16   U.S.A.

IPUMS-USA
Integrated Public Use Microdata Series for the United States

Steven Ruggles
Catherine A. Fitch
Patricia Kelly Hall
Matthew Sobek

The Integrated Public Use Microdata Series (IPUMS) is a high-precision individual-level database describing the characteristics of the U.S. population between 1850 and 1990.[1] It is the world's largest public-access individual-level database on a human population. We know simultaneously for every individual in the sample all of their personal and household characteristics. This makes the IPUMS a rich and powerful tool for exploring the social history of the United States.

We have worked for the past decade to create the database, document it and provide the innovative web-based

Steven Ruggles is Distinguished McKnight University Professor of History at the University of Minnesota and Director of the Minnesota Population Center. Catherine A. Fitch and Patricia Kelly Hall are Coordinators of the IPUMS project. Matthew Sobek is Managing Editor of the Historical Statistics of the United States, Millennial Edition, at the Center for Social and Economic Policy, University of California, Riverside.

For more detailed information on the IPUMS and for direct access to the data and documentation, see: http://www.ipums.org.

tools for electronic dissemination described in this article. The documentation and data extraction system are free to all users and can be accessed from the IPUMS homepage on the web at http://www.ipums.org.

Source Material

Preliminary samples now exist for all census years from 1850 to 1990, with the exception of 1890 (because census manuscripts were destroyed in a fire) and 1930 (which is unavailable until 2002 because of privacy restrictions). Samples of 1-in-100 or higher density are now available for three historic census years (1850, 1880 and 1920) as well as for every census from 1940 to 1990. Four nineteenth and early twentieth century census samples are available in preliminary form (1860, 1870, 1900, and 1910), but we are now expanding them; work on these samples will continue until 2004. We expect to create a sample of the 1930 census between 2002 and 2007. A description of these samples is presented in the Table 16–1.

Confidentiality Provisions

The confidentiality of individual information contained in the U.S. census is guaranteed by statute for 72 years from the date of enumeration. During the 72-year embargo, only designated employees of the Census Bureau have access to information, which contains individual names. Since 1960, the Bureau has created anonymized public use samples for social research and planning purposes. In addition to stripping names from sample records, any possibility of identifying individuals by their unique personal characteristics is precluded by masking geographic identifiers for those living in places with population totals below certain thresholds. Additional privacy protection is provided by the use of "topcodes" on sensitive questions such as income. Although the method of employing topcodes varies among censuses, the operating principle is the same: individual responses to these questions which fall above a certain value are suppressed; only the cut-off—or topcoded value—is repoorted for these individuals.

Table 16–1. Characteristics of National Census Microdata Files by Year of Creation

Census Year	Year Released	Sample Densities	Creator	Number of Records (thousands) Household	Person	Number of Variables	IPUMS File Size
1850	1994	1/100	Menard/Ruggles	37	198	92	79 Mb
1860	1999	1/200	Ruggles	13	71	94	28 Mb
1870	1998	1/500	Ruggles	16	86	94	34 Mb
1880	1990	1/1000	Ruggles	10	50	123	20 Mb
1880	1994	1/100	Ruggles	107	503	123	204 Mb
1900	1980	1/760	Preston	208	870	115	361 Mb
1910	1989	1/250	Preston	89	366	125	152 Mb
1920	1999	1/100	Ruggles	257	1,037	122	433 Mb
1940	1984	1/100	Winsborough/Census	391	1,351	174	584 Mb
1950	1984	1/100	Winsborough/Census	461	1,922	170	798 Mb
1960	1973	1/100	Census/DUALabs	579	1,780	141	790 Mb
1970	1972	(6) 1/100	Census Bureau	4,464	12,180	206	5,576 Mb
1980	1983	1/20, (2) 1/100	Census	6,595	15,871	276	7,526 Mb
1990	1992	1/20, 1/100	Census Bureau	6634	15,000	252	7,247 Mb
2000	2002	1/20	Census Bureau	c. 7000	c. 14,000	251	6,764 Mb
						TOTAL	32,400 Mb

In the 1980s, public use samples were constructed for the 1940 and 1950 census by Bureau employees under contract with social scientists from the University of Wisconsin. These data were also anonymized but had different size thresholds than the Bureau's samples.

The statutory confidentiality restrictions are still in effect for the 1930 census although we expect to begin construction of this sample when the 72-year confidentiality embargo expires in 2002.

The censuses for 1850 through 1920 are no longer protected by confidentiality restrictions. Thus, the historic samples for these years created by the Minnesota Historical Census Project and other researchers contain individual names and street addresses where shown on the original enumeration sheets.

Procedural History

Historical Sample Creation

When the Historical Census Project began in 1989, our main goal was to create large nationally representative samples of historical censuses. To date, we have entered data on some 2.5 million persons from the censuses of 1850, 1860, 1870, 1880, 1900, 1910 and 1920.

The samples for the period 1850 through 1920 are generally of 1-in-100 density. In addition, we are creating oversamples of some population subgroups of special interest. In the 1860 and 1870 samples, for example, black households are sampled at a density of 1-in-50 instead of 1-in-100. These are the years of transition from slavery to freedom for the great bulk of the black population and the 1870 census was the first to include complete information on the black population. Similarly, we are sampling American Indians at the rate of 1-in-10 for the census years 1900 and 1910, since these are the first censuses to include full information on the Indian population. In 1910, because we have a good existing sample at a 1-in-250 density created by Samuel Preston, we will end up with a overall sample density of 1-in-87. This high density is appropriate, since 1910 is the richest

census before 1940, especially for studies of immigration, nuptiality, fertility and mortality.

With the exception of the supplemental Preston sample for 1910, every one of the samples for the period 1850 through 1920 has an identical sample design and identical data processing procedures, including consistency checks, and verification procedures. We are now merging our data dictionaries for occupations, birthplaces, languages, family relationships, and institutions, so that in future versions of the database we can ensure that identical responses receive identical codes. The use of common software, procedures, and data dictionaries means that we can be confident that any differences in time periods reflect true differences in the source material. There are, to be sure, variations in the quality of the underlying manuscript data; the U.S. censuses improved dramatically over the period 1850 through 1920. Nonetheless, the consistent treatment of the manuscript data at least ensures that we have not introduced any additional noise.

A key feature of the samples constructed by the Historical Census Project is their use of a common coding system for all variables. For example, researchers do not have to recode hundreds of birthplace values in order to study migration at the turn of the century; a response of "Ireland" in the birthplace variable has been given the same numeric code in 1880, 1900, 1910 and 1920. This has meant that users can run the identical analysis on multiple census years with no additional work, and has led to a substantial body of multi-year studies (see the Bibliography at the end of this chapter).

The Census Bureau Samples

The first public use census microdata sample in the world was created as a byproduct of the 1960 census (U.S. Bureau of the Census 1964). In an effort to meet the needs of scholars who increasingly requested specialized tabulations, the Census Bureau created a 1-in-1000 extract from the basic data tapes they had used to make tabulations for the published census volumes. To preserve confidentiality, the Census Bureau removed names, addresses, and other potentially identifying information.

The 1960 public use sample was an immediate success. Not only did it allow researchers to make tabulations tailored to their specific research questions, but it also enabled them to apply new methods to the analysis of census data, especially multivariate techniques. But the sample did have two significant limitations. First, the sample size was relatively small. The 1-in-1000 sample density yielded about 180,000 person records. Given the modest capacity of computers in 1964, this was a lot of cases; but as researchers began to use the sample for detailed analysis of small population subgroups, its limitations became apparent. Second, the 1960 public use sample provided highly limited geographic information. In its zeal to preserve 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.

Both of these problems were addressed by the 1970 public use samples. The 1-in-1000 density of the 1960 sample was increased dramatically; the Census Bureau provided six independent public use samples for 1970, each of which had a 1-in-100 density. Users who required an exceptionally large number of cases could combine the samples to obtain a six-percent density, or about 12 million person records. In addition, the 1970 samples provided a variety of alternate geographic codes, although the Census Bureau still did not identify any places of less than 250,000 population.

In conjunction with the 1970 public use samples, the Census Bureau released a new version of the 1960 public use sample. They enlarged the sample density from 1-in-1000 to 1-in-100, and at the same time reorganized the coding schemes and record layouts to be compatible with the samples from 1970. This compatibility made it relatively easy for investigators to pool data from 1960 and 1970, and thus incorporate change into their analyses.

By the late 1970s, the public use samples for 1960 and 1970 had become one of the essential tools of American social scientists. It was in this climate that Halliman Winsborough and a group of other researchers at the University of Wisconsin developed the idea of creating historical public use samples for

earlier census years. They obtained funding from the National Science Foundation and contracted with the Census Bureau to create 1-in-100 samples for the censuses of 1940 and 1950 (U.S. Bureau of the Census 1984a, 1984b).

The Census Bureau also released public use samples for the 1980 and 1990 censuses. These samples include considerably greater geographic and subject content detail than either the 1960 or 1970 public use samples. Therefore, we now have a continuous series of Census Bureau microdata samples for six census years consisting of anonymized records spanning the period from 1940 through 1990. The Bureau will produce a seventh sample for the 2000 census soon.

The Integrated Public Use Microdata Series

Unfortunately, the samples for the 1940, 1950, 1980 and 1990 censuses broke with the precedent of 1960 and 1970: each sample was coded entirely differently and had separate documentation, which made it difficult to use more than one sample at a time. Therefore, in 1992 we proposed to develop a common format and documentation for both the Minnesota samples and the Census Bureau samples, as well as two small historical samples created by Samuel Preston for 1900 and 1910.

The first version of this database—the Integrated Public Use Microdata Series—was completed in 1993. To code variables into a common format without losing information, we developed a composite coding system. The first one or two digits of the code provide information available across all samples. The next one or two digits provide additional information available in a broad subset of samples. Finally, where needed to preserve all original census information, trailing digits provide detail only rarely available. In this fashion, we supply a lowest common denominator across all samples and yet retain full detail.

In addition to common codes, the IPUMS created common constructed variables. Especially popular have been the family interrelationship variables, which are "pointers" allowing researchers to link husbands and wives and parents and children without resorting to higher-level programming. We also provide constructed variables describing the composition of each unit,

measures of socioeconomic status, and a variety of constructed geographic variables to aid comparison across time.

An important feature of the IPUMS is the allocation of virtually all missing and illegible data. We have used logical editing and "hot deck" probabilistic editing procedures to allocate missing, illegible and inconsistent data in all the censuses from 1850 to 1920. The Census Bureau used the same allocation techniques for the period since 1940. Allocation of missing and inconsistent data significantly increases the precision of sample estimates and makes the samples easier to use. Data quality flags identify the allocations so that a researcher can choose not to use them. We have documented the missing data allocation procedures for all census years. For 1980 and 1990, the Census Bureau never published the allocation procedures, but we were able to reconstruct them from a series of internal memos provided to us by the Bureau.[2]

By connecting the historical samples to the modern ones by means of the IPUMS, we have made the historical samples more widely used. Even the nineteenth century material is widely used for policy-relevant research, because it is highly compatible with the late twentieth-century material. This in turn has made it easier to persuade government agencies that the historical work is worth funding.

Variable Availability

The census has always contained certain core questions that are generally comparable over the entire time span of the database. Other questions have come and gone. Table 16–2 describes many of the subject areas covered by the census since 1850. Several of these topic areas correspond to multiple variables in the database. Subjects such as personal income and migration status are confined to the second half of the twentieth century. Available geographic detail, on the other hand, is actually superior in the historical samples, which did not have to

---

[2] IPUMS-2000 Volume 3, *Counting the Past*, documents the allocation procedures we employed for the period 1850 to 1920 and describes the methods used by the Census Bureau from 1940 to 1990.

suppress such information to ensure confidentiality.[3] In fact, the detail available for most variables is typically superior in the pre-1940 historical samples, which sought to retain as much information as possible. Most of the samples prior to 1940 even include the names of the respondents and in some years their street address. Despite containing names, however, the samples are of little use for genealogy, since they contain only one percent of the population.

Documentation

We believe one of the most valuable contribution of the IPUMS is the documentation. The IPUMS provides researchers with comprehensive documentation for every phase of analysis when using these data, with standards at or exceeding those of the U.S. Census Bureau. IPUMS documentation is not confined to codebooks describing the labels associated with each census category. We provide a wide variety of ancillary information to aid in the interpretation of the data, including enumerator instructions, details of sample designs and sampling errors, procedural histories of each census, descriptions of data-entry procedures, and documentation of error correction and other post-enumeration processing. Documentary material not readily available in text form, such as census forms and maps, is easily accessible in graphic images on the web site.

Rapid growth in the technology available for electronic delivery of information has allowed us to expand the IPUMS

---

[3] Available geographic detail varies widely among the modern samples subject to confidentiality rules. The smallest unit available in 1940 and 1950 is State Economic Area: groups of counties within states that shared a similar economic base. The IPUMS replicates these groupings in the samples before 1940. The sample with the lowest level of geographic detail is 1960, which provides no information below the level of state). 1970 provides "county groups" which combine counties into units that meet a 250,000 population threshold. 1980 and 1990 define geographic units of varying composition that have at least 100,000 population. As long as the units have the requisite population in 1980 and 1990, the geographic units can be as precise as sections of cities or combinations of specific communities within counties.

Table 16–2. Availability of Select Subject Areas in the IPUMS, 1850-1990

	1850	1860	1870	1880	1900	1910	1920	1940	1950	1960	1970	1980	1990
Household Record													
State	X	X	X	X	X	X	X	X	X	X	X	X	X
County	X	X	X	X	X	X	X	·	·	·	·	·	·
Metropolitan area	X	X	X	X	X	X	X	X	X	·	X	X	X
City	X	X	X	X	X	X	X	X	X	X	·	X	X
Size of place	X	X	X	X	X	X	X	X	X	·	·	X	X
Urban/rural status	X	X	X	X	X	X	X	X	X	·	·	X	X
Farm	X	X	X	X	X	X	X	X	X	X	X	X	X
Ownership of dwelling	·	·	·	·	X	X	X	X	·	X	·	X	X
Mortgage status	·	·	·	·	X	X	X	·	·	·	·	X	X
Value of house or property	·	·	·	·	·	·	·	X	·	X	X	X	X
Monthly rent	·	·	·	·	·	·	·	X	·	X	X	X	X
Total family income	·	·	·	·	·	·	·	·	X	X	X	X	X
Person Record													
Relationship to household head	·	·	·	X	X	X	X	X	X	X	X	X	X
Age	X	X	X	X	X	X	X	X	X	X	X	X	X

Note: X = available in that census year.

Table 16–2.   **Availability of Select Subject Areas in the IPUMS, 1850-1990**   (continued)

	1850	1860	1870	1880	1900	1910	1920	1940	1950	1960	1970	1980	1990
Person Record (continued)													
Sex	X	X	X	X	X	X	X	X	X	X	X	X	X
Race	X	X	X	X	X	X	X	X	X	X	X	X	X
Marital status	·	·	·	X	X	X	X	X	X	X	X	X	X
Age at first marriage	·	·	·	·	·	·	·	X	·	X	X	X	X
Times married	·	·	·	·	·	X	X	X	X	X	X	X	·
Children ever born	·	·	·	X	X	X	·	X	X	X	X	X	X
Birthplace	X	X	X	X	X	X	X	X	X	X	X	X	X
Parents' birthplaces	·	·	·	X	X	X	X	X	X	X	X	X	·
Years in the United States	·	·	·	·	X	X	X	·	·	X	X	·	X
School attendance	X	X	X	X	X	X	X	X	X	·	X	X	X
Educational attainment	·	·	·	·	·	·	·	X	X	X	X	X	X
Literacy	X	X	X	X	X	X	X	·	·	·	·	·	·
Occupation	X	X	X	X	X	X	X	X	X	X	X	X	X
Industry	·	·	·	·	·	X	X	X	X	X	X	X	X
Wage and salary income	·	·	·	·	·	·	·	X	X	X	X	X	X
Migration status	·	·	·	·	·	X	·	X	X	X	X	X	X
Veteran status	·	·	·	·	X	X	·	X	X	X	X	X	·
Name	X	X	X	X	X	X	X	·	·	·	·	·	·

Note: X = available in that census year.

documentation while still giving researchers navigation tools to move through it efficiently. Hypertext links throughout the site allow nonlinear access to information based on the user's needs and interests. A full-text search engine enables users to find a specific item quickly and easily.

Variable descriptions form the core of the IPUMS documentation. They explain how the variable may have changed across time and include an availability table showing which samples contain the variable. Universe statements identify characteristics of the households or persons whose responses are reported for that variable. Every variable discussion includes links to documents that explain how the data were originally collected—such as census forms, census questions and enumerator instructions—as well as how they were transformed into microdata—sample design, data-entry procedures, and editing and allocation protocols. These links enable researchers to access documentation at varying levels of complexity depending on their needs.

Every variable in the IPUMS has an extensive comparability discussion in the documentation that points out consistencies and inconsistencies across all samples in the database. Important differences are highlighted, along with warnings about likely errors and strategies for enhancing compatibility for specific comparisons. In some cases, entire essays are necessary. All this material is presented in a hypertext documentation system that allows users to locate relevant information quickly and easily.

Whereas the Internet made distribution of data files on physical media unnecessary, the World Wide Web made it possible to present the documentation on-line in electronic form. Researchers can now navigate the documentation without poring through multiple paper volumes.

Value Added Over the Original Census Samples

Much of the value of the IPUMS for historical and long-term analysis arises from compatible variable treatment and integrated documentation across census years. The techniques we used to achieve this compatibility are described in two special issues of *Historical Methods* (1995, 1999). Beyond the

issues of compatibility, however, the IPUMS offers a number of entirely new features. Most of these are minor conveniences, but others are more substantial. The paragraphs that follow describe some of the most important reasons why the IPUMS is easier to use than any previous census microdata.

The IPUMS retroactively identifies metropolitan areas in the database in every census year applying the criteria used in 1950, when the concept was first employed by the census. We also have for the first time identified the county composition of every metropolitan area from 1850 to 1990, and included all maps associated with the varying geographic divisions identified in the original samples.

We include a set of consistently constructed family interrelationship pointer variables for all years; they greatly aid in the design of family-related measures. Of the original samples, only 1910 had such a feature, and the links were not always consistently made, since in many cases they were performed by hand rather than computer. With researchers using the same set of interrelationship variables, the IPUMS removes the chance that results of two researchers may simply be an artifact of two different linking procedures.

We emphasize the utility of the IPUMS for historical applications, and indeed that was foremost in our minds when we envisioned the project. But researchers interested in only the most recent census years—even only 1990—have much to gain by using the IPUMS. We have corrected dozens of small errors in the Census Bureau samples and documentation, and added convenient features such as an unweighted version of the 1990 sample.

IPUMS geographic codes are much easier to use than those provided by the Census Bureau. For example, the 1980 and 1990 samples do not identify city of residence. Users are forced to recode a complex seven-digit code to reconstitute cities (five digits in 1980) and the Bureau does not even document which cities are fully identifiable by means of this cumbersome procedure. The IPUMS includes a consistent code for city across all samples, and we use it to identify city of residence five years ago and city of work in addition to city of current residence. We

also discovered a way to identify urban-rural status in the 1990 sample, which otherwise lacked this basic variable.

A second example of the advantages of the IPUMS over the Census Bureau samples concerns the family and poverty variables. The 1970, 1980 and 1990 Census Bureau samples do not identify families unrelated to the householder, such as families composed of boarders or lodgers. Moreover, the 1970 and 1980 samples fail to identify most subfamilies, owing to a programming error. Our family interrelationship variables are a significant improvement in both these areas. The failure to identify families unrelated to the householder also led to a major flaw in the 1990 poverty variable. Children unrelated to the householder were simply ignored for poverty purposes in the 1990 sample; there are no unrelated children in poverty. In addition, the parent of such a child was assigned the poverty status (percent of poverty level) for a single adult, not a family of two or more. So the original 1990 PUMS greatly understates poverty for a group of persons who are of particular interest to policy analysts. The IPUMS rectifies these deficiencies.[4]

Electronic Dissemination

The timing of the IPUMS project coincided with a rapidly changing technological environment. The Internet was already fully functioning at the time of our initial preliminary data release in late 1993. Because it allowed us to make data directly accessible to researchers without delays, it has been our only means of data delivery from the beginning of the project.

With the release of the IPUMS, we had compiled an impressively large data series with an equally imposing body of documentation. Unfortunately, the great size of the files and the scale of the documentation still posed a barrier to usage. To some degree, the IPUMS was in danger of being the exclusive

---

[4] A poverty variable calculated with the 1990 formulas and adjusted for the inflation is included in the data. However, since there is considerable scholarly debate about the appropriateness of using the same foodbasket and multiplier assumptions across time, we urge users to use this variable with caution while we evaluate other possible measures to incorporate in the data.

province of researchers associated with large data centers or who were otherwise rich in terms of computing resources.

Technological developments, however, presented an opportunity to circumvent a serious potential bottleneck. In the course of our work on IPUMS we developed a rudimentary but effective system for extracting only select variables and cases from the database. The system could reduce multiple gigabytes of data to whatever size file was necessary and manageable for a given research project. It was clear that such a system was a logical complement to the IPUMS that would greatly expand its accessibility to researchers around the world.[5]

Data Access: The IPUMS Extract System

IPUMS data are free and are delivered by means of an interactive data extract system available from the IPUMS website at http://www.ipums.org. The IPUMS data extraction system allows researchers to select only those subpopulations and variables needed for a particular analysis. The extraction system creates a single data file containing multiple census years with identical custom record layouts. This system allows researchers to fashion smaller extracts of the data specifically oriented to their own research agenda and suited to their available computing power and storage capacity. In a simple four-step process, users select the samples, density, variables and case selection criteria most appropriate for their research. When the extract is ready, an email message notifies them that their request has been filled and directs them to a web page from which they can download their files. New features of the system allow users to recall previous extract requests and make modifications. Extract-specific SPSS, SAS, or Stata command files facilitate the process of moving the data extract into some of the more common statistical software packages.

---

[5] In recognition of the scientific power of the internet, the National Institutes of Health and the National Science Foundation have both supported further development of the electronic dissemination of the IPUMS.

The combination of free and open access, a user-friendly access system, and integrated comprehensive hypertext documentation has attracted many users. Since 1995, we have distributed over two terabytes of IPUMS data to users around the world. Our automated data-extraction system has prepared approximately 10,000 custom extracts of IPUMS data since May 1996, and is now processing approximately 500 data extract requests per month.

Research Possibilities, Expert Users, and Publications Using the IPUMS

Among key research areas that can be studied using the IPUMS are economic development, poverty and inequality, industrial and occupational structure, household and family composition, the household economy, female labor force participation, employment patterns, population growth, urbanization, internal migration, immigration, nuptiality, fertility, and education.

In a very brief period, the IPUMS has become one of the most widely used databases in American social science and has contributed significantly to the national and international visibility of the University of Minnesota. Even though the database has only been widely available for five years, there is already a substantial body of IPUMS-based research. To date, the database has been used in 87 articles, 3 books, and 25 Ph.D. dissertations as well as hundreds of conference papers and research reports. Many of these articles appeared in leading journals such as the *American Economic Review*, the *American Sociological Review*, the *American Historical Review*, *Social Forces* and *Demography* (see bibliography).

Data Expansion

The 1880 LDS Database

Volunteers working with the Church of Jesus Christ of Latter-Day Saints (LDS) have invested approximately two million hours transcribing information from the 1880 U.S. Census of Population. This database—which includes the entire

U.S. population enumerated in that census—has the potential to become our most important resource for the study of the economic and social organization of late-nineteenth century American society. The LDS holds the copyright to these data but has agreed to make them freely available for academic use in exchange for modest assistance in cleaning the data. The late nineteenth century is a critical period in the study of fertility decline, urbanization, immigration, household composition, and occupational structure.

The LDS 1880 database includes a wealth of information on these topics that can only be fully explored through the creation of a new microdata set. The complete 1880 database provides the necessary cases to study the demography and social structure of dispersed population subgroups and small localities. The database will also allow the construction of cross-tabulations on a wide range of topics that were not covered by census publications or were incompletely tabulated, filling a gap in aggregate data series for counties and cities. Perhaps even more important is the potential for longitudinal and multi-level, multivariate analyses opened up by the availability of the database. The 1880 census database will not only constitute an invaluable resource in its own right, but will also enhance the value of the previously created historical microdata samples; used in combination these microdata will constitute our most important resource for the study of nineteenth-century social structure.

Additional Census Samples and Other Data

The two remaining gaps in the IPUMS' twentieth century census data are for 1930 and 2000. We expect to begin work on construction of a 1930 sample in 2002 when the 72-year embargo period on confidentiality expires. The new IPUMS-International project (See McCaa and Ruggles, Chapter 20) includes funding to incorporate the public use sample of the 2000 census as soon as the tapes are available.

Data from other sources beyond the decennial census will also be added as part of the IPUMS International project. The monthly Current Population Surveys from 1964 through 2003 will be harmonized and incorporated with the IPUMS. In

addition we will add the American Community Survey, the Census Bureau's new annual demographic snapshot of the American population, from 2000 through 2003.

An Invitation to New Users

The IPUMS database and documentation aim to simplify use of census samples in combination, while the data extraction system allows users to create customized datasets with minimal effort. As we continue to add features to the extraction system, we seek to circumvent any need for programming knowledge, except for the most sophisticated analyses.

We are historians and therefore recognize that time is the most important variable. One of our missions at the Historical Census Project is to demonstrate to historians who have traditionally not used quantitative evidence in their research that the census—the people's data—is a rich source for interpreting the past. With the help of new technology, we have overcome most of the data access and management obstacles that have been such an enormous barrier to researchers not specifically trained to use quantitative evidence. With the IPUMS, we have made census data easy to understand, easy to use in a time series and accessible from virtually anywhere in the world. We invite new users to visit our web site, register on-line, and explore 150 years of social history in the United States.

Bibliography

The number of publications using or about the IPUMS continues to grow. For a current listing of these publications, see the MPC web site at: http://www.ipums.org.

Publications about the IPUMS

Block, William C and Dianne L. Star. 1995. "Data Entry and Verification." *Historical Methods* 28: 63-65.
Dillon, Lisa Y. 1996. "Integrating Nineteenth-Century Canadian and American Census Data Sets." *Computing in the Humanities* 30: 381-392.

Ferrie, Joseph P. 1996. "A New Sample of Males Linked from the Public-Use-Microdata-Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules." *Historical Methods* 29: 141-156.

Gardner, Todd. 1995. "Creation of the Public Use Census Files for 1850, 1880, and 1920: Software Development." *Historical Methods* 28: 59-62.

Gardner, Todd, Matthew Sobek and Steven Ruggles. 1999. "IPUMS Data Extraction System." *Historical Methods 32*: (3) 119-124.

Goeken, Ronald, Marjorie Bryer and Cassandra Lucas. 1999. "Making Sense of Census Responses: Coding Complex Variables in the 1920 PUMS." *Historical Methods 32*: (3) 134-138.

Goeken, Ron and Matthew Mulcahy. 1995. "Geographic Coding." *Historical Methods* 28: 74-76.

Gutmann, Myron P. and Douglas Ewbank. 1999. "The 1910 Black and Hispanic Oversamples." *Historical Methods 32:* (3) 156-158.

Gregory, J. 1995. "The Southern Diaspora and the Urban Dispossessed - Demonstrating the Census Public Use Microdata Samples." *Journal of American History 82*: 111-134.

Hacker, J. David, Steven Ruggles, Andrea R. Foroughi and Walter L. Sargent. 1999. "Public Use Microdata Samples of the 1860 & 1870 U.S. Census of Population." *Historical Methods 32: (3)* 125-133.

Hall, Patricia Kelly, Catherine Fitch, Margot Canaday, Lisa Ebeltoft-Kraske, Carrie Ronnander and Kathleen M. Thomas. 1999. "IPUMS Metadata: Documenting 150 Years of Census Microdata." *Historical Methods 32:* (3) 111-118.

Ito RM, Gratton B, Wycoff J. 1997. "Using the 1940 and 1950 Public Use Microdata Samples - A Cautionary Tale." *Historical Methods* 30: 137-145.

Kallgren, Daniel and David Beck Ryden. 1995. "Data Consistency Checking." *Historical Methods* 28: 66-69.

Magnuson, Diana L. 1995. "Population Schedule Content." *Historical Methods* 28: 11-26.

Magnuson, Diana L. and Miriam L. King. 1995. "Enumeration Procedures." *Historical Methods* 28: 27-32.

Ronnander, Chad. 1999. "The Classification of Work: Applying 1950 Census Occupation and Industry Codes to 1920 Responses." *Historical Methods 32*: (3) 151-155.

Ruggles, Steven et al. 1995. *The Minnesota Historical Census Projects.* Winter Special Issue, *Historical Methods* 28.

Ruggles, Steven and Patricia Kelly Hall, eds. 1999*. IPUMS: Integrated Public Use Microdata Series,* Summer Special Issue, *Historical Methods 32.*

Ruggles, Steven, Matthew Sobek and Todd Gardner. 1996. "Distributing Large Historical Census Samples on the Internet." *History and Computing 9*: 145-159.

_____. 1996. "Disseminating Historical Census Data on the World Wide Web." *Iassist Quarterly 20*: 4-18.

Ruggles, Steven and Russell R. Menard. 1995. "The Minnesota Historical Census Projects." *Historical Methods 28*: 6-10.

Ruggles, Steven, J. David Hacker and Matthew Sobek, 1995. "Overview of the Integrated Public Use Microdata Series." *Historical Methods 28*: 33-39.

Ruggles, Steven. 1995. "Sample Designs and Sampling Errors in the Integrated Public Use Microdata Series." *Historical Methods 28*: 40-46.

_____. 1995. "Family Interrelationships." *Historical Methods 28*: 52-58.

_____. 1993. "Historical Demography from the Census: Applications of the American Census Microdata Files." In David Reher and Roger Schofield, eds., *Old and New Methods in Historical Demography* (Oxford University Press), 383-393.

_____. 1991. "The U.S. Public Use Microdata Files as a Source for the Study of Long-term Social Change." *IASSIST Quarterly 15*: 20-27.

_____. 1991. "Comparability of the Public Use Files of the U.S. Census of Population, 1880-1980." *Social Science History 15*: 123-158.

_____. 1991. "Integration of the Public Use Files of the U.S. Census of Population, 1880-1980." *American Statistical Association Proceedings of the Social Statistics Section* (American Statistical Association) 365-370.

Sobek, Matthew. 1991. "Class Analysis and the U.S. Census Public Use Samples." *Historical Methods 24*: 171-181.

_____. 1995. "Occupation and Income Scores." *Historical Methods 28*: 47-51.

Sobek, Matthew and Lisa Dillon. 1995. "Occupational Coding." *Historical Methods 28*: 70-73.

Sobek, Matthew and Steven Ruggles. 1999. "The IPUMS Project: An Update." *Historical Methods 32:* (3) 102-110.

Published Research Using the IPUMS Database

Alyan, N. 1999. "The Role of Capital Intensity and Technology Usage in Upgrading Skills in the US labor Market*." Technological Forecasting and Social Change 61:* 59-74.

Carter, Susan B. and Richard Sutch. 1996. "Fixing the Facts, Editing of the 1880 United States Census of Occupations: With Implications for Long-Term Labor-Force Trends and the Sociology of Official Statistics." *Historical Methods 29*: 5-24.

_____. 1996. "Myth of the Industrial Scrap Heap—A Revisionist View of Turn-of-the-Century American Retirement." *Journal of Economic History 56*: 5-38.

Cattan, Peter and Guillermo Grenier. Forthcoming. "Latino Immigrants in the Labor Force: Trends and Labor Market Issues." In National Council of La Raza's *State of Hispanic America*.

Costa, Dora L. 1998. *The Evolution of Retirement: An American Economic History, 1880-1990.* Chicago: University of Chicago Press.

_____. 1999. "A House of Her Own: Old Age Assistance and the Living Arrangements of Older Nonmarried Women." *Journal of Public Economics* 72: (1) 39-59.

_____. 1998. "Unequal at Birth: A Long-term Comparison of Income and Birth Weight." *Journal of Economic History* 58: (4) 987-1009.

_____. 1997. "Displacing the Family: Union Army Pensions and Elderly Living Arrangements." *Journal of Political Economy* 105: 1269-92.

Darity, William A. 1998. "Intergroup Disparity: Economic Theory and Social Science Evidence." *Southern Economic Journal* 64: 805-826.

Darity William A., Jason, Dietrich, David K. Guilkey. 1997. "Racial and Ethnic Inequality in the United States: A Secular Perspective." *American Economic Review* 87: 301-305.

Dillon, Lisa Y. Forthcoming. "Women and the Dynamics of Marriage, Household Status, and Aging in Victorian Canada and the United States." *The History of the Family, An International Quarterly,* Special Issue on Canadian Family History.

_____. 1998. "Parent-Child Co-Residence Among the Elderly in Victorian Canada and the United States: A Comparative Study." In Edgar-André Montigny and Lori Chambers, eds. *Family Matters: Papers in Post-Confederation Canadian Family History.* Toronto: Canadian Scholars Press.

Ellis, M. M. Reibel, and R. Wright. 1999. "Comparative Metropolitan Area Analysis: Matching the 1980 and 1990 Census Public Use Microdata Samples for Metropolitan Areas." *Urban Geography* 20: (1) 75-92.

Elman Cheryl. 1999. "Labor Markets and Opportunity Structures - Modeling the Context of Retirement in the Historical United States" *Research on Aging 21*: (2) 205-239.

_____. 1998. "Intergenerational Household Structure and Economic Change at the Turn of the Twentieth Century." *Journal of Family History 4:* 417-440.

Elman Cheryl and G. Myers. 1997. "Age- and Sex-Differentials in Morbidity at the Start of an Epidemiological Transition: Returns from the 1880 US Census" *Social Science & Medicine 6*: 943-956.

Fairlie R.W. and W.A. Sundstrom. 1999. "The Emergence, Persistence, and Recent Widening of the Racial Unemployment Gap." *Industrial and Labor Relations Review 52*: (2) 252-270.

Ferrie, Joseph. 1997. "The Entry Into the U.S. Labor Market of Antebellum European Immigrants, 1840-60," *Explorations in Economic History 34*: 295-330.

_____. 1995. "The Geographic Mobility of Antebellum European Immigrants to the U.S. after their Arrival At New York, 1840-60" *Research in Economic History 15*: 99-148

_____. 1995. "Up and Out or Down and Out? The Occupational Mobility of Immigrant Non-Persisters in the Nineteenth Century U.S." *Journal of Interdisciplinary History 26* (Summer): 33-55.

_____. Joseph. 1999. *'Yankeys Now': European Immigrants in the Antebellum U.S., 1840-60* (New York: Oxford University Press).

_____. 1994. "The Wealth Accumulation of Antebellum European Immigrants to the U.S., 1840-60" *Journal of Economic History 54*: 1-33.

Friedberg L. 1999. "The effect of old age assistance on retirement." *Journal of Public Economics 71*: (2) 213-232.

Gardner, Todd. Forthcoming. "Suburbanization in the United States 1850-1940." *Journal of Urban History.*

_____. 1999. "Metropolitan Classification for Census Years Before World War II." *Historical Methods 32:* 139-150.

Geib-Gundersen, Lisa and Elizabeth Zahrt. 1996. "A New Look at U.S. Agricultural Productivity Growth, 1800-1910." *The Journal of Economic History 56*: 679-687.

Gisser M. and A. Davila. 1998. "Do Farm Workers Earn Less? An Analysis of the Farm Labor Problem." *American Journal of Agricultural Economics 80*: 669-682.

Gutmann, Myron P. W.P. Frisbie, and K.S. Blanchard. 1999. "A New Look at the Hispanic Population of the United States in 1910." *Historical Methods 32*: (1) 5-19.

Hacker, J. David. 1999. "Child Naming, Religion, and the Decline of Marital Fertility in Nineteenth-Century America." *The History of the Family: An International Quarterly*, 4: (3) 339-65.

Hill, Mark E. Forthcoming. "Multivariate Survivorship Analysis Using Two Cross-Sectional Samples." *Demography*.

Irwin J.R. O'Brien A.P. 1998. "Where have all the sharecroppers gone? Black occupations in postbellum Mississippi." *Agricultural History 72*: 280-297.

Lee C.H. 1998. "Long-Term Unemployment and Retirement in Early-Twentieth-Century America." *Journal of Economic History 58*: 844-856.

Kallgren, Dan. 1999. "Place, Race and Poverty in the Pattern of Southern School Attendance" *Historical Geography* 27.

Krueger, Kurt V. 1999. *Healthy Life Expectancy, 1996*. Shawnee Mission, KS: Expectancy Data.

Maloney, Thomas N. Forthcoming. "Personnel Policy and Racial Inequality in the Pre-World War II North." *Journal of Interdisciplinary History*.

McCaa, Robert. 1993. "Gender in the Melting Pot: Marital Assimilation in New York City, 1900-1980." *Journal of Interdisciplinary History 24*: 207-231.

Moehling, C.M. 1999. "State Child Labor Laws and the Decline of Child Labor." *Explorations in Economic History 36*: 72-106.

Preston, Samuel H., M.E. Hill and G.L. Drevenstedt. 1998. "Childhood Conditions that Predict Survival to Advanced Ages Among African-Americans." *Social Science and Medicine 47*: 1231-1246.

Ruggles, Steven. 1997. "The Rise of Divorce and Separation in the United States, 1880-1990." *Demography 34*: 455-466.

_____. 1997. "Reply to Oppenheimer and Preston." *Demography 34*: 475-479.

_____. 1997. "The Effects of AFDC on American Family Structure, 1940-1990." *Journal of Family History 22*: 307-25

_____. 1996. "The Effects of Demographic Change on Multigenerational Family Structure: United States Whites 1880-1980." In Alain Bideau, A. Perrenoud, K.A. Lynch and G. Brunet, eds., *Les systèmes demographiques du passé* (Lyons: Centre Jacques Cartier), 21-40.

_____. 1996. "Living Arrangements of the Elderly in America, 1880-1980." In Tamara K. Hareven, ed., *Aging and Generational Relations Over the Life Course: A Historical and Cross-Cultural Perspective* (Berlin: de Gruyter), 254-271.

_____. 1994. "The Transformation of American Family Structure." *American Historical Review 99*: 103-128.

_____. 1994. "The Origins of African-American Family Structure." *American Sociological Review 59:* 136-151.

_____. 1993. "Confessions of a Microsimulator: Problems in Modeling the Demography of Kinship." *Historical Methods 26*: 161-169.

Ruggles, Steven, and Ronald Goeken. 1992. "Race and Multigenerational Family Structure, 1900-1980." In Scott J. South & Stuart E. Tolnay, eds., *The Changing American Family: Sociological & Demographic Perspectives* (Westview Press), 15-42.

Schoeni Robert F. 1998. "Reassessing the decline in parent-child old-age coresidence during the twentieth century. *Demography 35*: 307-313.

Shoemaker, Nancy. 1999. *American Indian population recovery in the Twentieth Century* (Albuquerque: University of New Mexico Press).

Sobek, Matthew. 1996. "Work, Status, and Income: Men in the American Occupational Structure since the Late Nineteenth Century." *Social Science History 20:* 169-207.

Thorvaldsen, Gunnar. 1998. "Marriage and Names among Immigrants to Minnesota." *The Journal of the Association for History and Computing* 1(2).

Tolnay, Stewart E. 1996. "Structural Change and Fertility Change in the South, 1910 to 1940." *Social Science Quarterly 77*: 559-576.

_____. 1997. "The Great Migration and Changes in the Northern Black Family, 1940 to 1990." *Social Forces 75:* 1213-1238.

_____. 1998. "Migration Experience and Family Patterns in the 'Promised Land'" *Journal of Family History 23*: 68-89.

_____. 1998. "Educational Selection in the Migration of Southern Blacks, 1880-1990." *Social Forces, 77*: (2) 487-514.

_____. 1999. *The Bottom Rung: African American Family Life on Southern Farms.* Urbana: University of Illinois Press, 1999.

Tolnay, Stewart E. and K.D. Crowder. 1999. "Regional origin and family stability in northern cities: The role of context." *American Sociological Review 64*: (1) 97-112.

Weiss, T. 1999. "Estimates of white and nonwhite gainful workers in the United States by age group, race, and sex - Decennial census years, 1800-1900" *Historical Methods 32*: (1) 21-35.

Wyly, E.K. 1998. "Containment and mismatch: Gender differences in commuting in metropolitan labor markets." *Urban Geography 19*: 395-430.

Dissertations (completed and in-progress)

Alexander, Trent. "Great Migrations: Race in the Southern Move to the Industrial Midwest, 1920-70." (Dissertation in progress at Carnegie Mellon University, Department of History).

Bartley, Alan. "A Valuation of Specific Crime Rates." (Dissertation in progress at Vanderbilt University).

Brown, Kenneth R. "Drunk on the Eve of Reconstruction: Faith-Based Economic Development and Entreprenuership, the Gospel According to the Hartford Avenue Baptist Church of Detroit, Michigan." (Dissertation in progress at the Unviersity of Michigan).

Bures, Regina M. 1998. "Residential Mobility, Migration, and Life-Course Change: A Study of Family, Work, and Mobility in Later Mid-Life." (Ph.D., Brown University)

Digman, Jason. "Which Way to the Promiseland? Changing Patterns in Southerns' Migration, 1865 to 1920." (Dissertation in progress at the University of Illinois at Chicago, Department of History).

Dillon, Lisa Y. 1997. "Between Generations and Across Borders: Living Arrangements of the Elderly and their Children in Victorian Canada and the United States." (Ph.D., University of Minnesota).

Ehrlich, Steven Richard. 1999. "Migration, Housing Consumption, and Capital Gains in the Pre-taxpayer Relief Era." (Ph.D., University of Pennsylvania).

Gardner, Todd. 1998. "The Metropolitan Fringe: Suburbanization in the United States Before World War II." (Ph.D., University of Minnesota).

Glick, Jennifer Elyse. 1997. "Changing Household Structure Patterns in the Mexican Origin Population: Life Course, Family Survival Strategy, and Culture Incorporation Determinants." (Ph.D. University of Texas at Austin).

Goeken, Ronald. 1999. "Unmarried Adults and Residential Autonomy: Living Arrangements in the United States, 1880-1990." (Ph.D. University of Minnesota).

Hacker, J. David. 1999. "The Human Cost of War: White Population in the United States, 1850-1880." (Ph.D. University of Minnesota).

Hill, Mark E. 1999. "Studies in African American Social Demography." University of Pennsylvania.

Kallgren, Daniel. 1995. "The Individual, the Family and the Community in the Rise of American School Attendance." (Ph.D. University of Minnesota).

Kessler, Alan. Labor market Competition Between Immigrants and Dmestic Lborers. (Dissertation in progress at University of Texas at Austin).

Kauffman, Kathy. "Outsourcing the Hearth: Immigration and the strategic allocation of labor by American families." (Dissertation in progress at Columbia University).

Jones, William P. African American Lumber Workers in the South Between 1929 and 1950. (Dissertation in progress at University of North Carolina at Chapel Hill).

Magnuson, Diana L. 1995. "The Making of a Modern Census: The United States Census of Population, 1790-1940." (Ph.D., University of Minnesota). Currently Assistant Professor of History, Bethel College, St. Paul, Minnesota.

Moehling, Carolyn M. 1997. "Work and Family: Intergenerational Support in American Families, 1880-1920." (Ph.D., Northwestern University). Finalist, Allen Nevins Prize. Currently Assistant Professor, Ohio State University.

Nørgaard, Helle. "Globalizing Cities: Socio-Economic Restructuring of New York City." (Dissertation in progress at University of Copenhagen, Institute of Geography).

Smith, Mark W. 1999. "Essays on the Political Economy of State Welfare Programs." (Ph.D. Yale University).

Sobek, Matthew. 1997. "A Century of Work: Gender, Labor Force Participation, and Occupational Attainment in the United States, 1880-1990." (Ph.D., University of Minnesota).

Suzik, Jeff. "'Work as Paradise': The Civilian Conservation Corps and the Social Gendering of Male Youth, 1933-42." (Dissertation in progress at Carnegie Mellon University, Department of History).

Washington, Mary. 1998. "White, Black, Or Mulatto: A Sociological Exploration of The Meaning Of Racial Classification In The United States Census Of 1880." (PhD., Johns Hopkins University)

Wickni, Frank. "Do Student's Respond to Outside Incentives? Evidence from Microdata." (Dissertation in progress at the University of Oregon, Economics Department).

Zahrt, Elizabeth. 1997. "There's No Place Like Home? Reservation Institutions and American Indian Migration, 1985-1990." (Ph.D., University of California-Riverside).

Zimring, Carl. "Recycling for Profit: The Growth of the American Scrap Trade, 1830-1960." (Dissertation in progress at Carnegie Mellon University, Department of History).