing] a report that would provide an array of research questions that the proposed data sources can uniquely address, an indication of the importance of such research to ongoing scientific issues and policy concerns, and recommendations about the process of creation of such data files." (Mason et al 1977, p. 9). The two-and-a-half-day workshop produced documents demonstrating that "in principle the creation of the proposed public use data files is in the public interest." Massive social and demographic changes in the American experience in the mid-twentieth century demanded attention and could be better understood by social scientists "only if the proposed data files are produced and utilized" (Mason et al 1977, p. 11).

In the wake of the workshop, NSF agreed to consider a proposal to create 1/100 public use samples for the 1940 and 1950 Censuses. The proposal went through multiple rounds of review and was evaluated by a total of seventy-five reviewers (University of Wisconsin-Madison Archives 1985, p. 5). The grant, funded in March 1978, went to a team at the Center for Demography and Ecology at the University of Wisconsin. The Principal Investigator was Halliman Winsborough, with Co-Principal Investigators Karl Taeuber and Robert Hauser. Most of the work was to be conducted at the Census Bureau under a subaward because the data were protected by census confidentiality rules. The project was considerably more expensive

than the original Census Bureau estimates; the award was for $7,218,075 (about $39 million in 2026 dollars), which was apparently the largest social science grant ever made up to that date.[1]

The proposal called for direct data entry from the microfilm of the enumeration forms using interactive terminals connected to a central computer (U.S. Census Bureau 1984, p. xiii). Ordinarily in this period, data entry was conducted by transcribing information onto paper coding sheets, keypunching the coded data onto cards, and then transferring the data to magnetic tape through a batch process. Reducing the number of steps was expected to yield efficiencies in data capture. Because of the novelty of the approach, however, this strategy also introduced complexity and uncertainty.

From the outset, Winsborough was unenthusiastic about undertaking the project: “We knew ahead of time that this wasn’t going to be fun, and it wasn’t.” (University of Wisconsin-Madison Archives 1985, Day 1, 1:12:40). Winsborough explained, “the project began with a fight with the Census Bureau about how to do it, which we won, because we were right” (1:25:36). The logistical and technological hurdles were significant and consumed many months of fits and starts for both the Bureau and the team at the University of Wisconsin-Madison. The hurdles are described in detail in the minutes and status reports of the 40-50 PUS Project materials (MIN-8 - MIN-94). Workflow issues quickly emerged because of geographic distance and communication issues between Madison (WI), Pittsburg (KS), and the Census Bureau (MD) (University of Wisconsin Madison Archives, pp. 7-8). The project was organized into two separate staffs. In Madison, Halliman Winsborough and his team were responsible for financial oversight, equipment acquisition and “development of specifications for sampling, data transcription, editing, allocation, and other data processing procedures, and preparation of technical documentation” (U.S. Census Bureau 1984, p. xiii). The personnel Records Service Branch in Pittsburg housed the staff responsible for producing the 1940 and 1950 public use microdata products. The Census Bureau staff was responsible for operations including computer programming, staffing the Pittsburg site, managing data processing procedures and preparing technical documentation (U.S. Census Bureau 1984, p. xiii).

The first two years of the project (1978-1980) were consumed with “developing procedures for sampling and data entry, purchasing and installing the computer system in Pittsburg, Kansas, and writing the sampling and data-entry computer programs for the 1940 public use sample” (U.S. Census Bureau 1984, p. xiii). Winsborough wrote in December 1979, “In reviewing the nature of the delays and puzzling about their cause, I conclude that some of our difficulty is because our entrance to the Bureau structure is in the middle of a hierarchy which has considerable division of labor by levels….Specifically, we must gain greater control of the process that established time deadlines for detailed parts of the Bureau’s work on this project. Obviously it is a delicate matter to “intervene” in internal matters within the Bureau….We hope that this “Skinnerian” mode of management will move things along at a more rapid pace….If

[1] Much to the consternation of Winsborough, NSF later reduced the budget to $6,724,550, which created concern that the project could not be finished. In the end, all data preparation was completed, but the planned documentation was not. (“40/50 Project Budget Summary for Total Project Period of 4 Years,” May 1981). University of Wisconsin-Madison Archives 1985, p. 7.

these efforts have not proved effective by early spring, I believe it will be necessary to consult … with the Advisory Committee about next steps, for I am about at wits end" (Winsborough to Aborn, December 27, 1979; University of Wisconsin-Madison Archives, p. 8).

The data capture plan called for the purchase of a dedicated state-of-the-art computer, a Digital Equipment Corporation VAX 11/780, at a cost of about $250,000 (about $1.2 million in 2026 dollars). Necessary peripheral devices such as disk drives, tape drives, terminals, and printers added substantial additional expense. The plan was to have the University of Wisconsin purchase the computer and loan it to the Census Bureau for the duration of data entry.

Once data entry was completed, the machine would be shipped to Madison to replace an obsolete IBM machine. Referring to the plan to install the computer at the Census Bureau facility in Kansas, use it for data entry, and then ship it to Wisconsin where it was used for general social science research, Winsborough remarked that it was "kind of a cuckoo's egg, you know" (University of Wisconsin-Madison Archives, p. 3).

The acquisition of the VAX computer proved to be difficult. Pricing, approvals, delivery, set up, training, maintenance, and use outputs all had to be navigated across multiple staffs (MIN-46 - MIN-57). In April 1979: "Hal announced that we have informed Digital Equipment Corporation of our intent to purchase a VAX; have prices in hand for the main equipment (which will run $210,000 - 250,000 depending on what we settle on); and are now securing prices on peripherals and telephone line rental to Washington. There will be a fee of $1,100/month for maintenance contact from Kansas City to Pittsburg. Delivery date should be 90 days from date of ordering equipment" (MIN-46). Four months later the purchase of the VAX was still in process: Winsborough "summarized the current status of the arduous task of acquiring the project computer. Paperwork is still going through the state bureaucracy with the current projection of securing the governor's signature in about 15 days" (MIN-63). In October 1979, the project minutes were discouraging: "The project is now about 8 months behind schedule. In order to finish within the limits of the grant we must proceed at a vigorous time schedule. The [Advisory] Committee is very concerned with the timetable" (MIN-71).

The VAX was finally installed at Pittsburg and operational in February 1980 (MIN-84 - MIN-86) but communication continued to vex staff at the three locations: "The question of communication between Pittsburg and Madison, via Suitland has proven to be complicated. It may be that the problem of controlling the amount of data sent to the buffer will only be resolved by additional equipment. Since we can be delayed by nearly a week each time we must send things by mail, so a computer link would be a great time saver" (MIN-87).

Once the equipment and software were in place, the transcription process went rapidly. At the peak of production, fifty-one clerks worked during the day shift and thirty-seven clerks at night (U.S. Census Bureau 1984, p. xiv; Status Report No. 14, 1981). Each census year was divided into twenty subsamples consisting of randomly selected microfilm rolls. Thus, each subsample represented a clustered but representative sample of the nation as a whole.

Work began on the 1940 census, and the first sixteen subsamples were completed between October 1980 and July 1981. At that point processing of 1940 was paused, because the investigators were concerned that there would not be enough time to complete 1950 (U.S. Census Bureau 1984, p. xiv). The 1950 census was processed between July 1981 and May 1982. The final four subsamples of the 1940 census were then processed, and data capture was completed by July 1982. Once the data capture was complete, the VAX was shipped to Madison where editing, allocation, and coding procedures were already underway (University of Wisconsin-Madison Archives 1985, p. 4). The 1940 sample was completed in October 1983 and the 1950 sample in March 1984.

1900 Public Use Sample (1976-1980)

The 1900 public use sample was conducted at the same time as the 1940/1950 project, but it was conceived completely independently. Samuel H. Preston, the Principal Investigator, was entirely unaware of the 1940/1950 efforts. The catalyst for the 1900 project was the availability of new source materials. In late 1973, the National Archives and Records Administration opened access to microfilm of the 1900 census enumeration forms to qualified researchers. The motivation for this development is debated, but Bogue (1979) noted that “the Speaker of the House of Representatives at the time was an enthusiastic genealogist.”

One of Preston’s graduate students was searching for a dissertation topic and discovered that the 1900 population census manuscripts had recently been made available to qualified researchers. The student’s review of the enumeration forms revealed two important questions asked of every woman: “mother of how many children” and “number of these children living.” Preston recalled, “He told me that these questions had been asked and I went nuts. I couldn’t believe it. They had never been tabulated, nor had the Census Bureau issued the results. And here was a chance to study child mortality at a very pivotal time in U.S. history.” Preston determined that the best way to leverage the availability of these data was to create a public use sample of the census records (Ruggles and Magnuson 2025).

In September 1976, Preston secured funding from the National Science Foundation to create a national 1 in 1000 sample of the population of the United States in 1900. Compared with the 1940/1950 project, the 1900 sample was conducted on a shoestring: the total award was just $128,500, including the indirect costs (National Science Foundation 1976).

Like the 1940/1950 project, the 1900 sample used direct data entry. According to Preston, several individuals were key to the success of the project (Ruggles and Magnuson 2025). R. Christian Johnson had been “interested in the mechanics of creating a [1900] public use sample and he thought it would be desirable to do direct data entry from the microfilm ... That turned out to be a really good decision. Avoiding the potential errors that you get when you put something on paper and then translate that to a computer file.” Johnson “originated many of the methods used in collecting and coding the data, and contributed his administrative skills” (Graham 1980).

Stephen Graham provided statistical expertise on “every aspect of the creation of the sample” and “designed and programmed the computer systems and supervised data collection and coding” (Graham 1980). Leslie Kish, a survey methodologist at the University of Michigan consulted on sampling methodology. Additional expertise was drawn from Avery Guest (geographic and urban coding), James McCann (occupations), and Sheila Johannsen (household structure).

The 1900 Public Use Sample created by Preston’s team (and Robert Higgs after Preston left the University of Washington) began in the fall of 1976. The first order of business was addressing the technical details of the project and creating the workflow. The 1900 census records were released with restrictions: only approved historical, legal, or genealogical researchers could use the microfilm. Moreover, all research had to be conducted at a National Archives facility, and reproductions were prohibited (Kratz 2022). That meant that data entry had to be conducted at a National Archives facility. Fortunately, a branch of the National Archives was located in Seattle, Washington.

Working out data entry, design, programming, testing, and developing a sampling scheme consumed the first few months of the project (Graham, pp. 2-3). The data entry phase of the project began “in earnest” in the spring of 1977.

It was not practical in 1976 to use a computer in the National Archives facility since microcomputers were not yet widely available. The workflow for the data entry therefore used a Hewlett-Packard 2645A terminal equipped with two small cassette tape drives (Figure 1). The data entry operator keyed data from the microfilm to the tape, using the terminal’s Cathode Ray Tube screen to sight check for accuracy and then record the entry on the cassette (Graham 1980). Thus, Johnson’s idea for direct data entry from the census manuscript data on microfilm to a digital format bypassed the manual coding step that had been the norm with data collection projects.

Figure 1. Hewlett Packard 2545A



Source: https://terminals-wiki.org/wiki/index.php/HP_2645A

The workflow for the data entry phase of the project continued through the summer of 1978 and consisted of a weekly schedule of transmitting the tape cassettes over phone lines to a central computer at the University Washington, editing and cleaning at the University of Washington by project staff, edits returned to the data entry staff at the archive branch, manual lookup of edits

by the data entry staff, corrections made, corrections forwarded to the central computer where they were accumulated on tape (Graham 1980). At each of these workflow points project staff were cognizant of the need to control for errors.

After the data entry phase of the project was completed the coding stage began and continued through late October 1979 (Graham 1980). Codes were assigned for geographic, household, and occupational structure. Coding was a labor-intensive manual task conducted in part by researcher assistants “examining cases and deciding which codes to apply” and computers “in some cases” to organize the data and apply consistency checking algorithms.

Near the end of 1979 the first version of the 1900 public use sample was released to members of the project for their preliminary analysis and feedback. In July of 1980 documentation was completed and the second version of the 1900 public use sample was released to the public that summer.

The original plan for the 1900 sample called for 76,000 person records, representing 1/1000 of the U.S. population. Data capture was more efficient than anticipated, however, and the sample was expanded to 1/760th of the U.S. population in 1900, made up of 27,069 households and 100,438 person records (Graham, p. 1).

1910 Public Use Sample (1983-1989)

By 1979 Samuel Preston had moved to the University of Pennsylvania and was deliberating how he could use the 1910 census for historical child mortality research (Ruggles and Magnuson 2025). Under the new “72-year rule,” the microfilm of the 1910 enumeration forms was to be released on April 1, 1982 (Kratz 2022). Like 1900, the 1910 population census schedule also asked questions of married women regarding their children: “mother of how many children,” “number born,” and “number now living.” Drawing on Preston’s experience creating the 1900 public use sample, he successfully applied for and received federal funding to create a 1910 public use sample (Strong et al. 1989). The project received $511,065 from the National Institute of Child Health and Human Development for a 1-in-500 sample and supplemental funding of $118,000 from the National Science Foundation to expand the sample to 1-in-250 (Preston 1984, pp. 5-6).

The 1910 project had two great advantages over the 1900 project. First, in 1977 the National Archives relaxed the rules for access to census enumeration forms and began selling copies of the microfilm (Kratz 2022). Accordingly, the 1910 grant proposal budgeted for direct purchase of census microfilm reels, thus greatly simplifying the logistics of data entry operator access to the source data (Preston 1982). Second, the availability of microcomputers dramatically simplified data capture. Gone was the dumb terminal and phone line set up and in its place three IBM personal computers with 5¼ inch floppy disks, along with an external 10 MB hard disk, a dot-matrix printer, and a 1200 Baud modem. Altogether, the equipment was budgeted at $28,020,

or about $100,000 in 2026 dollars (Preston 1982). Table 2 summarizes the data-entry technology used for the Preston and Winsborough historical census projects.

Table 2. Technology used in Winsborough and Preston projects

PUS	Project Years	Data Entry Years	Technology
1900	1976-80	1977-1978	Hewlett-Packard 2645A dumb terminal, cassette tape, terminal, phone line
40/50	1978-84	1980-1982	Direct data entry into central VAX computer, 51+ CRT terminals
1910	1982-89	1982-1987	IBM personal computers, data entry saved to floppy disk

Project staff for the 1910 sample included three computer programmers, two data entry supervisors, three data entry operators, two business administrators, and two secretaries. Academic staff included Preston (as PI), Ann Miller (consultant on occupation/industry coding), Harold Lentzner (census historian), and Elaine Allen (statistical consultant) (Strong 1989, p. 41).

The original 1910 NICHD proposal called for the sample to be completed in 1985. The expansion of the sample funded by NSF was to be released in 1986. In practice, everything took longer and cost more than anticipated. We lack documentation of the specifics of the difficulties. The data were finally released in late 1989, three years after the funding had expired.

Discussion

As described above, in 1970 the Census Advisory Committee on Population Statistics proposed "public use samples tapes for 1940, 1950, and 1960 on a basis comparable with the code structure used in the public use sample for 1970." A new 1960 public use sample was created that was entirely compatible with the 1970 samples, sharing the same record layouts, variable names, and coding schemes for all variables.

For the past 35 years, we have told a simple story about the origins of the Winsborough and Preston historical microdata samples (Ruggles 1991a, 1991b, 2005, 2014; Magnuson and Ruggles 2022). We maintained that after the Census Bureau produced compatible microdata for 1960 and 1970 there was an outpouring of research on social and economic change in that period, and this independently inspired Winsborough and Preston to extend the series backwards in time. We now know that this origin story is incorrect. The calls for microdata

samples of 1940 and 1950 began in the late 1960s, well before the 1970 sample or the harmonized version of the 1960 sample existed. The Preston samples were created to address a particular research problem, and the ability to investigate historical trends was an incidental benefit.

The 1940 and 1950 samples did eventually appear, but they were not produced on “a basis comparable with the code structure used in the public use sample for 1970”; they were completely different. Indeed, even though they were prepared by the same team at the same time, the 1940 and 1950 samples were completely different from one another. So too were the two samples created under the direction of Preston. And when the Census Bureau produced new samples for 1980 and 1990, they also were each completely different from everything that had come before.

By the late 1980s, both Winsborough and Preston were burned out on the creation of historical census microdata samples. As Winsborough commented, “the project made me very tired.” (University of Wisconsin-Madison Archives 1985, Day 1, 1:04). They both provided Steven Ruggles (University of Minnesota) with copies of their grant proposals and encouraged him to fill in the gaps. Collaborating with historical demographer Russell Menard, Ruggles took ideas from both proposals and wrote a proposal to create a 1-in-1000 sample of the 1880 census. For example, the 1880 proposal proposed using subsamples, missing data allocation, and verification procedures modelled on the 1940/50 project, and the data-entry protocol and constructed pointer variables closely followed the 1910 project. With so many strong borrowed ideas, the proposal received nearly a perfect score. The reviewers, however, insisted on one key change: they wanted Ruggles and Menard to raise the sample size from 1-in-500 to 1-in-100 (with a commensurate increase in the budget) (Ruggles and Menard 1985; Ruggles and Menard 1990). Figure 2 summarizes the timeline of historical census microdata projects from 1970-1997.

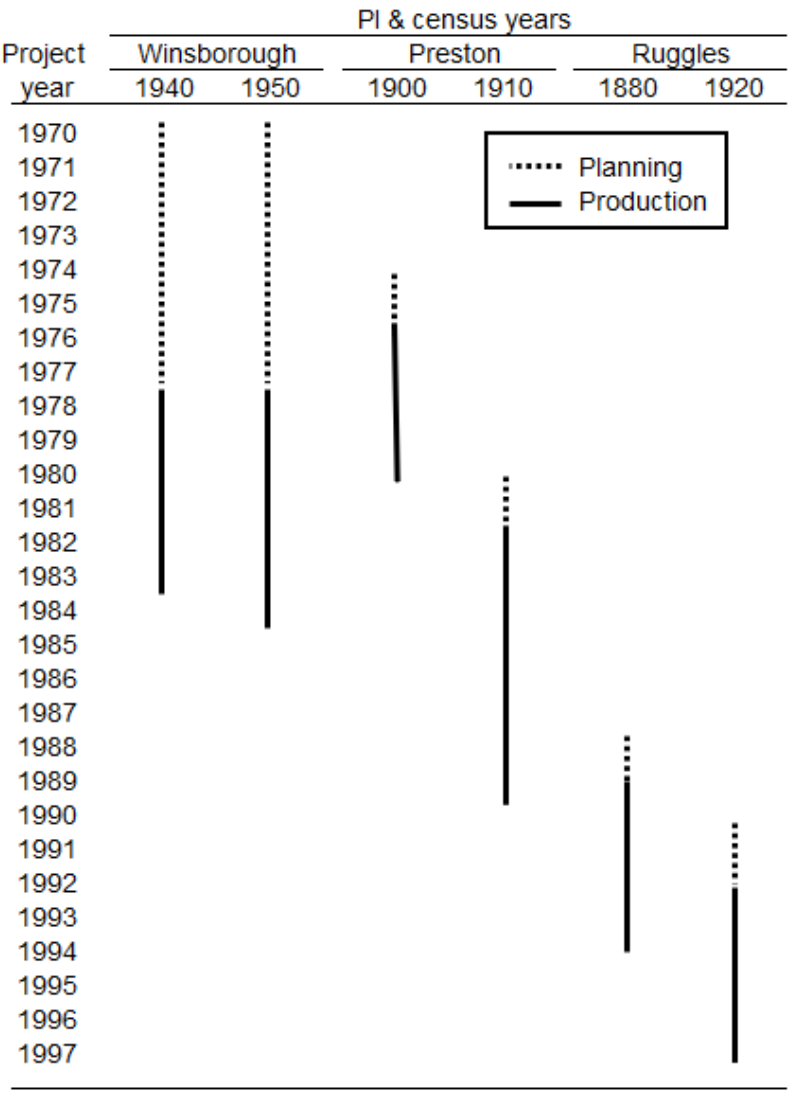
Two years later, Ruggles proposed a similar project to create a 1-in-100 sample of the 1920 census. The project capitalized on the 72-year rule, which promised that the enumeration forms for 1920 would become publicly available in April 1992 (Ruggles and Menard 1995, p. 7).

By 1991, there were samples for 1880, 1900, 1910, 1940, 1950, 1960, 1970, and 1980, and the new samples for 1920 and 1990 were imminent. All these samples were incompatible with one another, with the sole exception of 1960 and 1970. This made the data exceedingly difficult to use as a series. Researchers attempting to analyze change over time had to harmonize the data first, which was cumbersome, error-prone, and expensive. To address these issues, Ruggles proposed to make an integrated version of the datasets with one consistent record layout, common variable names, harmonized coding schemes, a single set of documentation, and no loss of information (Ruggles 1991a).

The harmonized version of the data was called the Integrated Public Use Microdata Series (IPUMS). In subsequent years IPUMS expanded further, and it now comprises thousands of

datasets and billions of records from a wide range of sources, including full-count census microdata for the period from 1790 to 1950.

Figure 2. Timeline of historical census microdata projects, 1970-1997



References

Bogue, Allan G. (1979). “Data Dilemmas: Quantitative Data and the Social Science History Association.” *Social Science History* 3(3/4): 204-226.

Duncan, O. D. (1969). *Toward Social Reporting: Next Steps*. Russell Sage Foundation. https://doi.org/10.7758/9781610448802

Graham, S.N. (1980). *1900 Public Use Sample User's Handbook.* The Center for Studies in Demography and Ecology, University of Washington, Seattle, Washington.

Interagency Committee on Population Research. (1973). *Analysis of Federal Population Research, Fiscal Year 1973.* Washington. https://babel.hathitrust.org/cgi/pt?id=mdp.39015084907982&seq=6

Kratz, Jessie. (2022). "Census Records: The 72-Year Rule." Pieces of History, A Blog of the U.S. National Archives. https://prologue.blogs.archives.gov/2022/01/20/census-records-the-72-year-rule/

Magnuson, D.L. and Ruggles, S. (2022). "Challenges of large-scale data processing in the 1990s: The IPUMS Experience." *IEEE Annals of the History of Computing* 44(4): 71-83. DOI: 10.1109/MAHC.2022.3214736

Magnuson, D.L. and Ruggles, S. (2026). "'Equally Available to Anyone': The Invention of Public Use Microdata at the U.S. Census Bureau." *Demography* 63(4): 1333-1346. https://doi.org/10.1215/00703370-12785170

Mason, W.M., Taeuber, K.E., Winsborough, H.H. (1977). "Old Data for New Research: Report of a Workshop on Research Opportunities and Issues in the Design and Construction of Public Use Samples from the 1940 and 1950 Censuses and from Current Population Surveys from 1960 Forward, Held in Madison, Wisconsin, June 28-30, 1976." CDE Working Paper 77-3. Center for Demography and Ecology, University of Wisconsin.

National Science Foundation. (1976.) Award #7620002. https://www.nsf.gov/awardsearch/show-award/?AWD_ID=7620002

Preston, S.H. (1984). "Curriculum Vita Sept 2024." https://www.pop.upenn.edu/bio/samuel-preston. Accessed 2026-09-16.

Ruggles, S. (1991a). "The U.S. Public Use Microdata Files as a Source for the Study of Long Term Social Change." *IASSIST Quarterly* 15(2): 20-27.

Ruggles, S. (1991b). "Comparability of the Public Use Files of the U.S. Census of Population, 1880-1980." *Social Science History* 15(1): 123-158. DOI: https://doi.org/10.1017/S0145553200021039

Ruggles, S. (2005). "The Minnesota Population Center Data Integration Projects: Challenges of harmonizing census microdata across time and place" *2005 Proceedings of the American Statistical Association, Government Statistics Section*, Alexandria, VA: American Statistical Association, pp. 1405-1415.

Ruggles, S. (2014). " Big Microdata for Population Research." *Demography*. 51(1): 287-297. https://doi.org/10.1007/s13524-013-0240-2

Ruggles, S. and Magnuson, D.L. (2025). Samuel Preston Interview, University of Minnesota, December 16, 2025.

Ruggles, S. and Magnuson, D.L. (2020). “Census Technology, Politics, and Institutional Change, 1790-2020,” *Journal of American History* 107(10): 19-51. https://doi.org/10.1093/jahist/jaaa007

Ruggles, S. and Menard, R.R. (1990). “A Public Use Sample of the 1880 U.S. Census of Population.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 23(3): 104-115.

Ruggles, S. and Menard, R.R. (1995). “The Minnesota Historical Census Projects.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 26(1): 6-10. https://doi-org.ezp3.lib.umn.edu/10.1080/01615440.1995.9955308

Strong, M.A., Preston, S.H., Miller, A.R., Hereward, M., Lentzner, H.R., Seaman, J.R., and Williams, H.C (1989). *User’s Guide Public Use Sample 1910 United States Census of Population*. Population Studies Center, University of Pennsylvania, Philadelphia, Pennsylvania.

University of Wisconsin-Madison Archives. (1985). Halliman H. Winsborough Interview #301. Oral History Project.

U.S. Census Bureau. (1963). *U.S. Censuses of population and housing: 1960. 1/1,000, 1/10,000: Two national samples of the population of the United States* (Description and technical documentation). Washington, DC: U.S. Department of Commerce, Bureau of the Census.

U.S. Census Bureau. (1984a). Census of Population, 1940 [United States]: Public Use Microdata Sample (ICPSR 8236).

U.S. Census Bureau. (1984b). Census of Population, 1950: Public Use Microdata Sample Technical Documentation. Washington, D.C.

IPUMS Archive, University of Minnesota

Preston, S.H. (1982). NICHD Grant Application, 1982, February 17. IPUMS Archive, University of Minnesota.

“40/50 Project Budget Summary for Total Project Period of 4 Years,” May 1981. Box 2, Folder 1, Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

MIN-8," Box 1, Folder 3. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

MIN-46 - MIN-57, Box 1, Folder 3. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

MIN-63, Box 1, Folder 3. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

MIN-71, Box 1, Folder 3. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

MIN-84 - MIN-86, Box 1, Folder 3. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

MIN-87, Box 1, Folder 3. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

Status Report No. 14. (1981, January 12). Box 2, Folder 1. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

Status Reports No. 3-19. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

Winsborough to Aborn, December 27, 1979 (correspondence), Box 2, Folder 1. Halliman H. Winsborough Collection, IPUMS Archive, University of Minnesota.

National Archives and Records Administration

Census Advisory Committee on Population Statistics. (1970, February 20). "Summary of Fourteenth Meeting–February 20, 1970. NARA, RG 29, Box 3, Census Advisory Committee on Population Statistics:1948-1980, CAC 6th Meeting 12/13/1966 to CAC Meeting 09/21/1975.

Census Advisory Committee on Population Statistics. (1971, March 12). "Summary of Sixteenth Meeting–March 12, 1971. NARA, RG 29, Box 3, Census Advisory Committee on Population Statistics:1948-1980, CAC 6th Meeting 12/13/1966 to CAC Meeting 09/21/1975.

Turner, M.L. (1970). Memorandum from Marshall L. Turner to Herman P. Miller, 1940 and 1950 Public Use Samples, February 13, 1970. NARA, RG 29, Box 10, General Correspondence Relating to the Development and Planning of the 1970 Census, 1961-1969. P88 NN3-29-81-54.

U.S. Census Bureau. (1970, February 20). Census Advisory Committee on Population Statistics, Summary of Fourteenth Meeting–February 20, 1970. NARA, RG 29, Box 3, Census Advisory Committee on Population Statistics:1948-1980, CAC 6th Meeting 12/13/1966 to CAC Meeting 09/21/1977.

U.S. Census Bureau. (1970, September 17). 1940 and 1950 Public Use Record Samples, PA-77. NARA, RG 29, Box 3, Census Advisory Committee on Population Statistics:1948-1980, CAC 6th Meeting 12/13/1966 to CAC Meeting 09/21/1977.

U.S. Census Bureau. (1970, October 23). Census Advisory Committee on Population Statistics, Summary of Fifteenth Meeting–October 23, 1970, PA-80. NARA, RG 29, Box 3, Census Advisory Committee on Population Statistics:1948-1980, CAC 6th Meeting 12/13/1966 to CAC Meeting 09/21/1977.