

IPUMS

Working Papers

“Equally Available to Anyone”: Creating the First Census Microdata Sample at the U.S. Census Bureau

Diana L. Magnuson†
University of Minnesota

Steven Ruggles
University of Minnesota

December 2024

Working Paper No. 2024-07

DOI: <https://doi.org/10.18128/IPUMS2024-07>

†Address correspondence to Diana Magnuson: magn0031@umn.edu. Diana L. Magnuson is Curator and Historian at the Institute for Social Research and Data Innovation, University of Minnesota. Steven Ruggles is Regents Professor of History and Director of IPUMS, University of Minnesota.

**“Equally Available to Anyone”:
Creating the First Census Microdata Sample at the U.S. Census Bureau**

Diana L. Magnuson and Steven Ruggles, University of Minnesota

Abstract

The availability of microdata for social science research has grown exponentially since its creation in the 1960s. Powered by parallel technological advances over the last sixty years, creation and manipulation of microdata for transformative historical and policy research has also exploded. This paper explores the creation of the first census microdata sample at the U.S. Census Bureau in the early 1960s. The role of the social science community outside the Bureau, and in particular prominent members of the Population Association of America, played a significant part in encouraging the development of microdata. Drawing on historical evidence from the National Archives—including Census Bureau internal memoranda, technical advisory committee minutes and correspondence, academic association minutes, as well as a variety of secondary and primary source materials gathered from public and private collections—our research aims to uncover the history of the people and social science behind the development of the U.S. Census microdata.

The Invention of Microdata

Until the mid-twentieth century, population statistics were available only in aggregated form, disseminated in printed books of statistical tables. These data described the characteristics of administrative units (such as states, counties, or municipalities) or statistical units defined by a statistical agency (such as census tracts). In 1962, the U.S. Census Bureau released the first microdata file. Instead of describing characteristics of *places*, the new data format described the characteristics of *people*. Moreover, instead of printed books, the microdata file was published in a machine-readable format, designed for further processing by computers or tabulating machines.

The first census microdata was created as a byproduct of the 1960 U.S. census.¹ The expansion of social science research in the late 1950s and early 1960s led to growing demand for special tabulations designed to answer specific research questions. The Census Bureau prepared these special tables on a reimbursable basis, but the costs were too high for the typical university researcher.²

To meet the demand for customized tabulations, the Census Bureau drew a 1-in-1000 extract of the basic data tapes they had used to create tabulations for the published census volumes. To preserve confidentiality, the Census Bureau removed names, addresses, and other potentially identifying information. In 1962, the Census Bureau released the data on 7 IBM tapes or 11 Univac tapes for a charge of \$1,500.³ Very few research centers had access to computers, so the Census

¹ U.S. Census Bureau, *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* (U.S. Department of Commerce, Bureau of the Census, undated).

² Solomon Fabricant, "Report of the Census Advisory Committee," *American Economic Review*, vol. 55(1/2) (March 1, 1965): 619-620. Rebecca S. Kraus, "Statistical Déjà Vu: The National Data Center Proposal of 1965 and Its Descendants," *Journal of Privacy and Confidentiality*, vol. 5, no. 1: 1-37, <https://doi.org/10.29012/jpc.v5i1.624>.

³ H.G. Brusman, Letter to Dr. Joshua Lederberg, April 22, 1963. Joshua Lederberg Papers, National Library of Medicine. <http://resource.nlm.gov/101584906X18048>. J.W. Duncan and W.C. Shelton, *Revolution in the United States Government Statistics: 1926-1976*. U.S. Department of Commerce Office of Statistical Policy and

Bureau also made a 1-in-10,000 version available on 18,000 punch cards that investigators could analyze using older unit record machines.⁴

The Power of Microdata

The 1960 public use sample revolutionized analysis of the American population and led to an outpouring of new census-based research. Before 1960, aggregated data was “destroy[ing] useful information and hindering scholarly potential in the emerging computer age.”⁵ Sociologist and statistician Otis Dudley Duncan put it,

The importance of this innovation can hardly be overestimated. We have known for a long time that certain essential social indicators are available in principle from the Federal statistical system. Yet all too often efforts to put information into an appropriate form are frustrated by the inadequacy of the published summary tables for the purpose at hand. With access to the unit records, the social scientist may specify in detail how variables are to be manipulated so as to produce an optimal estimate of the magnitude desired.⁶

Researchers William M. Mason (demographer), Karl E. Taeuber (sociologist), and Halliman H. Winsborough (sociologist) concurred, explaining that the 1960 census microdata sample was a “development of profound significance to social research” because it gave the research community “freedom to retabulate or manipulate without the constraints imposed by a fixed set of printed volumes.”⁷ The new sample allowed researchers to make tabulations tailored

Standards (Washington, D.C.: GPO, 1978). P.M. Hauser, “The 1960 census as an instrument for demographic research,” *Population Index*, vol.26, no.3 (July 1960): 201-202.

⁴ W.M. Mason, K.E. Taeuber, and H. Winsborough. “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.” Madison, WI, June 28-30, 1976. Center for Demography and Ecology Working Paper 77-3. U.S. Census Bureau, *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* (U.S. Department of Commerce, Bureau of the Census, undated).

⁵ Richard Ruggles and Nancy D. Ruggles, “The Development of Public Use Samples of Longitudinal Data for U.S. Manufacturing Establishments, 1966-1978” (Yale University, January 1980): 2.

⁶ O.D. Duncan, “Developing social indicators,” *Proceedings of the National Academy of Sciences* (1978) 71:5096-5102 [citation on p. 5097].

⁷ W.M. Mason, K.E. Taeuber, and H. Winsborough. “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.” Madison, WI, June 28-30, 1976. Center for Demography and Ecology Working Paper 77-3.

to their specific research questions. For example, suppose one wants to study the relationship of teen marriage to high school drop-out rates. The U.S. census never published a table on marital status by school enrollment by age, so that topic cannot be investigated using the aggregated census data. With microdata, however, researchers can analyze any combination of characteristics.

The new microdata format enabled construction of customized variables based on multiple records. The 1960 census file was organized into a hierarchical format: individuals are nested into families, and families are nested in households. Within families, the relationships between individuals are known. This structure makes it easy to develop new measures that combine characteristics of more than one person. For example, by comparing a husband's ethnicity with that of his wife, one can measure ethnic intermarriage. Investigators can develop a virtually limitless number of customized measures that combine characteristics from multiple persons, such as number of own children attending school, age differences of siblings, or co-residence of kin.

Social scientists quickly realized that the new data format enabled new methods, especially individual-level multivariate analysis.⁸ In a statement before the Subcommittee on Economic Statistics of the Joint Committee, economist Richard Ruggles put it this way: "The ability of social scientists to obtain highly disaggregated data permits them to use techniques of analysis which are inherently much more powerful and can separate out the structural changes of the system from the changes in behavior of individual units."⁹ Thus investigators could move beyond the simple three-way or four-way tables provided by aggregate census data and control for many characteristics simultaneously using regression-based techniques. Beyond individual-level analysis, microdata

⁸ J.B. Neuroth, "The National Data Center proposals between macro modelling and micro targeting," *Cogent Arts & Humanities* (2023) 10:2286077, <https://doi.org/10.1080/23311983.2023.2286077>.

⁹ "The Coordination and Integration of Government Statistical Programs," Hearings Before the Subcommittee on Economic Statistics of the Joint Economic Committee, Nineteenth Congress, First Session of the United States, Statement of Richard Ruggles, Yale University, before the Subcommittee on Economic Statistics of the Joint Economic Committee, May 17, 1967.

enabled multi-level analyses controlling for individual, family, and community level-variables simultaneously.

How did this revolutionary innovation come about? This paper explores the people and social forces behind the first microdata sample. While we cannot pinpoint one individual as responsible for census microdata, we can identify the principal actors and general issues surrounding its creation.

Two broad themes emerge from our research. First, the creation and development of machine-readable U.S. census microdata occurred in a dynamic technological context of invention, expansion, and dissemination. The pressures of post-war American economic expansion on the federal government and its citizens demanded access to data to analyze the present and future impact of these changes.¹⁰ Second, the relationship between Census Bureau leadership and users of census data in academic disciplines was dynamic, mutually supportive, and actively nurtured. Census Advisory committees were formed within several major (census and survey data consuming) academic organizations, for the purpose of coordinating communication around census issues. In fact, the idea of microdata “apparently was first suggested at a Census Advisory Committee meeting.”¹¹

Historical Context

The history of microdata is best understood within the context of public access to tabulated U.S. census data. For over 150 years, tabular census data was disseminated to the public in printed publications.¹² This seemingly tidy distributive form belied a complex history.

¹⁰ Margo J. Anderson, “Reapportionment, Funds Allocations, and the Census,” chapter in *The American Census: A Social History* (New Haven: Yale University Press, 2015): 209-223.

¹¹ J.W. Duncan and W.C. Shelton, *Revolution in the United States Government Statistics: 1926-1976*. U.S. Department of Commerce Office of Statistical Policy and Standards (Washington, D.C.: GPO, 1978): 142.

¹² Janice S. Fryer, revised by Leonard M. Gaines, “Dissemination of data: printed publications,” *Encyclopedia of the U.S. Census*, Margo J. Anderson, editor (CQ Press, Washington, D.C. 2012): 188-191.

Between 1790 and 1840, the census form was organized around gathering data at the household level. Correspondingly, tabulation of each decennial census through 1840 was decentralized and carried out in the field by assistant U.S. marshals who forwarded their grand totals to their respective U.S. marshals. 1850 marked a significant watershed in enumerating the American people both in form and substance. Six separate schedules were used: 1) free inhabitants; 2) slave inhabitants; 3) mortality; 4) agriculture; 5) industry and 6) social statistics. Schedule No. 1 was focused on the individual, with each person in the household listed on a separate line and information collected on eleven demographic questions and two household questions.

The reformatting and expansion of the 1850 census necessitated a redistribution of tabulation work. Assistant marshals were now instructed to transfer their completed raw schedules to the Census Office in Washington where temporary clerks conducted the tabulation work. Almost immediately, the Census Office was in a data capture crisis created by a tabulation bottleneck that continued to worsen exponentially with the succeeding two decennial censuses.¹³ The growth of the free population through the abolition of slavery and immigration, the increase in the number of questions on the population schedule, and the demand by the public for published statistics, all put mounting pressure on the Census Office to find a solution to the tabulation problem.

Ahead of the 1890 decennial census, Superintendent of the Census Robert P. Porter issued a call for ideas to address the tabulation crisis. Several inventive ideas were submitted, and the winner was census employee Herman Hollerith and his electrical punch card tabulation system. In consequence, from 1890 through 1950, censuses were processed by punched cards and a series of

¹³ Steven Ruggles and Diana L. Magnuson, "Census Technology, Politics, and Institutional Change, 1790-2020," *The Journal of American History*, vol. 107, issue 1 (June 2020): 19-51, <https://doi.org/10.1093/jahist/jaaa007>.

increasingly modernized tabulating machines. In-house invention and innovation characterized the productive work of the Machine Shop (later Census Mechanical Laboratory).¹⁴ Despite these gains, the modern punch card too, would ultimately produce its own set of bottlenecks.

The demands of data capture and processing in the 1940s and 1950s fueled significant technological advances at the Census Bureau during this period. Nascent electronic computers seemed to offer potential solutions to Bureau problems. The Census Bureau funded commercial computer development to address their data processing needs. Key innovations included the use of magnetic tape for data storage instead of punch cards and microfilming paper enumeration forms. Concurrently, the Bureau collaborated with the National Bureau of Standards to design and build an optical sensing system “that used a photoelectric cell to read marks directly onto magnetic tape.”¹⁵ The Film Optical Sensing Device for Input to Computers (FOSDIC) was the result. The stage was set for a transformation in the dissemination possibilities for tabular census data.

Across this period from 1790-1950, the temporary Census Office and later the permanent Census Bureau (beginning 1902), struggled to keep pace not only with the constitutional demands of the tabulation but also the constituent user appetite for tabulated census data. One strategy for addressing these needs was to actively cultivate relationships between the Census Bureau in Washington and users of the data in academia, federal agencies, and organizations.

Beginning with the 1840 census, statistical experts were consulted by census administrators during the developmental stage leading up to the decennial census. The advice of these experts

¹⁴ Steven Ruggles and Diana L. Magnuson, “Census Technology, Politics, and Institutional Change, 1790-2020,” *The Journal of American History*, vol. 107, issue 1 (June 2020): 19-51, <https://doi.org/10.1093/jahist/jaaa007>.

¹⁵ Steven Ruggles and Diana L. Magnuson, “Census Technology, Politics, and Institutional Change, 1790-2020,” *The Journal of American History*, vol. 107, issue 1 (June 2020): 19-51, <https://doi.org/10.1093/jahist/jaaa007>.

A. Ross Eckler, *The Bureau of the Census* (Praeger Publishers, New York, 1972).

was heeded to varying degrees across the nineteenth and into the twentieth centuries, but the connection between the census-invested community and census officials in Washington was firmly established. In 1919, the American Statistical Association created a standing Census Advisory Committee with the purpose of “advising the Directors of the Bureau on program on policies.”¹⁶ By the 1960s, the broader social science community was playing an active consulting role as census of population and housing products were being developed. The expertise of a broad swath of social scientists was solicited and acted upon by the Census Bureau. Ahead of the 1960 census, the Bureau relied upon newly formed advisory groups as well as long-standing permanent advisory committees for “advice and assistance” in the development of the population and housing census.¹⁷

The Role of the Social Science Community

The social science community played a significant role in promoting the development of publicly accessible census data products in the twentieth century. The second half of the twentieth century demanded more, and more flexible, data to address some of the most pressing social and economic concerns of the day.¹⁸ Social scientists along with the local, state, and federal governing entities were trying to understand the parameters of these issues and offer attainable solutions. Annual academic conferences like the Population Association of American (PAA) and the American Statistical Association (ASA) were natural hubs of information and idea exchanges between Bureau staff and academic researchers. Many Bureau staff came up through graduate programs in the social sciences, and many had ties to universities throughout their careers.

¹⁶ U.S. Bureau of the Census, *The 1950 Censuses – How They Were Taken* (Washington, D.C. 1955): 5. Diana L. Magnuson, “The Making of a Modern Census: The United States Census of Population, 1790-1940” (University of Minnesota, PhD Dissertation, 1995): 74-78.

¹⁷ U.S. Bureau of the Census, *1960 Censuses of Population and Housing: Procedural History* (Washington, D.C. 1966), Appendix C: Principal Conferences and Meetings on the 1960 Census Program, pp. 295-300.

¹⁸ Conrad Taeuber, “Providing Relevant Data,” *The American Sociologist*, vol. 6, Supplementary Issue (June 1971): 62-65.

At the May 1956 annual meeting of the Population Association of America, held in Ann Arbor at the Department of Sociology, University of Michigan, seven sessions of papers and discussions were held, in addition to other annual meeting business. The subjects of the sessions, paper abstracts, and “Discussants’ Remarks” were recorded in *Population Index*. Among the sixteen quoted discussant remarks at the May 1956 meeting was this suggestive comment: “The results of the 1960 Census would be made more useful to researchers and other consumers if a method were devised for obtaining duplicate cards for particular population groups or areas. The researcher could then use his own tabulating equipment and could ‘explore’ his data at minimum cost.”¹⁹ While the discussant is unnamed, they expressed an emerging concept among the social science community: access to population data in a new form.

The “Report of the Committee on the 1960 Census” was published in the October 1957 issue of *Population Index*. The committee was appointed by Joseph J. Spengler, president of the Population Association of America in May 1956, “pursuant to a letter of 23 May from Robert W. Burgess, Director, Bureau of the Census.” Burgess’ letter communicated that the “Bureau of the Census would welcome the advice of the Association and its members on such matters as subjects to be included in the next Population Census as well as the definitions of such items and any others on which you may wish to take initiative.”²⁰ The committee created working groups for particular subject areas and the report indicated that there was strong cooperation between the Bureau and the committee, and “particularly valuable” was the appointment of Bureau staff to serve as liaisons with each working group.²¹ The report made clear that:

¹⁹ “The 1956 Meeting of the Population Association,” *Population Index*, vol. 22, no. 3 (July 1956): 181.

²⁰ Otis Dudley Duncan, “Report of the Committee on the 1960 Census Population Association of America,” *Population Index*, vol. 23, no. 4 (October 1957): 293-305.

²¹ Census Bureau staff serving as liaisons included Howard G. Brunsman, Paul C. Glick, Henry D. Sheldon, Henry S. Shryock, Jr., and Conrad Taeuber. Otis Dudley Duncan, “Report of the Committee on the 1960 Census Population Association of America,” *Population Index*, vol. 23, no. 4 (October 1957): 293.

The principal criterion observed by the several work groups and by the Committee was the use of census data for scientific research in demography and closely related disciplines. While mindful of the many uses that must be served by the census, these groups considered that their recommendations would be most valuable if presented from this point of view.²²

The recommendations of the working groups focused on enumeration procedures, additional or refined census questions, and post-enumeration tabulation issues.

In December 1957 a census monograph was reviewed by economist Sherman J. Maisel in *American Economic Review*, and Maisel discerned that:

...this work may play a still more significant role if it leads the Census to re-examine the present inadequate methods of making data available for research purposes. The author is forced to apologize on the average of every five or six pages because the census data cannot really be used to test adequately the theories presented.

Within the monograph the author wrote that, “an attack on the problem via census data can be made only with the aid of bold assumptions and catch-as-catch-can techniques.” The Census Bureau only published data in “highly aggregated groups of families.” The review was distributed internally at the Census Bureau to “the Executive Staff” by A. Ross Eckler, Assistant Director of the Census. Assistant Director for Demographic Fields, Conrad Taeuber, scrawled across the bottom of the memo to A. Ross Eckler: “A.R.E. But there is a better way than Maisel’s proposal – made elsewhere – that we give him punch cards.”²³ Bureau leadership and the broader social science community recognized the potential power of census data to understand and analyze contemporary social and economic problems.

In her 1959 presidential address, PAA president Dorothy Swaine Thomas expressed concern that the 1960 Census—which for the first time stored individual-level data on magnetic

²² Otis Dudley Duncan, “Report of the Committee on the 1960 Census Population Association of America,” *Population Index*, vol. 23, no. 4 (October 1957): 293.

²³ Sherman J. Maisel, *American Housing and Its Use* by Louis Winnick, review in *American Economic Review*, vol. 47, issue 6 (December 1957): 1076-1077. Memorandum from A. Ross Eckler to Executive Staff, December 26, 1957, NARA, RG29, Office of the Director, Box 4 MLR A1 389F NN3-029-99-019.

tape rather than on punch cards—might impede demographic research. She reasoned that storage on tape was so expensive that the individual-level data would be quickly erased, and it would then be impossible for demographers to obtain special tabulations from the Census Bureau. Accordingly, she pleaded with the Census Bureau to preserve the microdata in machine-readable form for later analysis, regardless of expense.²⁴ At the annual ASA meetings, seven presidents intentionally drew attention to the dynamic relationship between statistics, data, and modern national problems in their presidential addresses at the annual ASA meetings in the decade of the 1960s.²⁵ Clearly there was demand for public access to machine-readable data among the social science community.

The summary of the 1960 Meeting of the Population Association of America teased “there will probably be available on punch cards or magnetic tape a national sample of households and person, from which identifying information will have been deleted. This sample tabulation may be purchasable from the Census at nominal cost for local use. A larger sample, perhaps a million

²⁴ “The 1959 Meeting of the Population Association,” *Population Index*, vol. 25, no. 3 (July 1959): 191-211. More from Thomas’ address: “In conclusion, it was suggested that the principles enunciated by Francis A. Walker in connection with the especially difficult 9th census of 1870 might be applicable to the comparably difficult 18th census of 1960, namely, that there should be ‘no least hesitation in undertaking any compilation, no matter how extended, which promised results that could be useful to any considerable class of the community or which had a *clear scientific value* (italics supplied)’; and that “especially since it costs so heavily to bring this material into the Census Office, would it be the falsest of all false economy to lose any portion of it which, when tested, is found to be trustworthy, for the sake of effecting a saving in the cost of tabulation”; and, finally, that the additional cost through increase in scope must be justified “in the interest of science, industry, and good legislation.” (p.211)

²⁵ Walter E. Hoadley, Jr., “Statisticians—Today and Tomorrow,” *Journal of the American Statistical Association*, vol. 54, no. 285 (March 1959):1-11. Rensis Likert, “The Dual Function of Statistics,” *Journal of the American Statistical Association*, vol. 55, no. 280 (March 1960): 1-7. Morris H. Hansen, “Cooperation Among Statistical and Other Societies,” *Journal of the American Statistical Association*, vol. 56, no. 293 (March 1961):1-10. Martin R. Gainsburgh, “Statistics We Live By,” *Journal of the American Statistical Association*, vol. 57, no. 297 (March 1962): 1-9. Philip Hauser, “Statistics and Society,” *Journal of the American Statistical Association*, vol. 58, no. 301 (March 1963): 1-12. Raymond T. Bowman, “The American Statistical Association and Federal Statistics,” *Journal of the American Statistical Association*, vol. 59, no. 305 (March 1964): 1-17. Frederick F. Stephan, “The Quality of Statistical Information and Statistical Inference in a Rapidly Changing World,” *Journal of the American Statistical Association*, vol. 62, no. 317 (March 1967): 1-9. A. Ross Eckler, “Statisticians and Shoemakers (“Who is Worse Shod Than the Shoemaker’s Wife,” from Heywood’s Proverbs),” *Journal of the American Statistical Association*, vol. 65, no. 329 (March 1970): 9-21.

households, will, it is hoped, be drawn for special tabulations by the Census for research purposes.”²⁶ Researchers only had to wait two more years for this promise to become a reality.

Moving to Microdata

In October 1959 Conrad Taeuber distributed an internal Bureau memo, “Making available census reports for individual persons for research purposes” in which he communicated two recommendations from a subgroup of the Population Association of America. Among those recommendations:

1. It is proposed that we draw a simple random sample of moderate size (perhaps 10-20,000 households) from all households and publish the individual returns, without specific geographic identification by place, county, or State. Publication might be in the form of tape, cards, or a printout from which the appropriate medium for tabulation might be prepared. The geographic code would show size of place, whether or not in an SMSA, and possibly broad region.

It is essential to this proposal that these materials would be equally available to anyone, and that any purchaser would be free to use them or subsamples of them as he saw fit. A standard credit to the Bureau and a disclaimer by the Bureau might be required whenever results from these materials are published, but the Bureau would not require that it clear for publication tabulations or analyses based in whole or in part on these materials.

2. The other proposal is that the Bureau draw off on tape the data for a sample of 500,000 – 1,000,000 households and hold it available for special cross tabulations. If necessary, the geographic detail...could be removed here also....

Either proposal would require the spelling out of details, such as methods of sample selection, weighting, drawing of subsamples, estimates of costs, etc.²⁷

Director of Census Robert W. Burgess circulated a memo in April 1960 to members of the Census Advisory Committee of the American Statistical Association, soliciting their views on a proposal to “develop a set of individual census returns which could be made available to research

²⁶ Frederick Osborn, “The 1960 Meeting of the Population Association,” *Population Index*, vol. 26, no. 3 (July 1960): 202.

²⁷ Memorandum from Conrad Taeuber, October 12, 1959, NARA RG 29, Office of the Director, Box 10 MLR A1 389F.

workers for their own use,” assuming legal counsel approved of the scheme.²⁸ Attached to Burgess’ memo was another memo on the “background,” “proposal,” and “some other considerations,” from A. Ross Eckler, Deputy Director of the Bureau of the Census to Kenneth F. McClure, Assistant General Counsel. Eckler’s memo indicated that the Bureau had “long sought ways of making Census data available to research workers for analysis requiring tabulations not included in the published volumes.”²⁹ In August of 1960, Eckler gave a talk at the meeting of the Associated University Bureau of Business and Economic Research (AUBBER) in which he revealed that:

We are studying the possibility of making generally available a set of tapes (or alternatively a deck of punch cards) containing information for a fairly large national sample of households, consisting of perhaps 60,000 units (about 1/10 of one percent of the total). All identifying information will be removed from the tape so as to eliminate the possibility of disclosure of personal information. Such a set of tapes or cards could be used by specialized organizations to make additional tabulations on subjects of a special interest to them.³⁰

Evidently, researchers had communicated to the Bureau that they needed access to “raw materials” that would enable them “to make tabulations which they cannot fully specify in advance.” In this period, researcher access to computers to run analysis at their home institution was difficult, and requesting the Census Bureau to run a special tabulation was cost prohibitive, impractical, and deemed a breach of confidentiality. The new “formula,” proposed by the Bureau could address all these issues in favor of researcher access.

In short, the proposed “formula” would draw a national random sample from the 1960 25% (long-form) sample, “perhaps 35,000 households,” and publish those returns “without name, address, or any specific geographic identification by place, county, or State.” The information for

²⁸ Memorandum from Robert W. Burgess, April 1, 1960, NARA RG 29, Census Advisory Committee on Population Statistics, 1948-1980, Box 1 A1-419.

²⁹ Memorandum from A. Ross Eckler to Kenneth F. McClure, February 18, 1960, NARA RG 29, Census Advisory Committee on Population Statistics, 1948-1980, Box 1 A1-419.

³⁰ A. Ross Eckler, “A Progress Report on Some Phases of the Census Bureau’s Program,” Associated University Bureaus of Business and Economic Research—AUBBER, Eugene, Oregon, August 1, 1960, <https://catalog.archives.gov/id/274957221>.

each individual would be “in code form,” geographic information coded by size and four broad regions, and occupation by “about 26 broad groups.” Publication of the data would be in magnetic tape, punchcards, or a printout “from which an appropriate medium for tabulation might be prepared by the purchasers.” It was further proposed that use of the data would obligate researchers “publish a standard credit to the Bureau” and clarify that the Bureau “assumed no responsibility for the tabulations or analysis.”³¹

By September 1960 the plans for the first microdata sample began to crystallize. As Taeuber described in a letter to Eleanor C. Isbell of the Social Science Research Council,

The Bureau plans to make available a set of tapes containing information from the individual records, and to do so in such a way that there is no breach of confidentiality and no possibility of identifying any individual. The sample will be a national sample and the geographic identification will be in terms only of broad geographic regions and size of place. A sample of approximately 60,000 households is contemplated. The plan is to offer to make the tapes containing the information for each person within the 60,000 households available on a cost basis. Any institution or person purchasing these tapes would be free to use them for any tabulations desired. Subsamples or subgroups might be drawn for special analysis. Preliminary discussion with a number of persons indicates that such material may be particularly useful in the preparation of theses.³²

Taeuber concluded that “It seems likely that a systematic exploration of the possibilities along this line” would reveal other sources of data that could be disseminated in the same manner.

In correspondence with Philip Hauser, Director of the Population Research and Training Center at the University of Chicago, Taeuber confirmed the preparation “of a national sample of about 60,000 census returns which would be made available to researcher workers for their own use.” The Population Technical Advisory Committee and ASA Census Advisory Committee had

³¹ Memorandum from A. Ross Eckler to Kenneth F. McClure, February 18, 1960, NARA RG 29, Census Advisory Committee on Population Statistics, 1948-1980, Box 1 A1-419.

³² Correspondence from Conrad Taeuber to Eleanor C. Isbell, September 15, 1960, NARA RG 29, Office of the Director, A. Ross Eckler, Correspondence Files – Taeuber Chron. 1940-1968, Box 9 MLR A1 389F.

both “favorably reviewed” the proposal.³³ By October 1960, the national sample was now being referred to as “our proposed public use sample” in correspondence between Taeuber and Hauser.³⁴ Seven months later, a memo was circulated to the branch chiefs of the Population Division, outlining the basic specifications of the sample, and requesting the branch chiefs’ review and recommendations to the proposed “One in a Thousand Sample.”³⁵ Originally projected to become available “early in 1962,” regular decennial census work delayed the release of the public use sample until later that year.³⁶

In May and June of 1961, the plans for the one-in-a-thousand sample were coalescing in:

A revised and simplified statement, regarding the content and arrangement of the one in a thousand sample. The following changes have been made in the content and presentation of this statement, as compared with the earlier drafts of March 15, 1961...The presentation proposes a 120-digit record for each person in a housing unit on electronic tape. It proposed that this record be available in a form compatible with the ‘more popular electronic computers.’ This proposal arises for several reasons. In the first place it is necessary to convert all codes into excess 3 on the magnetic tape before they can be converted into punchcards. The primary difference between magnetic tape that is usable on simpler computers will read excess 3 numbers but will not read binary numbers. Once this conversion has been completed to meet the needs of conversion to punchcards we will have a tape which will be immediately legible to the simpler Reminton Rand computers.³⁷

³³ Correspondence from Conrad Taeuber to Philip M. Hauser, September 21, 1960, NARA RG 29, Office of the Director, A. Ross Eckler, Correspondence Files – Taeuber Chron. 1940-1968, Box 9 MLR A1 389F.

³⁴ Correspondence from Conrad Taeuber to Philip M. Hauser, October 6, 1960, NARA RG 29, Office of the Director, A. Ross Eckler, Correspondence Files – Taeuber Chron. 1940-1968, Box 9 MLR A1 389F.

Correspondence from congrid Taeuber to Philip M. Hauser, October 18, 1960, NARA RG 29, Office of the Director, A. Ross Eckler, Correspondence Files – Taeuber Chron. 1940-1968, Box 9 MLR A1 389F.

³⁵ Memorandum from Howard G. Brunsman to Henry S. Shrylock, May 26, 1961 (attachment, “The One in a Thousand Sample,” dated May 10, 1961), NARA RG 29, Eighteenth Decennial Census Methodological Files Concerning Program Planning and Development, 1955-1965, 1960 Census Sampling Theory and Techniques, Container 1.

³⁶ Correspondence from Conrad Taeuber to Otis Dudley Duncan, June 7, 1961, NARA RG 29, Office of the Director, A. Ross Eckler, Correspondence Files – Taeuber Chron. 1940-1968, Box 9 MLR A1 389F. “The One-In-A-Thousand Sample,” October 12, 1961, NARA RG 29, Eighteenth Decennial Census Methodological Files Concerning Program Planning and Development, 1955-1965, 1960 Census Sampling Theory and Techniques, Container 1. “One-In-A Thousand Population Census Sample,” *Population Index*, vol. 28, no. 2 (April 1962): 126. “Two National Samples of the Population of the United States,” *Population Index*, vol. 28, no.3 (July 1962): 238-239.

³⁷ Memorandum from Howard G. Brunsman to Henry S. Shrylock, May 26, 1961 (attachment, “The One in a Thousand Sample,” dated May 10, 1961), NARA RG 29, Eighteenth Decennial Census Methodological Files

Our position in relation to customers for our data in the form of tape. With the increasing availability of computers in business organizations and in other government agencies, there may be some advantage to being in step with the majority. Of course, when conversion ceases to be a problem, this factor loses most of its importance.³⁸

In July 1961 A. Ross Eckler, Deputy Director, Bureau of the Census, reported to the American Library Association that in addition to the availability of special tabulation created by the Bureau, “there are opportunities for additional tabulations designed for the needs of a particular user. These may take the form of tabulation which the user carries out with the benefit of a specially designed punch card deck, or of a set of reels of tape containing the information for a well-designed sample of about 60,000 households. These decks of cards or reels of tapes, which may be purchased for private use, have been so designed that there can be no disclosure of the information pertaining to a particular household or individual.”³⁹

The Census Bureau issued a summary statement on the “One-In-A-Thousand Sample” on October 12, 1962 and indicated that, in an effort to “facilitate the more detailed analysis of the results of the 1960 Census,” the Bureau planned to “assemble and to make available a special subsample” of the 1960 census. Records would be made available on “10 reels of magnetic tape designed for use on the more popular types of electronic computers, as well as decks of 80-column punchcards designed for use on conventional punchcard tabulators.”⁴⁰ According to the report, the

Concerning Program Planning and Development, 1955-1965, 1960 Census Sampling Theory and Techniques, Container 1.

³⁸ Correspondence from Conrad Taeuber to Robert F. Drury, June 26, 1961, NARA RG 29, Office of the Director, A. Ross Eckler, Correspondence Files – Taeuber Chron. 1940-1968, Box 9 MLR A1 389F.

³⁹ A Ross Eckler, “Keeping Up With the Census,” Reference Services Meeting of the American Library Association, Cleveland, Ohio, July 11, 1961, <https://catalog.archives.gov/id/274957279>.

⁴⁰ U.S. Census Bureau, “The One-In-A-Thousand Sample,” October 12, 1961, NARA RG 29, Eighteenth Decennial Census Methodological Files Concerning Program Planning and Development, 1955-1965, Container 1.

materials were expected to be available in the spring of 1962, but this proved to be optimistic.⁴¹ By November 1961, consensus was reached on the form the public use sample would take. In a memorandum with the subject, "Tape specifications for the permanent file for the one per thousand sample," Richard A. Hornseth, Chief, Computer Programming Branch, Decennial Operations Division, stated "It appears that most needs will be best served if the contemplated permanent file for the one per thousand sample is put on the large plastic reels at 128 to the inch density with 2.4 inch interblock spacing and with no blockette spacing... we could use 1.2 inch blockette spacing in the permanent file and have a truly universal file compatible with the 1105, 1103, Univac, tape-to-tape (1410), tape-to-card (1410), and the buffered and unbuffered High Speed Printers."⁴²

The 1-in-1,000 and 1-in-10,000 public use samples finally became available to the public in the second half of 1962. Presentations by Census Bureau staff at the Annual American Statistical Association Meeting in December 1961 promised the production and availability of "a new national sample of one person in a thousand" by fall or late 1962.⁴³ In the April 1962 issue of *Population Index*, the announcement "One-In-A-Thousand Population Census Sample" described the new data products and concluded "It is anticipated that these materials will be available during the later part of 1962. The cost to each subscriber will depend upon the number of subscriptions.

⁴¹ U.S. Census Bureau, "The One-In-A-Thousand Sample," October 12, 1961, NARA RG 29, Eighteenth Decennial Census Methodological Files Concerning Program Planning and Development, 1955-1965, Container 1.1.

⁴² Memorandum from Richard A. Hornseth to Sigmund Schor, November 27, 1961, NARA RG 29, Eighteenth Decennial Census Methodological Files Concerning Program Planning and Development, 1955-1965, 1960 Census Sampling Theory and Techniques, Container 1. "Current and Prospective Developments in Federal Statistics: Bureau of the Census Program, prepared for submission to meeting of the American Statistical Association, Minneapolis, September 8, 1962, <https://catalog.archives.gov/id/274957399>.

⁴³ Charles B. Lawrence, Jr. "Availability of Magnetic Computer Tapes Containing Census Information," Summary remarks for the meeting on census tracts at the 121st Annual Meeting of the American Statistical Association, New York City, December 28, 1961, <https://catalog.archives.gov/id/274957294>. Richard M. Scammon and A. Ross Eckler, "Ten Major Contributions of the 1960 Census to Marketing Planning—And How We Hope to Make the Census Tools Even More Useful," paper to be presented at the 121st Annual Meeting of the American Statistical Association, New York City, December 29, 1961, <https://catalog.archives.gov/id/274957301>.

It is expected that the cost will be between \$4,000 and \$7,000 each.”⁴⁴ *Population Index* provided an update on the public use sample products in the July 1962 issue. “In the April 1962 issue ... announcement was made that the Bureau of the Census is planning to make available, on a cost basis, reels of magnetic tape or a set of punchcards containing a one-in-a-thousand sample of the population of the United States, based on the returns of the 1960 Census. The Census Bureau has now completed its plans for this project. In addition to the one-in-a-thousand (0.1 per cent) sample, a one-in-ten-thousand (0.1 percent) sample will also be prepared.”⁴⁵ Internally, the Census Bureau publicized a new tape service making computer tapes available to the public for a fee. Plans to expand this new service were “scheduled for production next fall” and would “contain detailed information on a sample of one person per thousand of the U.S. population.”⁴⁶

Conclusion

Our research clearly indicates that the first microdata sample was an outcome of interactions between Census Bureau staff and members of the academic community, especially the PAA and the ASA. *Population Index* noted “These samples were developed in response to strong recommendations by a number of social scientists that such material would be a valuable tool for researchers.”⁴⁷

The most important player in the design and execution of the first microdata sample was Conrad Taeuber. Like most Census Bureau leaders of the era, Taeuber had close ties to PAA. Among other roles, he served as the fourth Secretary of the Association (1939-42) and the 12th

⁴⁴ “One-in-a-Thousand Population Census Sample,” *Population Index*, vol. 28, no. 2 (April 1962):126.

⁴⁵ “Two National Samples of the Population of the United States,” *Population Index*, vol. 28, no. 3 (July 1962): 238-239.

⁴⁶ “Computer Tapes Now Available to the Public, *Census Bulletin*, Vol. XII, no. 3 (February 2, 1962): 4, <https://www.census.gov/library/publications/1962/comm/cb-volume-12-3.html>.

⁴⁷ “Two National Samples of the Population of the United States,” *Population Index*, vol. 28, no. 3 (July 1962): 238-239.

President (1948-49). It is therefore unsurprising that Taeuber was closely attuned to the needs of the demographic research community. In 1991, Taeuber received the PAA Robert J. Lapham award for his contributions to the development of demographic research infrastructure.

This research demonstrates that census microdata emerged from the fruitful interaction of academic social scientists with Census Bureau statisticians, and especially the interaction of Taeuber with the demographic user community at PAA.

The cost of the 1960 microdata was high, and few social scientists had access to the computing equipment and software needed to analyze the tapes. Nevertheless, there was significant use very early on. By 1965, 40 copies of the data had been sold.⁴⁸ The first paper using the 1960 sample at PAA did not appear until 1964, and the earliest publication appeared in 1965.⁴⁹ By 1968, there were at least seven publications and six dissertations, in addition to numerous conference presentations based on the data.⁵⁰

⁴⁸ Paul G. Glick, "Census Data as a Source for Theses and Dissertations in the Field of Sociology," *The Milbank Memorial Fund Quarterly*, vol. 43, no. 1 (January 1965): 17-30.

⁴⁹ John C. Belcher, "The One-Person Household in the United States: Based on the One-in-1000 Sample of the 1960 Census," *Population Index*, vol. 29, no. 3 (July 1963): 313. James D. Cowhig and Calvin L. Beale, "Levels of Living Among Whites and Nonwhites" in *White-Nonwhite Differentials in Health, Education, and Welfare*, U.S. Department of Health, Education, and Welfare (February-October 1965): 56-65.

⁵⁰ Publications: Glen G. Cain, *Married Women in the Labor Force: An Economic Analysis* (Chicago, Illinois: University of Chicago Press, 1966). Finis Welch, "Measurement of the Quality of Schooling," *The American Economic Review*, vol. 56, no. 1/2 (March 1966): 379-392. Jack Ladinsky, "Sources of Geographic Mobility Among Professional Workers: A Multivariate Analysis," *Demography*, vol. 4, no. 1 (1967): 293-309. A.E. Bayer, "Differential Fertility of Nativity-Parentage Groups in the United States: The Assimilation of European Female Foreign Stock," *Sociological Inquiry*, vol. 37, no. 1 (January 1967): 99-108. S. Mitra, "Income, Socioeconomic Status, and Fertility in the United States," *Obstetrical & Gynecological Survey*, vol. 22, no. 2 (April 1967): 270-272. John C. Belcher, "The One-Person Household: A Consequence of the Isolated Nuclear Family?" *Journal of Marriage and Family*, vol. 29, no. 3 (August 1967): 534-540. D. McAllister, "The demand for rental housing: An investigation of some demographic and economic determinants," *The Annals of Regional Science*, vol. 1, issue 1 (December 1967): 127-142. Dissertations: Glen G. Cain, "Labor Force Participation of Married Women," Ph.D. dissertation, The University of Chicago (June 1964). Alan E. Bayer, "The Assimilation of American Family Patterns by European Immigrants and Their Children," Ph.D. dissertation, The Florida State University (August 1965). Finis Welch, "The Determinants of the Return to Schooling in Rural Farm Areas, 1959," Ph.D. dissertation, The University of Chicago (June 1966). Daniel Carl Rogers, "Private Rates of Return to Education in the United States: A Case Study," Ph.D. Dissertation, Yale University (1967). Albert McCarroll Marckwardt, "Differentials of Recent Internal Migration in the United States," Ph.D. dissertation, The University of Michigan (1968). Lucy Bunzel Mallan, "Financial Patterns in Households with Working Wives,"

Despite the early uptake of the microdata, it quickly became apparent that there were substantial limitations of the sample, including the lack of geographic identifiers and the small sample size, which limited the potential for analysis of population subgroups. The Census Bureau rectified these issues in the next 1970 census by releasing six independent 1-in-100 samples, providing data on a total of six percent of the population. The new samples provided much more detail, especially for geographic variables.

A key innovation of the early 1970s was preparation of a new version of the 1960 census microdata. The new version was ten times the size of the original 1960 sample. Most important, the new 1960 sample released in January 1973 was reorganized to be as compatible as possible with the 1970 sample, with identical coding schemes for most variables and an identical record layout. This made it easy to compare 1960 and 1970 and to assess the massive social and economic changes that occurred in that tumultuous decade.

As Dudley Kirk expressed in 1963,

A further significant contribution of the cards and tapes is their meaning for posterity. In addition to the fixed census tabulations, published and unpublished, the one-in-a-thousand sample represents a permanent record of the 1960 population in a viable form that can be readily manipulated in the future to gain historical series in terms of problems and trends now unforeseen.⁵¹

With the rapid decline in the cost of computing, together with the emergence of the first statistical packages that simplified coding, the public use samples became core resources of U.S. social science.

Ph.D. dissertation, Northwestern University (June 1968). For conference presentations, see for example, "The 1968 Meeting of the Population Association," *Population Index*, vol. 34, no. 3, (July 1968):263-327.

⁵¹ Dudley Kirk, "The 1960 Census—Research Potential," *American Behavioral Scientist*, vol. 6, issue 9 (May 1963): 14-18.