

Frozen Film and FOSDIC Forms

Restoring the 1960 U.S. Census of Population and Housing

STEVEN RUGGLES
MATTHEW SCHROEDER
NATASHA RIVERS
*Minnesota Population Center
University of Minnesota*

J. TRENT ALEXANDER
TODD K. GARDNER
*U.S. Census Bureau
Washington, DC*

Abstract. In this article, the authors describe a collaboration of the Minnesota Population Center (MPC), the U.S. Census Bureau, and the National Archives and Records Administration to restore the lost data from the 1960 Census. The data survived on refrigerated microfilm in a cave in Lenexa, Kansas. The MPC is now converting the data to usable form. Once the restored data are processed, the authors intend to develop three new data sources based on the 1960 census. These data will replace the most inadequate sample in the series of public-use census microdata spanning the years from 1850 to 2000, extend the chronological scope of the public census summary files, and provide a powerful new resource for the Census Bureau and its Research Data Centers.

Keywords: census, data recovery, demography, IPUMS, microdata, Minnesota Population Center, public-use sample

During the past several years, the United States Census Bureau has devoted significant effort to recovering machine-readable data from the long forms of the 1960–80 Censuses of Population and Housing. These data, combined with comparable material from the censuses of 1990 and 2000, create a rich series of data on the U.S. population in the last four decades of the twentieth century.

The purpose of this article is to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views or opinions in the article are the authors' own and do not necessarily reflect the views or opinions of the U.S. Census Bureau.

Address correspondence to Steven Ruggles, Minnesota Population Center, University of Minnesota, 225 19th Ave. S, Minneapolis, MN 55455, USA, or Todd K. Gardner, U.S. Census Bureau, Center for Economic Studies, 4600 Silver Hill Rd., Washington, DC 20233, USA. E-mail: ruggl001@hist.umn.edu or todd.k.gardner@census.gov

Data from 1970 and 1980 were successfully verified and converted to modern formats. The 1960 data set, however, could not be completely retrieved; information from some parts of the country—most notably Cook County (Chicago), Illinois—was lost.

This article describes an unusual collaboration of the Minnesota Population Center, the Census Bureau, and the National Archives and Records Administration (NARA) to restore the lost data from the 1960 Census. The data survived on refrigerated microfilm stored on shrink-wrapped pallets in a cave in Lenexa, Kansas. We are now converting the data to usable form through optical mark recognition. Once we have finished processing the data, we intend to develop three new data sources based on the 1960 census. These data will replace the most inadequate sample in the series of public-use census microdata spanning the years from 1850 to 2000, extend the chronological scope of the public census summary files, and provide a powerful new resource for the Census Bureau Research Data Centers.

We begin by describing the pioneering history of the 1960 census. It was the first census to provide data in electronic form; today, however, those data are sadly out of date. We explain the limitations of the existing 1960 microdata and small-area summary files. We then describe the National Historical Census Files Project—a collaborative effort to recover, verify, document, and disseminate all surviving individual-level data for the period since 1960—which uncovered the serious flaws in the surviving 1960 data. We then describe how we extracted data from the microfilm and describe the steps we are taking to restore the lost records. We conclude with a brief description of the planned data products.

Innovations of the 1960 Census

The 1960 census played a pivotal role in the development of quantitative social science. It was the first census to provide researchers with machine-readable data on computer tapes. Indeed, the 1960 census was the first large-scale machine-readable data source to be widely used for social science research.

In 1963, the Census Bureau produced a 1-in-1,000 sample of the data tapes they had used to create tabulations for the published census volumes (U.S. Census Bureau 1963b). The Census Bureau removed detailed geographic codes and other potentially identifying information and made the sample available to the research community. There were three formats: punch cards, Remington-Rand magnetic tape suitable for use on Univac computers (13 tapes), and IBM magnetic tape (7 tapes) (U.S. Census Bureau 1963a). Despite the high cost of data processing, the 1-in-1,000 sample was an extraordinary success; it revolutionized analysis of the U.S. population and led to an explosion of new census-based research. As Otis Dudley Duncan (1974, 5097) 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.

The new sample not only allowed researchers to make tabulations tailored to their specific research questions, it also allowed them to apply new methods—such as multivariate techniques—to the analysis of census data. A quick JSTOR search reveals more than 50 citations of this early sample in the top journals of economics, sociology, and demography within the first decade after the data became available. The 1960 samples were widely used for training a new generation of quantitatively oriented social scientists. The use of census microdata grew rapidly over the next three decades, and these data have become an indispensable component of social science infrastructure.

What made the 1960 sample so compelling was not simply its availability in machine-readable form. Equally important were innovations in sample design that yielded far richer data than had previous censuses. The Census Bureau first made use of sampling in the 1940 census. In 1940, each sheet of the census enumeration schedule had 40 rows, with one row for each respondent, and 34 columns, with a different census question in each column. This was essentially the same layout as had been used since 1850. What was different in 1940 was that two rows on every page were highlighted, and the individuals enumerated on those rows were asked a set of 17 supplemental questions. This yielded a systematic geographically stratified sample representing 5 percent of the popula-

tion. The Census Bureau adopted a similar form in the 1950 census, except that the sample density increased to 20 percent, and many more questions were asked on a sample basis.

Sampling allowed the census to expand the number of questions; by 1950, the census included 64 questions, more than twice the number asked in 1930. But there were limitations to the approach. The addition of so many detailed questions strained the conventional door-to-door enumeration methodology. If the sample individual was not present when the enumerator visited the household, the respondent might not know the answers to some of the detailed questions on the form, such as nonwage income or highest grade completed. Another liability of this sampling approach was that only one individual in each household ordinarily would have been asked the supplemental questions. This means, for example, that one cannot compare the income or education of husbands with that of their wives, since only one member of a couple would have been asked those questions.

These problems were resolved in 1960 by a redesign of the census form and new enumeration procedures. There was a separate census form for each household. The Census Bureau mailed each householder an “Advance Census Report” form to fill out before the census taker arrived. The advance report form contained a sharply restricted set of 6 population questions (name, relationship, sex, race, date of birth, and marital status) and 13 housing questions. The enumerator went from house to house and collected the forms in person. If a household failed to complete its advance report form, or filled it out incorrectly, the census taker did a conventional interview. The census enumerators carried out sampling on a household basis rather than an individual basis.¹ The enumerators designated every fourth household visited as a sample household, and gave the respondent a sample form containing 28 additional questions for each person in the household and 33 additional housing questions. The enumerators requested that respondents complete the form and mail it to their local census office in a postage-paid envelope. When they received the forms, Census Bureau personnel checked the sample forms for consistency and completeness and conducted telephone or in-person inquiries to complete unanswered questions when necessary.

This two-stage procedure and the Advance Census Report greatly reduced the complexity of fieldwork, and testing suggested that the accuracy of responses to the long-form questions was considerably higher than in 1950. The increased efficiency allowed expansion of the census to a total of 81 questions, approximately the same number as were asked on the 2000 census long form. In addition to improving the reliability of the census and expanding its content, the new census procedures meant that every individual in a sample household was asked the full roster of sample questions. This substantially increases the usefulness of microdata from 1960 in comparison

with that from 1940 or 1950, since analysts can simultaneously access the characteristics of multiple household members.

Limitations of the 1960 Microdata Sample

Despite its deep impact on the social sciences, the initial 1960 sample had two serious limitations. First, the sample size was relatively small. The 1-in-1,000 sample density yielded about 180,000 person records. Relative to the capacity of computers in 1964, this was an enormous number of cases. As computing costs declined and researchers began to use the sample for detailed analyses of small population subgroups, however, its limitations became apparent. Second, the 1960 public-use sample provided little geographic information. To ensure confidentiality, the Census Bureau stripped off all information on places below the state level. This meant, for example, that it was impossible to extract a subsample of the New York City population. The restricted geographic information applied not only to place of residence, but also to the variables on migration and journey to work.

In 1973, the Census Bureau responded to user demand by enlarging the 1960 sample from 1-in-1,000 to 1-in-100 (U.S. Census Bureau 1973). The 1960 sample nevertheless remains much smaller than the microdata samples available for subsequent census years. For the censuses of 1970 through 2000, the Census Bureau has released samples covering between 6 percent and 9 percent of the population. The number of microdata records currently available for each census year is shown in figure 1.

Starting with the 1970 census, the Census Bureau also addressed the problem of poor geographic identification. In 1970, the microdata identified geographic areas of 250,000 or more, and beginning in 1980 the samples identified areas as

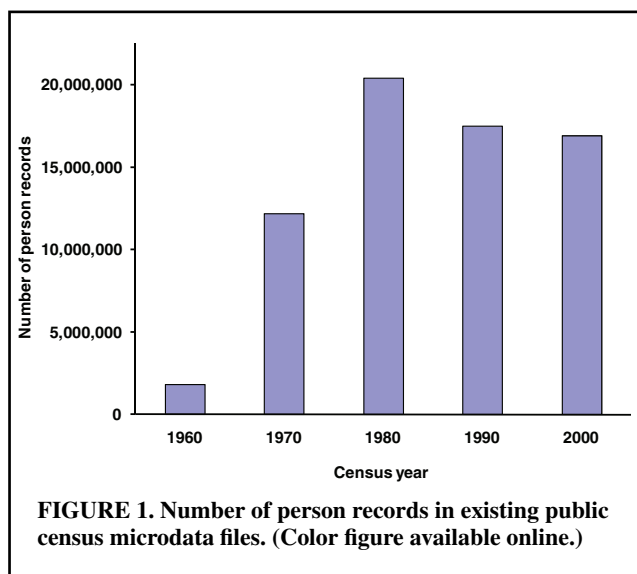


TABLE 1. Number of Geographic Units Identified in Public Use Microdata Samples

Year	Places identified
1960	51
1970	408
1980	1,154
1990	1,726
2000	2,071

small as 100,000. The number of geographic areas identified in the census microdata files rose dramatically, from just 51 in 1960 to 1,726 in 1990 (see table 1).

When the Census Bureau proposed reducing geographic precision in Census 2000 as a means of reducing disclosure risk, the research community responded vigorously and unanimously, writing hundreds of letters and messages to protest the change. In a survey of data users conducted by the Minnesota Population Center, 65 percent of faculty researchers indicated that any reduction of geographic detail would be “catastrophic” to research in their field. Several researchers made a direct comparison to the limitations of the 1960 sample. For example, Patricia Beeson, an economist at the University of Pittsburgh, wrote, “1960 data cannot be used for substate analysis, and it would be catastrophic for researchers examining local areas to also lose 2000” (Ruggles, Fitch, and Sobek 2000, 70). In the end, the outcry from researchers was persuasive, and the Census Bureau provided information on more than 2,000 places in the Census 2000 microdata.

At first, the very large samples that became available in the 1980s were too expensive for most researchers to process. With the advent of comparatively inexpensive UNIX workstations, however, the cost of computing declined rapidly during the first half of the 1990s; by early in the first decade of the twenty-first century, even desktop personal computers were capable of processing the largest census microdata samples. Since 1996, online data dissemination tools developed at the Minnesota Population Center have provided researchers with easy access to large microdata extracts. Accordingly, the largest census microdata files—once available to only a few researchers at great expense—are now accessible to virtually all social scientists.

Today, most researchers rely on the largest files available for the period they are studying. Between 1996 and 2005, nearly 80 percent of *Demography* articles based on recent census microdata used the highest density samples available.²

Most of these analyses depend on information for small population subgroups, ranging from same-sex couples to the grandchildren of immigrants. In many instances, the large samples also permit the use of innovative methods; to take just one example, these files have allowed demographers to

carry out multilevel contextual analyses by making it feasible to assess the characteristics of small geographic areas.

The seminal 1960 public-use sample is now obsolete. Because of small sample size and poor geographic information, the 1960 sample simply cannot support the exciting studies demographers and economists are undertaking with more recent census data. Among the 52 studies using U.S. census microdata published in *Demography* during the past decade, only 11 could be replicated using the 1960 data; in the other cases, the topic or methodology requires larger samples or more detailed geography. As a result, researchers using microdata increasingly must skip the 1960 census year. The limitations of the existing 1960 sample virtually guarantee that many analyses of long-term social and demographic change encompass only the period from 1970 forward, thereby vitiating one of the greatest potential strengths of the IPUMS series of U.S. public-use census microdata.

Limitations of the 1960 Summary Files

The small area statistics available for the 1960 census are even weaker than the microdata. There are currently three sources of electronic summary statistics for 1960:

1. In 1971, the DUALabs company prepared and disseminated a set of machine-readable tract-level data from the 1960 census (U.S. Census Bureau 1971). These data were never widely used, partly because until recently they were available only in obsolete DUALabs compressed format, and partly because DUALabs produced only 63 tables for each tract.³
2. Beginning in 1968, the Inter-university Consortium for Political and Social Research (ICPSR) undertook a program to keypunch state- and county-level statistics from printed sources (ICPSR 1973). During the past decade, that work has been corrected and supplemented by Michael R. Haines (Haines and ICPSR 2005). Although

widely used, the Haines–ICPSR file includes only 32 tables for each state and county.

3. The Civil Works division of the U.S. Army Corps of Engineers had the Census Bureau prepare a special tabulation on county characteristics in the late 1960s, and these data were preserved at Lawrence Berkeley Laboratories. From 2001 to 2005, the University of California, Berkeley, data archive managed to convert most of the data from an obsolete compressed format, but not all counties could be recovered (University of California, Berkeley 2006). Most of the Berkeley data is redundant with the Haines/ICPSR file, but there is some additional detail on educational attainment and employment.

Table 2 compares the available summary data in 1960 with that in later census years. A data element is one cell of a table; thus, for example, a table of age by sex with 15 age groups would have 30 data elements. The year 1960 has far fewer data elements than subsequent census years. Furthermore, in 1960 those elements are available for only three geographic levels: states, counties, and tracts. For each subsequent census, the Census Bureau provided additional geographies, such as place, minor civil division, metropolitan area, and public-use microdata area. Moreover, in 1960 only metropolitan areas were divided into tracts, so the tract file covers only about 70 percent of the population. The best measure of the total quantity of small-area data available in each census year is the number of gigabytes each census consumes in the National Historical Geographic Information System database, which incorporates all extant aggregate population and housing data in a consistent format. As shown in the final column of table 2, the 1960 census has only a tiny fraction of the data available for the more recent census years.

The 1960 summary files lack the most basic tables required to understand geographic patterns of demographic and economic change. For example, the 1960 tract files combine information on all races except for whites into a single category of “nonwhite.” No information is available on the

TABLE 2. Selected Measures of Machine-Readable Census Summary Files

Census year	Data elements		Geographic levels	NHGIS (gigabytes)
	Counties	Tracts		
1960	393 ^a	1,068	3	0.3
1970	10,804	3,157	14	14.4
1980	11,495	11,495	27	53.7
1990	21,218	21,218	54	237.3
2000	37,273	37,273	79	1,222.8

Note. NHGIS = National Historical Geographic Information System.

^aexcluding Berkeley file.

number of blacks, so standard indices of residential segregation cannot be constructed for 1960. Critical variables such as educational attainment, school enrollment, and income are available only as frequency distributions; they are not crossclassified by age or gender. The content of the 1960 county files differs from that of the tract data, but in general even less detail is available for counties.

The period from 1960 to 1970 witnessed a dramatic spatial reconfiguration of metropolitan areas. The development of interstate highways and mass migration to suburbs had numerous long-term consequences, and not all of them benefited the health and well-being of the population. New spatial tools developed for the National Historical Geographic Information System at the Minnesota Population Center can help us understand such transformations. Without adequate small-area data for 1960, however, such studies are effectively limited to the period since 1970.

National Historical Census Files Project

For the past eight years, the Census Bureau and the Minnesota Population Center (MPC) have been engaged in a collaborative data recovery project, under the direction of Todd Gardner, with far-reaching implications for social science research. The goals of the National Historical Census Files Project are to recover, preserve, document, harmonize, and disseminate all surviving machine-readable population census microdata from 1960 through 1980 (Gardner 2001).

These surviving historical census microdata include long-form population and housing records for the 1960–2000 censuses and short-form records for the period from 1970 to 2000. Altogether, the collection includes over one billion person records and 400 million household records. This represents 16 times the number of records contained in the existing public-use microdata files for those census years. The files include far greater geographic detail than is available in public-use census microdata; in most cases, the specific block of residence is identified. The Census Bureau is making the data from 1970 to 2000 available to qualified researchers with approved projects only through its network of 12 Census Bureau Research Data Centers, which provide the necessary security for confidential data.⁴

The first goal of the project was to produce a clean and complete ASCII-format version of the data with thorough documentation. This involved two basic tasks: The first was transferring all raw files for both long-form and short-form data to a server where the data could be verified. We were fortunate to locate two independent conversions of the data. One was fairly easy to access, but the other existed only on tapes readable by an outdated Unisys mainframe computer. All of the tapes had to be transferred to a Linux server and then merged. The second basic task was comparing

the two independent conversions record by record, field by field.

Not only was the process laborious, but it also required more storage space than was available at the time. The MPC and the Maryland Population Research Center purchased a 1.8-terabyte RAID device for the Census Bureau to carry out this work. Once that was in place, the verification process went smoothly. All discrepancies were easily resolved, and the data transfer was certified a success.

We then tabulated figures using the verified microdata files, and this uncovered serious problems. The population totals for 1970, 1980, and 1990 matched the published numbers, but we found discrepancies for 1960. In 17 counties, the long-form data set contains fewer cases than were recorded in the published tables. The shortfall is consistent in both copies of the 1960 data, meaning that data loss occurred many years ago, which was confirmed by the National Archives and Records Administration (Adams and Brown 2000, 16).

The largest problem is in Cook County, Illinois, which is missing information on 1,150,124 persons. Also, 12 other Illinois counties account for most of the other missing cases. Because the 1960 data were not nationally representative, they could not be used to create new data products.

The National Historical Census Files represent a new class of source material for social scientists. We anticipate that the availability of consistent microdata for the entire population over a broad time span will have a profound effect on the practice of social science research, comparable in its impact to the first release of census microdata in 1964. Epidemiologists can use these data to assess the impact of neighborhood change on health and well-being. The availability of high-density microdata will enable entirely new approaches and methods in the study of residential segregation. Among other key substantive areas are the decline and renaissance of central cities, immigrant and ethnic settlement patterns, suburbanization and urban sprawl, rural depopulation and agricultural consolidation, the identification of concentrated poverty, transportation, the transformation of electoral politics, geographic criminal justice studies, and environmental justice. Analysts of small population subgroups—such as American Indian tribes, specific occupation groups, and particular immigrant groups—will for the first time have sufficient cases to carry out their analyses. Without data from 1960, however, the extraordinary potential of this unique data series is greatly compromised.

Recovering the Lost Data

Fortunately, the corrupted 1960 data tapes were not the only machine-readable source of data from the census. The 1960 census was the first fully computerized enumeration. Census information was converted to digital form by means of an innovative optical scanning system, the Film Optical

Sensing Device for Input to Computers (FOSDIC). Census enumerators received completed census forms from respondents and transferred the information to bubble-coded optical mark recognition forms (figure 2 and figure 3).

On the front of the FOSDIC form, the upper panel recorded housing information (figure 2), and the lower panel recorded information on the first individual in the household (figure 3). On the reverse of the form was room for two more individuals. For households larger than three, additional forms were used as needed, and the "continuation" bubble was filled in field 3 of the household form. The household form in 1960 had four variations with slight differences in the questions asked. The individual form included open-ended questions, such as occupation, industry, mother tongue, and place of residence five years ago. These were classified and coded onto the FOSDIC forms by census staff specially trained to code specific fields (see fields labeled "For office codes" on figure 3). In addition to the household and person forms, coders prepared "breaker sheets" to be inserted between each enumeration district to provide geographic identifiers.

Once the forms were complete, they were microfilmed, and the film was scanned by FOSDIC machines (figure 4). The Census Bureau developed the FOSDIC scanner in con-

junction with the National Bureau of Standards. Work on the device began in 1951, with the goal of reducing the 200,000 days of keypunching required for the 1950 census. The machine was capable of reading forms at the extraordinary rate of 100 frames (24,000 characters) per minute. The 1960 census used five FOSDIC machines, each of which was staffed by 30 operators and technicians working three eight-hour shifts. This crew of 150 digitized the short-forms in six months and the long form in nine months. By comparison, the 1950 census used nearly 2,000 keypunch operators for 14 months (U.S. Census Bureau 1966; Weik 1961, 288). The new technology was so successful that it was also used for the 1970, 1980, and 1990 censuses.

Although some of these electronic records were lost, fortunately the complete set of microfilmed enumeration manuscript forms survives. The long-form data covering 25 percent of the population is stored on approximately 30,000 100-foot reels of 16-millimeter film. Each reel includes information from approximately 10 enumeration districts, or 6,000 individuals. The film is stored in the cave shown in figure 5 at the Regional Records Services Facility of NARA in Lenexa, Kansas. It is maintained in a 35°F cold room on shrink-wrapped pallets. In accordance with

Form 60 PH-3 (5) U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS

1960 CENSUS OF POPULATION AND HOUSING

CONFIDENTIAL—The Census is required by the United States Constitution and further authorized by 13 U.S.C. § 9 (a), 221-6. The law requires that the information furnished will be accorded confidential treatment. The Census report cannot be used for purposes of taxation, investigation, or registration.

2. Page No. from PH-1 or PH-2
Mark below:
0 1 2 3 4 5 6 7 8 9

3. If this is a continuation of a household started on a previous sheet, mark below and omit all "H" items on this sheet.

H1. Sample key
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL AM AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT BU BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA DB DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS DT DU DV DW DX DY DZ EA EB EC ED EE EF EG EH EI EJ EK EL EM EN EO EP EQ ER ES ET EU EV EW EX EY EZ FA FB FC FD FE FF FG FH FI FJ FK FL FM FN FO FP FQ FR FS FT FU FV FW FX FY FZ GA GB GC GD GE GF GG GH GI GJ GK GL GM GN GO GP GQ GR GS GT GU GV GW GX GY GZ HA HB HC HD HE HF HG HH HI HJ HK HL HM HN HO HP HQ HR HS HT HU HV HW HX HY HZ IA IB IC ID IE IF IG IH II IJ IK IL IM IN IO IP IQ IR IS IT IU IV IW IX IY IZ JA JB JC JD JE JF JG JH JI JJ JK JL JM JN JO JP JQ JR JS JT JU JV JW JX JY JZ KA KB KC KD KE KF KG KH KI KJ KK KL KM KN KO KP KQ KR KS KT KU KV KW KX KY KZ LA LB LC LD LE LF LG LH LI LJ LK LL LM LN LO LP LQ LR LS LT LU LV LW LX LY LZ MA MB MC MD ME MF MG MH MI MJ MK ML MM MN MO MP MQ MR MS MT MU MV MW MX MY MZ NA NB NC ND NE NF NG NH NI NJ NK NL NM NN NO NP NQ NR NS NT NU NV NW NX NY NZ OA OB OC OD OE OF OG OH OI OJ OK OL OM ON OO OP OQ OR OS OT OU OV OW OX OY OZ PA PB PC PD PE PF PG PH PI PJ PK PL PM PN PO PP PQ PR PS PT PU PV PW PX PY PZ QA QB QC QD QE QF QG QH QI QJ QK QL QM QN QO QP QQ QR QS QT QU QV QW QX QY QZ RA RB RC RD RE RF RG RH RI RJ RK RL RM RN RO RP RQ RR RS RT RU RV RW RX RY RZ SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SR SS ST SU SV SW SX SY SZ TA TB TC TD TE TF TG TH TI TJ TK TL TM TN TO TP TQ TR TS TT TU TV TW TX TY TZ UA UB UC UD UE UF UG UH UI UJ UK UL UM UN UO UP UQ UR US UT UU UV UW UX UY UZ VA VB VC VD VE VF VG VH VI VJ VK VL VM VN VO VP VQ VR VS VT VU VV VW VX VY VZ WA WB WC WD WE WF WG WH WI WJ WK WL WM WN WO WP WQ WR WS WT WU WV WW WX WY WZ XA XB XC XD XE XF XG XH XI XJ XK XL XM XN XO XP XQ XR XS XT XU XV XW XX XY XZ YA YB YC YD YE YF YG YH YI YJ YK YL YM YN YO YP YQ YR YS YT YU YV YW YX YY YZ ZA ZB ZC ZD ZE ZF ZG ZH ZI ZJ ZK ZL ZM ZN ZO ZP ZQ ZR ZS ZT ZU ZV ZW ZX ZY ZZ AA AB AC AD AE AF AG AH AI AJ AK AL AM AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT BU BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA DB DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS DT DU DV DW DX DY DZ EA EB EC ED EE EF EG EH EI EJ EK EL EM EN EO EP EQ ER ES ET EU EV EW EX EY EZ FA FB FC FD FE FF FG FH FI FJ FK FL FM FN FO FP FQ FR FS FT FU FV FW FX FY FZ GA GB GC GD GE GF GG GH GI GJ GK GL GM GN GO GP GQ GR GS GT GU GV GW GX GY GZ HA HB HC HD HE HF HG HH HI HJ HK HL HM HN HO HP HQ HR HS HT HU HV HW HX HY HZ IA IB IC ID IE IF IG IH II IJ IK IL IM IN IO IP IQ IR IS IT IU IV IW IX IY IZ JA JB JC JD JE JF JG JH JI JJ JK JL JM JN JO JP JQ JR JS JT JU JV JW JX JY JZ KA KB KC KD KE KF KG KH KI KJ KL KM KN KO KP KQ KR KS KT KU KV KW KX KY KZ LA LB LC LD LE LF LG LH LI LJ LK LL LM LN LO LP LQ LR LS LT LU LV LW LX LY LZ MA MB MC MD ME MF MG MH MI MJ MK ML MM MN MO MP MQ MR MS MT MU MV MW MX MY MZ NA NB NC ND NE NF NG NH NI NJ NK NL NM NN NO NP NQ NR NS NT NU NV NW NX NY NZ OA OB OC OD OE OF OG OH OI OJ OK OL OM ON OO OP OQ OR OS OT OU OV OW OX OY OZ PA PB PC PD PE PF PG PH PI PJ PK PL PM PN PO PP PQ PR PS PT PU PV PW PX PY PZ QA QB QC QD QE QF QG QH QI QJ QK QL QM QN QO QP QQ QR QS QT QU QV QW QX QY QZ RA RB RC RD RE RF RG RH RI RJ RK RL RM RN RO RP RQ RR RS RT RU RV RW RX RY RZ SA SB SC SD SE SF SG SH SI SJ SK SL SM SN SO SP SQ SR SS ST SU SV SW SX SY SZ TA TB TC TD TE TF TG TH TI TJ TK TL TM TN TO TP TQ TR TS TU TV TW TX TY TZ UA UB UC UD UE UF UG UH UI UJ UK UL UM UN UO UP UQ UR US UT UU UV UW UX UY UZ VA VB VC VD VE VF VG VH VI VJ VK VL VM VN VO VP VQ VR VS VT VU VV VW VX VY VZ WA WB WC WD WE WF WG WH WI WJ WK WL WM WN WO WP WQ WR WS WT WU WV WW WX WY WZ XA XB XC XD XE XF XG XH XI XJ XK XL XM XN XO XP XQ XR XS XT XU XV XW XX XY XZ YA YB YC YD YE YF YG YH YI YJ YK YL YM YN YO YP YQ YR YS YT YU YV YW YX YZ ZA ZB ZC ZD ZE ZF ZG ZH ZI ZJ ZK ZL ZM ZN ZO ZP ZQ ZR ZS ZT ZU ZV ZW ZX ZY ZZ AA AB AC AD AE AF AG AH AI AJ AK AL AM AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT BU BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA DB DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS DT DU DV DW DX DY DZ EA EB EC ED EE EF EG EH EI EJ EK EL EM EN EO EP EQ ER ES ET EU EV EW EX EY EZ FA FB FC FD FE FF FG FH FI FJ FK FL FM FN FO FP FQ FR FS FT FU FV FW FX FY FZ GA GB GC GD GE GF GG GH GI GJ GK GL GM

The image shows a single page of a 1960 FOSDIC form, titled 'Individual panel'. It is a complex form with numerous questions and checkboxes. The questions are organized into columns and rows, with some questions having multiple-choice options and others having numerical scales. The form includes sections for:

- Personal Information:** Name, relationship to head, sex, color or race, month and year of birth, marital status.
- Education:** Highest grade attended, type of school, years of school.
- Work History:** Whether worked last week, how many weeks worked, type of work, class of worker.
- Income:** Wages or salary in 1959, net income after business expenses.
- Other Information:** Languages spoken, father's birthplace, when moved into house, when married, when did he get married, when did he get married for the first time, when did he get married for the first time, when did he get married for the first time.

 The form is filled out with handwritten marks, including checkmarks and numbers, indicating the responses for a specific individual. The bottom right corner of the form has the code '800-1089-011113'.

FIGURE 3. 1960 FOSDIC form: Individual panel. Source: U.S. Census Bureau, "Evaluation and Research Program of the U.S. Censuses of Population and Housing 1960: Background, Procedures, and Forms" (Series ER60, No. 1, U.S. Census Bureau, Washington, DC, 1963).

Census privacy rules, copies of the film are scheduled for public release in April 2032. The confidentiality of the 1960 data greatly complicated the logistics of the project, and it would have been difficult to accomplish without the thoughtful and efficient help of Census Bureau Safety Officer Glen Everhart.

The first task of data restoration was to identify the microfilm reels containing the missing 1960 cases. We knew which census tracts have missing cases, but we did not yet know which specific enumeration districts within these tracts were affected. Because the reels were organized and labeled by enumeration district, to identify the affected reels we had to identify the districts. This task would be simple if we had documentation on the count of long-forms for each enumeration district within each tract, as we could compare the totals with the surviving long-form data. Unfortunately, we could not locate information on enumeration district population counts, despite a thorough search of the Census Bureau archives and NARA collections.

The lack of tract numbers on the reel labels meant that we needed to identify and scan records from entire counties that contained one or more tracts with missing cases. We identified all counties with missing records and selected all microfilm reels containing enumeration districts within those

counties. In each state, about one in five microfilm reels were not labeled as belonging to any particular county or enumeration district. These reels contained an assortment of enumeration districts, many of which had been microfilmed for FOSDIC quality control purposes, and were all labeled "quasi" in the part of the label where most reels listed a specific range of enumeration districts. After looking at the records on a sample of the quasi reels, we determined that each of them contained records from all over the state.

The existence of the quasi reels greatly increased the potential cost of data recovery, particularly for states missing only a few records. In Hennepin County, MN, for instance, where the data file was missing about 100 cases, we realized that we would need to select all reels from Hennepin County as well as all quasi reels from the state of Minnesota (because the quasi reels contained some cases from Hennepin County). Recovering those 100 cases from Hennepin County would have involved scanning and entering data on hundreds of thousands of cases that were already in the data set and did not need to be recovered. We determined that, for all states other than Illinois, the existence of the quasi reels made the data recovery prohibitively expensive. This was not a serious problem, however, because more than 99 percent of the missing cases were in Illinois. In the end,

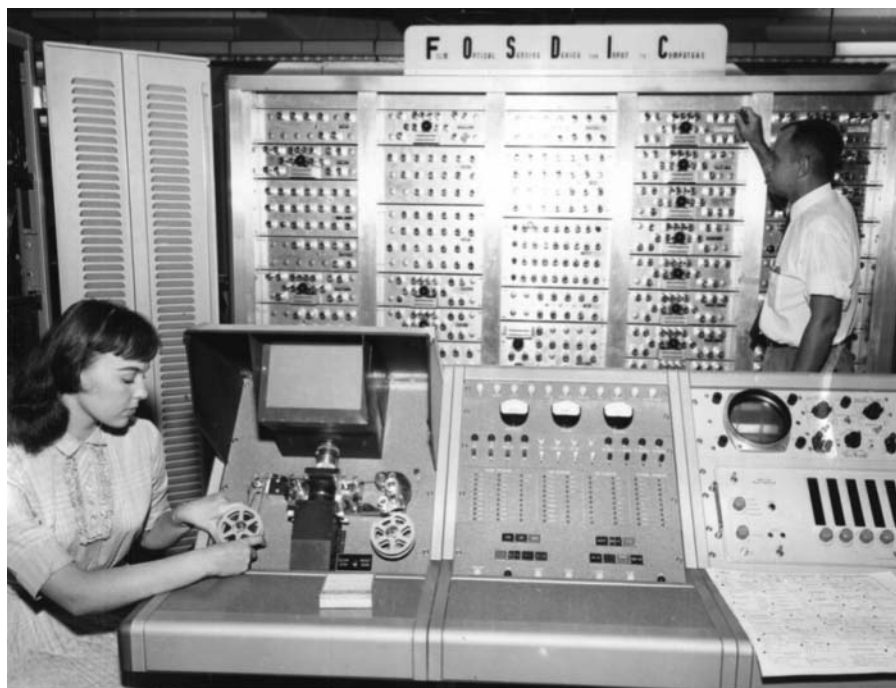


FIGURE 4. FOSDIC machine (partial view). The FOSDIC machines used for the 1960 census recorded the information from the microfilm on seven-track magnetic computer tapes that were readable by the Census Bureau's computers. *Source:* U.S. Census Bureau, "PhotoZone: Centennial Celebration," U.S. Census Bureau Public Information Office, 2004, <http://www.census.gov/pubinfo/www/photos/centennial.html>.

we selected only Illinois reels for data recovery, including Cook County and all of Illinois's quasi reels. The weights will be adjusted to correct the remaining discrepancies, all of which are minor.



FIGURE 5. The entrance to the Lenexa, Kansas, Federal Records Center. (Color figure available online.)

Our original proposal called for shipping the cold microfilm from the Lenexa, Kansas, facility to be duplicated at the National Archives in College Park, Maryland, and then sending the copy to the Census Bureau's National Processing Center in Jeffersonville, Indiana, for scanning. Fortunately, with the help of John Allshouse, a NARA staff member at the Lenexa facility, we able to greatly simplify the process. Allshouse provided office in space within the Lenexa, Kansas, Federal Records Center to scan the film on site. We purchased a NextScan Eclipse 500, and trained NARA staff members to operate it. After several months of fine-tuning and experimentation to obtain adequate resolution consistently, we began production in March 2008 and completed scanning 2 million images in October 2008, processing the microfilm at an average rate of five reels per hour.

The scanner produced digital images of the census forms. To turn these images into usable data, we needed to process the data using optical mark recognition (OMR) software. We carried out the OMR processing at the Census Bureau National Processing Center. To get the images from Lenexa to Jeffersonville without compromising security, Glen Everhart made repeated trips to act as a courier, bringing disk drives loaded with images in his carry-on luggage.

We used Cardiff TeleForm optical mark recognition software. The Formtran Company developed customized

templates for each of the six forms used on the microfilm (the breaker sheets with geographic information, each of the four variations of the household panel, and the individual panel). TeleForm is designed for OMR on tests and surveys that use modern forms, and the low-resolution images from the 1960 census were challenging. Once again, we spent months fine-tuning the software and procedures before production began in March 2009. Despite our efforts, the OMR was much slower than anticipated. The original FOSDIC machine processed 100 frames per minute; a half-century later, we were able to achieve only 10 frames per minute. Part of the problem was that many images could not be deciphered by the software, and we had to resort to manual data entry for these records. The Census Bureau operators are highly skilled, however, and we eventually achieved adequate productivity by adding a second workstation and identifying for manual entry only those cases missing from the existing internal data. The OMR processing was completed in December 2010.

From Jeffersonville, Indiana, the data will now move to the Census Bureau Center for Economic Studies (CES) in Suitland, Maryland. We must accomplish four tasks to restore the 1960 data.

1. Recode and reformat the ASCII files to match the existing long-form records for the rest of the country.
2. Merge the new data into the existing file, eliminating any redundant records.
3. Create weights compatible with the existing weights in the 25-percent long-form sample data, based on the ratio of complete count to sample records for 44 population subgroups in each smallest weighting area (usually the census tract) (U.S. Census Bureau 1966, 21–24, 81).
4. Edit and allocate missing and inconsistent data.

We will do this work both at CES and at the Minnesota Census Research Data Center (MnRDC) at the MPC. The MnRDC opened in 2010 and is one of 10 secure remote facilities operated by the Census Bureau where qualified researchers with approved projects analyze restricted access data. MnRDC was conceived from the outset as a means of collaborating with the Census Bureau to develop new restricted access and public-use data sets and to improve existing data collections. Accordingly, the Census Bureau is making special accommodations to allow us to complete most of the work on the 1960 microdata within the MnRDC. These accommodations include installation of our 250,000-line data editing and allocation software on the MnRDC servers; installation of a C++ compiler so that we can make modifications to that software as needed; and providing access on MnRDC servers to a half-terabyte of scanned images of 1960 census forms. This collaboration and flexibility will substantially reduce the costs for completing the 1960 project.

Editing and allocation is the most challenging aspect of the work that remains. We are adapting the editing and allocation

software developed at the Minnesota Population Center for the censuses of 1850 through 1930 to meet the needs of the 1960 census. This will not always produce results identical to those obtained originally; the editing and allocation procedures used for the 1960 census were sharply constrained by the limited memory and speed of contemporary computers. We will conduct appropriate tests to ensure that using modern software for part of the file does not introduce significant comparability issues.

New Data Products

We plan three new data products:

1. A restricted-access long-form file including the full 25-percent long-form data with full geographic identification. This file will be made available to authorized researchers through the Census Bureau's national network of secure Research Data Centers.
2. A new 5-percent Public Use Microdata Sample with improved geographic identifiers. Building on the cartography of the National Historical Geographic Information System, we have designed a geographic system approximating the Census 2000 Public Use Microdata Areas. Our plan for the new sample must be approved by the Census Bureau Disclosure Review Board; until that review is complete, we cannot be certain how much geographic detail will be permitted in the public file.
3. A set of 1960 summary files at the tract and county levels. We are selecting the data elements most commonly requested by users of later census years and will design the categories to maximize comparability over time.

In addition to these data products, we are exploring an improved version of the existing 1-percent 1960 public-use data file that would incorporate identifiers for State Economic Areas (SEAs). SEAs were developed by Donald Bogue (1951) for use with the 1950 census, and they are available in all the IPUMS samples for the period from 1850 through 1950. If this plan proves practical, 1960 will serve as a crosswalk between the older census geographies and the recent systems.

We anticipate that the restricted-access 1960 file will be available through the Research Data Centers by December 2011. We plan to complete the public-use microdata samples and summary files by June 2013. The public files will be disseminated by the Census Bureau, IPUMS, and the National Historical Geographic Information System.

The new data products for 1960 will fill a critical gap in U.S. population data infrastructure. The 1960 census is presently the weak link in the series of public-use microdata spanning the twentieth century. The small size and limited geography in the existing 1960 public-use microdata sample precludes analysis of cities or metropolitan areas and makes multilevel analysis impossible. The need for new small-area summary data is equally great. The existing summary files

are missing a wide range of important tabulations that are consistently available for all censuses from 1970 onward, and the tables for 1960 census tracts do not distinguish any racial groups except for whites.

The 1960 census was taken at an extraordinary moment in the nation's demographic and economic history. It was just three years after the peak of the baby boom and one year after the peak of the marriage boom. As we seek to understand the sources of the ongoing transformation of fertility and marriage behavior, the 1960 census is an essential starting point. The 1960 census also provides a baseline for understanding the spectacular economic transformation of the late-twentieth century. The decade from 1959 to 1968 saw the largest increase in real per-capita domestic product of any 10-year period since World War II. To understand the social and behavioral consequences of the new affluence, we need high-quality data from the outset of the boom.

The 1960 census is also crucial for the study of seismic late twentieth-century shifts in such areas as race relations, inequality, and immigration. The modern civil rights movement had just begun, making the 1960 census a key point of reference for the study of racial inequality and segregation. This nationally representative data set predates the passage of landmark civil rights legislation in the mid-1960s. Also, across the population as a whole, income inequality was near an all-time low in 1960, and yet race and gender wage differentials were large. Because the 1960 census was the last census taken before the 1965 Immigration Act abolished national-origin quotas, it provides a benchmark for analysis of the late twentieth century boom in immigration.

Improved data for 1960 will be an invaluable resource for studying inequality, the transformation of industrial and occupational structure, family and household composition, life-course transitions to adulthood, the household economy, internal migration, nuptiality, fertility, and educational attainment. For each of these topics, the census can provide insights unavailable from any other source. Used in combination with data for other census years, data for 1960 will enable us to disentangle period and cohort changes in life-course processes and open exciting new opportunities for multilevel analyses in a key period of social and economic transition.

NOTES

The New Data Resources from the 1960 U.S. Census project at the Minnesota Population Center, University of Minnesota, is funded by the National Institutes of Health, Grant Number 5RO1HD056215-04.

1. Household-level sampling had been considered for the 1940 census, but it was rejected because it did not fit as well with "the established census procedures" and because of concern about the impact of clustering on sample efficiency (Stephan, Deming, and Hansen 1940, 620).

2. This percentage excludes 11 articles that did not specify sample density.

3. In addition, for many years the tracts from New Jersey were thought to have been lost, but now most of the New Jersey tracts have been recov-

ered. Some of the missing New Jersey tracts are available in the Elizabeth Mullen Bogue (1975) data set. Except for New Jersey, the 1960 Bogue file is redundant with information in the DUALabs tract file (U.S. Census Bureau 1971).

4. A directory of these data centers can be found at <http://www.census.gov/ces/main/contact.html> (U.S. Census Bureau 2011).

REFERENCES

- Adams, M. O., and T. E. Brown. 2000. Myths and realities about the 1960 Census. *APDU Newsletter* 22:8–9, 16.
- Bogue, D. 1951. *State economic areas*. U.S. Census Bureau. Washington, DC: Government Printing Office.
- Bogue, E. M. 1975. Census tract data, 1960: Elizabeth Mullen Bogue file [computer file]. ICPSR version, University of Chicago, Community and Family Study Center [producer]; Washington, DC: National Archives and Records Administration [distributor].
- Duncan, O. D. 1974. Developing social indicators. *Proceedings of the National Academy of Sciences* 71:5096–5102.
- Gardner, T. 2001. The National Historical Census Files Project. Paper presented at the Biennial Conference of Official Representatives of the Inter-university Consortium for Political and Social Research, Ann Arbor, Michigan. October 25–28.
- Haines, M. R., and Inter-university Consortium for Political and Social Research. 2005. Historical, demographic, economic, and social data: The United States, 1790–2000 [computer file]. ICPSR Study Number 2896-v2. Colgate University, Hamilton, NY, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.
- ICPSR. See Inter-university Consortium for Political and Social Research.
- Inter-university Consortium for Political and Social Research. 1973. Historical, demographic, economic, and social data: The United States, 1790–1970 [computer file]. ICPSR Study Number 3. Inter-university Consortium for Political and Social Research, Ann Arbor, MI.
- Ruggles, S., C. A. Fitch, and M. Sobek. 2000. The public use microdata samples of the U.S. Census: Research applications and privacy issues. Report prepared for the Task Force on the 2000 PUMS and the Census 2000 Users' Conference on PUMS, Alexandria, VA. May 22.
- Ruggles, S., M. Sobek, T. Alexander, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander. 2004. Integrated public use microdata series: Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center.
- Stephan, F. F., W. E. Deming, and M. H. Hansen. 1940. The sampling procedure of the 1940 population census. *Journal of the American Statistical Association* 35:615–30.
- University of California, Berkeley. 2006. Census rescue project. University of California, Berkeley. <http://ucdata.berkeley.edu:7101/projects/censusrescue/>.
- U.S. Census Bureau. 1963a. Evaluation and research program of the U.S. censuses of population and housing 1960: Background, procedures, and forms. Series ER60, No. 1. U.S. Census Bureau, Washington, DC.
- . 1963b. *U.S. censuses of population and housing: 1960. One-in-a-thousand sample description and technical documentation*. Washington, DC: Government Printing Office.
- . 1966. *Censuses of population and housing, 1960: Procedural history*. Washington, DC: Government Printing Office.
- . 1971. *Census tract-level data, 1960 [computer file]*. Washington, DC: DUALabs.
- . 1973. *Technical documentation for the 1960 public use microdata sample*. Washington, DC: Government Printing Office.
- . 2004. PhotoZone: Centennial celebration. U.S. Census Bureau Public Information Office. <http://www.census.gov/pubinfo/www/photos/centennial.html>.
- . 2011. U.S. Census Bureau: Center for Economic Studies (CES): Contact information. U.S. Census Bureau. <http://www.census.gov/ces/main/contact.html>.
- Weik, M. H. 1961. A third survey of domestic electronic digital computing systems. Ballistic Research Laboratories Report 1115, U.S. Department of the Army Project 5B03–06-002, U.S. Department of Defense, Aberdeen Proving Ground, MD.