

Project Summary

There are five countries in the world that possess completely digitized individual-level censuses for the late nineteenth century: Canada, Great Britain, Iceland, Norway, and the United States. At the end of the nineteenth century, these five nations were closely connected by economic ties and migration flows. All were undergoing rapid social, economic and demographic transitions. The recent availability of an extraordinary volume of individual-level census data in all five countries creates unprecedented opportunities for international collaborative and comparative research. This proposal seeks funding to develop a web-based collaboration that will maximize the usefulness of these data for understanding the North Atlantic world as it entered the industrial age.

The North Atlantic database will provide a remarkable resource for the study of international migration, urbanization, industrialization and the fertility transition. It will allow the comparative investigation of a wide range of topics in social and demographic history, such as the employment of women and children, domestic service, boarding and lodging, marriage patterns, living arrangements of the aged and of children, industrial and occupational structure, life-course transitions to adulthood, the household economy, and internal migration. For each of these topics, the database will provide information on the interrelationships among variables that cannot be obtained from any other source. Most important, the availability of information on entire populations—not just samples—will open important new avenues of research in all five countries. It will allow study of small and dispersed population subgroups and will make possible new kinds of analysis using longitudinal, multi-level, and geographic approaches.

To take advantage of these opportunities, however, research teams in all five countries recognize that efforts must be closely synchronized. The goal of this project is to coordinate existing national projects by means of innovative tools for data development. No funds are requested for data collection. A dozen funding agencies in five countries have contributed substantial resources to develop these data. A small investment in collaboration among these countries will leverage these assets and ensure that the full potential of the data is realized. By working together, we can create a database of exceptional power. This fundamental social science infrastructure will stimulate broad-based comparative investigations of economic development and demographic change.

The North Atlantic Population Project entails eight tasks: (1) maintain close communication via the Internet, by telephone, and in person to develop consistent standards of census interpretation; (2) develop common classification systems for occupation, birthplace, family relationship, and group quarters type that balance the goals of international comparability, retention of detail, and ease of use; (3) create web-based software that will allow researchers in multiple countries to classify census responses in a common data dictionary and to continuously monitor and discuss coding decisions made by other countries; (4) develop a consistent set of constructed variables describing household composition, family interrelationships, urban and metropolitan residence, and other geographic characteristics; (5) document the comparability of census enumeration instructions, procedures and definitions across the five countries; (6) create machine-understandable metadata compliant with the Data Documentation Initiative metadata standard; (7) develop efficient web-based software to optimize and simplify access to the entire database; and (8) disseminate the database through mirrored websites in each country.

This is a collaboration of five organizations that have been central to the creation and dissemination of North Atlantic historical census data: the Minnesota Population Center, the UK Data Archive (Essex), the Norwegian Historical Data Centre (Tromsø), the Digital Archive of the Norwegian National Censuses (Bergen) and the Institute for Canadian Studies (Ottawa). All five organizations are leaders in the development of new methods for web-based electronic data dissemination, and the investigators in each country have published extensively on the international comparability and coding of historical census data.

Project Description

Objectives

There are five countries in the world that possess completely digitized individual-level censuses for the late nineteenth century: Canada, Great Britain, Iceland, Norway, and the United States. All these datasets were created for genealogical or historical purposes and are now being adapted for social science research.

These five nations at the end of the nineteenth century were closely connected by economic ties and migration flows. All were undergoing rapid social, economic and demographic transitions. The recent availability of an extraordinary volume of individual-level census data in all five countries creates unprecedented opportunities for international collaborative research. To take advantage of these opportunities, however, we will need to coordinate our activities closely.

We seek funding to develop a web-based collaboration that will maximize the usefulness of these data for understanding the North Atlantic world as it entered the industrial age. We are not requesting money to carry out data collection; we are requesting funds to coordinate existing national projects by means of innovative tools for data development. A dozen funding agencies in five countries have contributed substantial resources to develop national census databases for the late nineteenth century. A small investment in collaboration among these countries will leverage these assets and ensure that the full potential of the data is realized. By working together, we can create a database of exceptional power. This fundamental social science infrastructure will stimulate broad-based comparative investigations of economic development and demographic change.

The North Atlantic Population Project will carry out eight principal tasks:

- Maintain close communication via the Internet, by telephone, and in person to develop consistent standards of census interpretation;
- Develop common classification systems for occupation, birthplace, family relationship, and group quarters type that balance the goals of international comparability, retention of detail, and ease of use;
- Create web-based software that will allow researchers in multiple countries to classify census responses in a common data dictionary and to continuously monitor and discuss coding decisions made by other countries;
- Develop a consistent set of constructed variables describing household composition, family interrelationships, urban and metropolitan residence, and other geographic characteristics;
- Document the comparability of census enumeration instructions, procedures and definitions across the five countries;
- Create machine-understandable metadata compliant with the Data Documentation Initiative metadata standard;
- Develop efficient web-based software to optimize and simplify access to the entire database; and
- Disseminate the database through mirrored websites in each country.

This is a collaboration of five organizations that have been central to the creation and dissemination of North Atlantic historical census data. Each of the collaborators brings critical expertise and experience to the project. The Minnesota Population Center is widely known for its work on U.S. historical census data and has a well-established track record for timely completion of innovative infrastructure projects in the social and behavioral sciences. The UK Data Archive (Essex), one of the world's principal social science data repositories, has unrivaled experience in the improvement of historical British microdata. The Norwegian Historical Data Centre (Tromsø) and the Digital Archive of the Norwegian National Censuses (Bergen) have worked for two decades to create the most comprehensive electronic archive of historical census data anywhere. Finally, the Institute for Canadian Studies (Ottawa) has recently emerged as an

important center for the improvement of Canadian historical data. All five organizations are leaders in the development of new methods for web-based electronic data dissemination, and the investigators have published extensively on the international comparability and coding of historical census data.

Background and Significance

Introduction

By remarkable serendipity, complete machine-readable population censuses of five North Atlantic countries in the late nineteenth century will soon become available for social science research. The Church of Latter-Day Saints (LDS), in collaboration with local genealogical societies, laboriously digitized three of these censuses—for Britain, Canada and the United States—to provide a resource for genealogical research. That massive project involved some 4.8 million hours of work by thousands of volunteers and professionals, and resulted in a verified transcription of the census information on the approximately 85 million individuals who resided in those countries in 1880 or 1881.

The copyright for the British data is held by Her Majesty's Stationery Office, and the data were produced with the cooperation of the Public Record Office and The General Register Office for Scotland. Consequently, the completed dataset is the property of the British Government. The History Data Service of the UK Data Archive has recently obtained authorization from both London and Edinburgh to distribute the datasets to academic researchers. In the U.S. and Canadian cases, the transcriptions are copyrighted by the LDS. In the past twelve months, the Minnesota Population Center and the Institute for Canadian Studies at the University of Ottawa have both negotiated agreements with the LDS allowing us to freely distribute the data to academic researchers in exchange for cleaning the data.

The Norwegian and Icelandic cases are somewhat different. Over the past two decades, Norwegian researchers have invested more than half a million hours in digitizing historical population records. The national censuses of 1865 and 1900 are now complete, and the census of 1875 is well underway. Although the primary use of these materials to date has been for genealogical purposes, they were envisioned from the beginning as a source for social science research. The database is a collaborative product of the Norwegian Historical Data Centre (Tromsø) and the Digital Archive of the Norwegian National Censuses (Bergen). In Iceland, the censuses of 1860, 1870 and 1901 have been transcribed as part of an effort to construct genealogies for genetic research, and work is underway on the censuses of 1880 and 1890.

The result of all these labors is a transcription of the characteristics of 90 million persons who resided on the North Atlantic rim in the late nineteenth century. The census in each case provides information on age, sex, marital status, family relationships, occupation and birthplace, and allows the construction of a full complement of variables describing household composition, fertility, and neighborhood and community characteristics. In their present form, however, these data are of minimal use for social science research. There are literally millions of occupational titles, birthplaces, family relationships and geographic localities transcribed in four different languages. Before any of these data can be fully exploited, each variable must be numerically coded and classified. Efforts are already planned or underway in each country to carry out such classification. This proposal seeks funds to coordinate our work, so that we will be able to pool the datasets and carry out cross-national analyses of the North Atlantic population.

Initial discussions about the potential for creating an integrated database for the entire populations of the United States, Britain and Canada occurred in Ottawa in April 1999, at a meeting of the International Microdata Access Group (IMAG). The Norwegian and British projects were already underway, and

participants from those countries described their plans for converting their data into a form usable by social scientists. The U.S. and Canadian participants had just learned of the existence of the LDS transcriptions of data from the censuses of 1880 and 1881, and all immediately realized the potential for a powerful integrated social science database.

During the course of the next year, the Canadian and U.S. groups obtained permission from the LDS to disseminate the data and raised external funds to process the datasets. In June and October 2000 participants from each country met in Minneapolis to define the goals of the project and develop a detailed plan of work. The participants agreed that we should not simply create compatible datasets, but rather should develop a single fully integrated database with common coding systems, constructed variables, documentation and dissemination systems. We agreed that this ambitious plan for international collaboration would require additional funding.

The collaborators on this project have pieced together funding from numerous sponsors in four countries to support the painstaking tasks of data cleaning and coding. In Britain, the funders include the Economic and Social Research Council, the Leverhulme Trust, and the Essex University Research Promotion Fund; in Canada, the Social Sciences and Humanities Research Council, the Harold Crabtree Foundation, the Church of Jesus Christ of Latter-Day Saints, and the University of Ottawa Research Partnerships Programme; in Norway, the Norwegian Research Council, the Norwegian National Archives and the Faculty of Social Sciences of the University of Tromsø; and in the United States, the National Science Foundation and the National Institutes of Health. A comparatively modest infusion of support for international collaboration will leverage this investment and allow us to create an extraordinary resource for comparative social and economic research.

We envision the North Atlantic Population Project as the foundation for a long-term collaborative enterprise to reconstruct the population of this region from the mid-nineteenth century to the present. Table 1 describes the surviving individual-level censuses for each country. The five countries involved in this collaboration are fortunate to have extraordinarily rich collections of surviving individual-level census data; indeed, these are probably the five best-endowed nations of the world in this respect. Although many countries undertook censuses in the late nineteenth century, in almost all other cases the individual-level enumerator's returns were destroyed or lost. In the long run, we envisage a series of complete census transcriptions for each country. This would allow a longitudinal perspective on the North Atlantic population as it underwent industrialization, urbanization and demographic transition. Our proposed collaboration is an essential first step, and will establish the standards for future expansion of the database.

Advantages of a complete-count census database

Usable national census samples already exist for late-nineteenth century Britain, Canada and the United States; these samples are identified with an "S" in Table 1. The proposed database, however, will be far more powerful than these existing resources. The availability of information on the entire population will open important new avenues of research in all five countries. The paragraphs that follow describe some of the new methodological approaches that will be possible with complete count data.

Study of small, dispersed population subgroups. The minimum acceptable number of cases for census data is substantially larger than is needed for a typical rectangular sample survey. For many topics of study, the relevant individuals for analysis are a small subset of the sample population. For example, fertility analysis is ordinarily limited to the population of women 15 to 49 years old and studies of occupational structure are restricted to the employed population. The precision of census samples is further limited because such samples are invariably clustered by households. Some of the most important variables for analysis—such as region, size of locality, household structure and race—are perfectly or

Table 1. Availability of North Atlantic census microdata, 1835-2001

	1840	1860	1880	1900	1920	1940	1960	1980	2000										
Britain	M	S	M	M	C	M	P	M	M	M	M	M	M	Z	Z	S	Z		
Canada			M	S	C	Z	S	Z	Z	Z	Z	Z	M	S	S	S	S	Z	
Iceland	C	M	C	M	M	C	P	P	C	P	M	C	M	M	M		S		
Norway				C	P		M	C	M	M	M		M	M	S	S	S	S	Z
U.S.		S	S	S	C		S	S	S	S	Z	S	S	S	S	S	S	S	Z

M Manuscript individual-level census survives
S Machine-readable national sample of census exists
C Complete machine-readable transcription exists
P Complete machine-readable census transcription planned or in progress
Z Machine-readable national census sample planned

near perfectly homogeneous within clusters. When analyzing such characteristics, the number of households rather than the number of individuals determines sample precision (Ruggles 1995a). The number of cases needed to analyze a population subgroup depends on the type of subgroup, the type of analysis, population heterogeneity and desired precision. If high precision estimates are required, many thousands of cases of the subgroup of interest may be necessary.

Many small population subgroups—defined by race, ethnicity, occupation or even age—can only be studied with very high-density datasets. For example, the availability of complete-count data will allow study of the indigenous populations of Canada, Norway and the United States. The new database will even include sufficient indigenous women of childbearing age to allow in-depth fertility analysis. Similarly, existing sample data are insufficient to study immigrant groups in detail. A substantial proportion of the Icelandic population emigrated to Canada and the United States, but there are insufficient Icelandic cases in any of the existing samples to allow quantitative analysis. The number of Norwegians is larger, but is still insufficient for detailed analysis. The new database will also be sufficiently large to compare specific occupations across all five nations, such as sailors and fishermen, and will for the first time allow comparative study of centenarians.

Community studies. The community study has been one of the most fruitful analytical approaches in both history and sociology. The existing sample datasets lack sufficient cases to examine particular localities. Because it includes the entire population, the new database will allow historians and sociologists to extract customized datasets focusing on particular communities. The international dimension of the database will allow investigators to undertake comparative community studies. For example, an investigator could compare patterns in a Minnesota Norwegian community with the sending community in Norway. There is a large historical demand for local statistical data, and the North Atlantic database will immediately become an essential tool for community historians of all sorts. Even historians who make little use of quantitative analysis will be able to quickly and painlessly locate their study subjects in the manuscript census.

Longitudinal analysis. Perhaps the greatest limitation of the existing samples is that they are cross-sectional snapshots and do not allow one to trace individuals across time. This problem will be greatly alleviated by the new database. In Britain, Canada and the United States, there exist machine-readable samples of the census for multiple years. Thus, it will be possible to create a series of linked samples; in the case of Canada, for example, individuals in the 1871, 1891 and 1901 census samples can be linked to

the complete-count 1881 census. Thus, researchers will be able to construct three linked Canadian samples, covering 1871-1881, 1881-1891 and 1881-1901. As shown in Table 1, in Norway, Iceland and Britain there are existing or in preparation complete censuses from multiple census years. These datasets offer the potential to link individuals across more than a single pair of census years. Researchers will even be able to link some individuals across countries, especially from Norway and Iceland in 1865 to the United States in 1880 and Canada in 1881.

Historians have been linking individuals across censuses for decades, but the results are problematic. In most cases, linked census studies have been based on local populations because no complete census for a larger area has been available. These studies generally lose between 60 and 80 percent of the population each decade due to linkage failures (see for example Katz 1975, Knights 1991, Thernstrom 1964, Guest 1987, Ferrie 1996). Most linkage failure is attributable to the very high migration characteristic of the mid-nineteenth century. The availability of high-quality census files including entire populations will allow far more sophisticated matching than has previously been possible. Using the new database, for example, entire countries can be searched using characteristics such as age, sex, birthplace, birthplace of mother, and birthplace of father as well as name. The new database will allow a far higher rate of matches than have previous studies and will be able to provide samples thousands of times larger. Moreover, because the analyses will be based on representative populations at both ends of the record linkage, any biases in the linked population will be readily detectable.

Linked census data holds the promise of finally resolving some of the longest-running debates in nineteenth-century social history. Past studies of social and geographic mobility were ultimately inconclusive because of their exclusion of migrants and their small sample size. Scholars will be able to gauge the extent of social and geographic mobility, analyze the interrelationship of geographic and economic movement, and assess trends and differentials in social mobility far more reliably than heretofore (Thorvaldsen 1995). In addition, the linked samples will allow investigation of questions regarding family formation and dissolution. For example, they will allow us to answer several controversial questions surrounding the formation of multigenerational households in the nineteenth century (Ruggles 1994a, 2000).

Multilevel analysis. In recent years, multilevel analyses of the effects of local context on individual behavior have proven exceedingly valuable tools for research in historical sociology (see for examples Elman 1998; Kramarow 1995; Ruggles 1997a, 1997b). A key problem for such nineteenth-century research, however, is that the method requires independent variables tabulated for small geographic units, and such data are scarce before the twentieth century. The new North Atlantic sample will allow creation of a wide variety of contextual variables—such as racial or ethnic composition, female labor-force participation, and occupational structure—at any geographic level, including the block, the neighborhood, and the enumeration district.

Geographic Information Systems. Geographers are ordinarily unable to tap the power of microdata. The existing nineteenth-century microdata files are samples, so when they are used for small areas they provide insufficient precision for reliable mapping. Although some relatively high-density samples are available for the period since 1970, those microdata files suppress detailed geographic data. Therefore, geographers are forced to rely on complete count aggregate data that usually provide only basic summary statistics for small areas.

The North Atlantic census database will provide full geographic detail for every individual in the population. Digitized small-area boundary files are already in preparation for nineteenth-century Norway and Britain, and a pending proposal to the National Science Foundation would provide a similar resource for the United States. Thus, there is already a large scholarly investment in nineteenth-century

geographic information systems. What is lacking is a fine level of geographic detail in social, economic, and demographic characteristics. The North Atlantic census database will allow scholars to marry existing geographic boundary files to population characteristics, thus creating a powerful new analytic tool. Such fine geographic analysis will be especially potent in the analysis of topics such as early suburban development and racial and ethnic residential segregation (see for example Gardner 1998).

Substantive Research Areas

A cross-section encompassing the entire population of the North Atlantic world in the late nineteenth century will open up vast new terrain in the fields of history, economics, demography, and sociology. The censuses include a great deal of information on demography and social structure that can only be taken advantage of through the creation of a new microdata set. The late nineteenth century is a critical period in the study of fertility decline, urbanization, international migration, household composition and occupational structure. The database will allow the construction of cross-tabulations on a wide range of topics that were not covered by census publications or were incompletely tabulated. Perhaps even more important is the potential for longitudinal and multilevel multivariate analyses opened by the availability of the database. The North Atlantic 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.

A full discussion of the specific topics that could be addressed with a complete machine-readable database of the nineteenth century censuses of five countries would require many pages. The paragraphs that follow sketch only a few of the most obvious research applications of the new database.

Industrialization. The first Industrial Revolution may have begun in Lancashire, but by the late nineteenth century, the entire North Atlantic world was involved in manufacturing, the production of raw materials, or both. The North Atlantic database will allow unprecedented opportunities to explore economic structures within and between each nation during this critical transitional period.

For the first time, we will have consistently coded occupational data available for multiple nineteenth century countries, and it will be available at the individual level for the entire population. This will allow comparative analysis at the level of persons, families, communities or regions, and investigation of the geographic organization of economic activity. In four of the five countries, for example, mechanized textile manufacturing existed, and the census provides sufficient occupational detail to analyze the organization of the industry in each locality. All five nations were deeply involved in and interconnected by maritime industries. They competed in the rich North Atlantic fishery and in the transatlantic shipping trade. The North Atlantic database will not only reveal the structure of maritime industries, but also will allow the comparative investigation of maritime communities.

Fertility transition. At the time these censuses were taken, each of the North Atlantic countries was just beginning deliberate fertility limitation.¹ The North Atlantic database will allow study of differential fertility patterns in this critical period of demographic transition, to assess the importance of such factors as occupational class, ethnicity, region, literacy, local economy, size of locality and family structures. Study of this elemental shift in population structure has the potential to enhance our understanding of ongoing demographic change in the contemporary developing world.

¹ Until recently, demographers thought that American fertility decline began much earlier than in most of Northern Europe. Hacker (1999, 2000a), however, has shown using census-based own-child and back projection methods that deliberate fertility limitation in the United States began considerably later than was previously thought. Hacker places the fertility transition in America after the Civil War, which is the same period that it occurred in each of the other countries in the North Atlantic database.

Past comparative analyses of the European fertility transition have relied on aggregate vital statistics (Coale and Watkins 1986). This approach has two major disadvantages. First, aggregate vital statistics do not allow direct measures of child spacing or stopping behavior; only the *level* of fertility can be considered. Second, the aggregate approach does not allow control of individual-level socioeconomic characteristics.

The new database will allow analysis of fertility differentials through own-child methods (Cho, Retherford and Choe 1986). Own-child methods of fertility analysis require very large datasets, and are therefore especially well suited to complete population databases. Thus, the database will allow a new and more subtle generation of comparative studies of the first demographic transition.

Household and family composition. For more than a century, political theorists, sociologists and historians have been debating the relationship between industrialization and the family. In the 1970s, a series of British, Canadian and American studies argued that the harsh economic conditions of early industrial capitalism strengthened the interdependence of family members and led to a high frequency of complex households (Anderson 1972; Hareven 1978, 1982; Katz 1975; Foster 1974; Modell 1978). Each of these analyses focused on a single industrializing community, and so were unable to test the proposed association between industrial development and family or household composition.

In recent years, there have been numerous national and regional studies of family composition in the late nineteenth century based on sample data, but few have incorporated community-level economic measures (Sogner 1990, 1998; Gunnlaugsson and Garðarsdóttir 1996; Dillon 1997, 1998, 2000; Ruggles 1994b, 2000; Wall 1995). Comparisons across national boundaries have also been inhibited by inconsistencies in the construction of measures of household composition. Thus, there is presently little agreement about national similarities and differences in family and household composition in the late nineteenth century. Some of the most promising recent work has focused on relatively small population subgroups, such as the living arrangements of the aged or of unmarried mothers of young children, but only the largest samples are capable of supporting such investigations.

The North Atlantic database will include a common set of constructed variables to aid in the analysis of family and household composition and will thus allow consistent comparisons across all five countries. It will allow investigators to assess the impact of local context on family systems through multilevel analysis, and thus for the first time permit analysis of the effects of individual-level factors, local economic conditions, regional inheritance systems, and national characteristics on the nineteenth-century family.

International migration. The late nineteenth century saw international population movements on an unprecedented scale. The massive North Atlantic migration profoundly shaped both the receiving and contributing countries. The great majority of emigrants from Norway, Iceland and Britain went to Canada and the United States, and the influx transformed North American society. Many of these newcomers remained only a few years before returning to their homelands, often bringing home money and always bringing new ideas and experiences (Runblom and Norman 1976; Nugent 1992; Gjerde 1992; Thorvaldsen 1997).

The North Atlantic database will be a wonderful resource for the study of migration history. It will allow close and consistent comparisons of occupational structure, marriage patterns, fertility and family composition. Researchers will be able to identify and compare specific sending and receiving communities. In some instances, it will even be possible to follow individual migrants across the Atlantic and back again. In combination with new machine-readable ship lists and emigration registers, the database will open a new window on the implications of international population flows.

Educational applications of the database

In addition to scholarly research, we anticipate that the new database will make important contributions to teaching in the social sciences, helping to bring the excitement of discovery into the classroom. The detailed geographic analysis made possible by the new database makes it a suitable vehicle for introducing a quantitative dimension into secondary, undergraduate and graduate courses focusing on local history. Once the North Atlantic database is created, we plan to collaborate in the development of web-based instructional materials that capitalize on the fine detail available for local areas and small population subgroups.

Research Plan

Source Data

Table 2 describes the data files to be included in the North Atlantic database. We plan to include complete data from nine censuses: the three census transcriptions created by the LDS for Britain, Canada and the United States in 1880/1881, and six files covering the Icelandic and Norwegian censuses between 1860 and 1901. With respect to their structure, organization and available information, the nine censuses are remarkably comparable. In each case, the censuses describe the characteristics of individuals grouped into households, and the interrelationships of individuals within households can be determined. All countries defined households as a group of people sharing a common place of residence. There is a core set of variables common to virtually all datasets, including relationship of each individual to the household head, age, sex, marital status, occupation and birthplace. The geographic units identified vary from country to country, mainly because of differences in the political organization of each nation. In all countries, however, we can identify the location of all places with 5,000 or more persons, and we estimate that we can identify approximately 25,000 places across the five countries. The common core variables will allow us to construct a variety of new variables describing community and neighborhood characteristics, household composition, socioeconomic status, and family interrelationships.

Most of the censuses were taken on a *de jure* basis, under which individuals who were temporarily absent from home—such as migrant workers and travelers—were to be enumerated at their usual place of residence. The exception is the British census, taken under a *de facto* rule which specified that no one who was present on census night at a particular address could be left out of the tally, and that no person absent from home could be written in. In Norway and Iceland from 1870 onwards, persons were to be enumerated both at their usual place of residence *and* at the place they stayed on census day, and enumerators identified both temporary visitors and absent household members. These variations in enumeration rules pose only minor compatibility problems, but investigators of some household composition and migration issues will have to be aware of them.

Data quality is good. The United States census is probably the weakest of the group in this respect. The United States was the largest and most heterogeneous of the five countries. Moreover, the weak federal system of American government meant that census administration was decentralized, making it difficult to enforce uniform standards of enumerator training and accountability (Magnuson 1995). Nevertheless, the 1880 census had a comparatively modest undercount. The most recent demographic analysis indicates that net underenumeration of the 1880 census was 6.4 percent (Hacker 2000b; King and Magnuson 1995). Gross underenumeration may have been as high as ten percent, since some persons were double counted. Although coverage was not complete, the overall response rate to the American census in the late nineteenth century therefore compares favorably with modern survey data, such as the Current Population Survey. We lack comparable estimates for Britain, Canada, Iceland and Norway, but because of their smaller size, more homogeneous populations, lower geographic mobility and stronger central governments, census taking was considerably less challenging. As part of the project, we will

Table 2. Variables in the proposed North Atlantic census database

X= Variable taken directly from source

C= Constructed variable

Country	Britain	Canada	Iceland			Norway			United States
Census year	1881	1881	1860	1870	1901	1865	1875	1900	1880
Enumeration rule	<i>de facto</i>	<i>de jure</i>	<i>de jure</i>	both	both	<i>de jure</i>	both	both	<i>de jure</i>
Number of person records (000)	30,000	4,300	67	70	72	1,702	1,813	2,240	50,155
I. HOUSEHOLD RECORD									
Household characteristics									
County	X								X
City, town, village	X	X	X	X	X	X	X	X	X
Province/state		X	X	X	X	X	X	X	X
Parish	X		X	X	X	X			
Enumeration district	X	X	X	X	X		X	X	X
School district						X			
Address	X		X	X	X	X	X	X	
Microfilm reel or folio number	X	X	X	X	X	X	X	X	X
Census page number	X	X	X	X	X	X	X	X	X
Number and type of rooms								X	
Farm residence	C	C	X	X	X	X	X	X	C
Constructed household variables									
Record type (Household)	C	C	C	C	C	C	C	C	C
Household sequence number	C	C	C	C	C	C	C	C	C
Number of persons in household	C	C	C	C	C	C	C	C	C
Group quarters residence	C	C	C	C	C	C	C	C	C
Group quarters type	C	C	C	C	C	C	C	C	C
Urban/rural residence	C	C	C	C	C	C	C	C	C
Size of place	C	C	C	C	C	C	C	C	C
Metropolitan area	C	C	C	C	C	C	C	C	C
Community characteristics	C	C	C	C	C	C	C	C	C
Household type (UN system)	C	C	C	C	C	C	C	C	C
Household type (Hammel/Laslett)	C	C	C	C	C	C	C	C	C
Number of families	C	C	C	C	C	C	C	C	C
Size of primary family	C	C	C	C	C	C	C	C	C
Number of children under 18	C	C	C	C	C	C	C	C	C
Number of married couples	C	C	C	C	C	C	C	C	C
Number of secondary individuals	C	C	C	C	C	C	C	C	C
II. PERSON RECORD									
Individual characteristics									
Relationship to household head	X	C	X	X	X	X	X	X	X
Age	X	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X	X
Occupation	X	X	X	X	X	X	X	X	X
Marital status	X	X	X	X	X	X	X	X	X

(Table continued on following page)

Table 2. Variables in the proposed North Atlantic census database (continued)

X= Variable taken directly from source

C= Constructed variable

Country Census year	Britain	Canada	Iceland			Norway			United States
	1881	1881	1860	1870	1901	1865	1875	1900	1880
Individual characteristics (continued)									
Place of birth	X	X	X		X	X	X	X	X
Parental birthplaces									X
Citizenship/Nationality	X					X	X	X	
Ethnicity/Race		X				X	X	X	X
Religion		X				X	X	X	
Disability	X					X	X	X	
Surname	X	X	X	X	X	X	X	X	X
Given name	X	X	X	X	X	X	X	X	X
Absent or visiting on census day				X	X		X	X	
Constructed person variables									
Record type (person)	C	C	C	C	C	C	C	C	C
Person number in household	C	C	C	C	C	C	C	C	C
Socioeconomic scores	C	C	C	C	C	C	C	C	C
Surname similarity code	C	C	C	C	C	C	C	C	C
Location of spouse	C	C	C	C	C	C	C	C	C
Location of own mother	C	C	C	C	C	C	C	C	C
Location of own father	C	C	C	C	C	C	C	C	C
Number of own children	C	C	C	C	C	C	C	C	C
Number own children under 5	C	C	C	C	C	C	C	C	C
Age of eldest own child	C	C	C	C	C	C	C	C	C
Age of youngest own child	C	C	C	C	C	C	C	C	C

develop a comparative analysis of enumeration procedures and systematically gather evidence on underenumeration from all five countries.

Variable coding

The data for each country are presently in the form of alphabetic character strings that represent a transcription of the information collected from each individual in the late nineteenth century. The individual records are grouped into residential units corresponding to the modern census concepts of household and group quarters. There are approximately 90 million of these records, and the information is recorded in English, French, Icelandic or Norwegian. In their present form, the data have little social science application, because the number of variations of each variable is too great for researchers to digest. For example, we estimate that the data include approximately four million different occupational strings, one million birthplaces and 50,000 family relationships.

Each country has raised funds to classify these alphabetic strings into numerically coded categories. This work is already underway in Britain, and is scheduled to begin during the coming year in Canada, Norway and the United States. If it were not for the proposed collaboration, each country would code variables strictly according to their own conventions, and the result would be five separate and fundamentally incompatible datasets. Some variables—age, sex and marital status—can be made

comparable with little effort, but the complex variables will require close collaboration to develop common coding standards.

Occupational coding is the most challenging component of the project. The fine detail available in the occupational field is one of the reasons why the North Atlantic database has the potential to transform our understanding of historical social structure. At the same time, however, the complexity of occupational structure will demand meticulous care to ensure consistency. We have already agreed to adopt the Historical International Standard Classification of Occupations (HISCO) as our basic framework for occupational classification (van Leeuwen, Maas and Miles 1997; Edvinsson and Karlsson 1998). The HISCO system is a modification of the 1968 United Nations occupational classification system with extensions to accommodate historical occupations. An international committee with representatives from Belgium, Canada, England, France, Germany, the Netherlands, Norway, Sweden and the United States is near completion of the final description of the system. We will further modify and extend the system to accommodate the additional detail available in the North Atlantic database. Similarly, we intend to adapt United Nations classification systems as a framework for the other principal complex variables, such as birthplaces, family relationships, group quarters, and ethnicity.

To translate from character strings into numeric codes, we must construct a data dictionary that assigns a numeric code to each alphabetic variation that occurs in the data. This work is difficult enough in the context of a single country; for a project of this scale, it requires a team of expert coders who work in close cooperation, sharing coding decisions continuously. We envision a merged dictionary of unprecedented scale that would include the alphabetic strings from all five countries. This is uncharted territory. Until recently, constructing such a dictionary would have necessitated assembling experts with appropriate language skills and historical knowledge about each country in a single location. This would be prohibitively expensive. With the advent of Internet technology, however, we can develop cost-effective tools that allow us to distribute the task and work on a dictionary in multiple countries simultaneously.

A central goal of this project is to develop a web-based collaboratory that will allow us to coordinate coding operations. Working from a common data dictionary, researchers in each country will classify census responses for their country and will continuously monitor and debate coding decisions in each other country. The software will be developed by the Minnesota Population Center, which has extensive experience in web-based database management. In addition to the new software, we will take full advantage of off-the-shelf web-based meeting and communications technology to maintain continuous interaction among coders.

The data and dictionaries will be maintained in a SQL application with a web interface. The software will allow researchers in each country to:

- browse all unique alphabetic strings for each field;
- assign numeric codes to each string;
- review and approve numeric codes assigned by other countries;
- search fields for substrings;
- view cases in the context of their entire households or communities;
- edit alphabetic fields in individual person records; and
- display the number of occurrences of each unique string in each census.

The system is best explained by reference to the most complex variable, occupational title. Each country will be responsible for designating a HISCO code for each string that occurs in that country. As batches of occupations are completed, they will be reviewed and approved by at least one other country. In the

case of Icelandic, French, and Norwegian titles, coders will provide an English translation of each occupational string. When the existing HISCO codes lose too much detail, coders will propose extensions of the HISCO system.

In some cases, an identical string must be coded into two different occupational titles, depending on country. For example, the term “engineer” has distinctly different meanings in the United States and Britain. The software will allow each country to assign a different code to the same string, but will require that any such discrepancies be reviewed and approved by each country involved.

Where possible, the system will suggest codes based on existing data dictionaries. In the United States, for example, we have data dictionaries created for the national samples of the 1850, 1860, 1870, 1880, 1900, 1910 and 1920 censuses. These dictionaries include approximately 100,000 different occupational titles coded into both the 1950 and 1880 United States occupational and industrial classification systems. In many cases, the combination of these classifications will uniquely identify a specific HISCO code; in other cases, they will be sufficient to identify a broader HISCO occupational group. Similar dictionaries exist for Canadian, Norwegian, and British data. Thus, whenever an occupational string from the census transcription appears in an existing dictionary, the software will suggest the appropriate HISCO code or group of codes.

The web-based collaboratory will not eliminate the need for face-to-face communication. Accordingly, we have included modest funds for an annual meeting of all key personnel and additional topical meetings for subsets of participants.

Constructed variables

We will design and implement a consistent set of constructed variables describing household composition, family interrelationships, urban and metropolitan residence and other geographic characteristics. Constructed variables are identified by a “C” in Table 2. Software to create the constructed variables will be designed at Minnesota. The constructed variables fall into eight categories, detailed below.

1. We will create technical variables to aid in data management and analysis, such as record type, serial number, group quarters residence and household size.
2. We will construct variables describing urban/rural residence, size of place, and metropolitan residence.
3. We will create several variables describing the characteristics of districts and neighborhoods, such as population density, percent of employment in agriculture, percent of employment in manufacturing, and percent of employment in resource extraction. In addition, we will explore the potential for variables describing land use and climate.
4. We will provide geographic coordinates for the lowest level of geographic detail for which information is readily available.
5. We will construct a standard set of variables to describe the composition of families and households. These variables will replicate the most commonly-used historical and contemporary classification systems.
6. We will create variables to aid in the analysis of family interrelationships. Among the most useful of these are pointers to own parents and own spouse; for each individual, we will specify the location within the family of their own mother, father and spouse, if present. These variables will allow users of statistical software packages to attach the characteristics of immediate kin (spouse’s birthplace, father’s occupation, children’s characteristics, etc.) without the need to resort to programming (Ruggles 1995b).

7. We will add the variables required for basic own-child fertility analysis: number of own children present, number under five years old, age of eldest child and age of youngest child.
8. We will create indices of socioeconomic status, such as Duncan Scores (Duncan 1961), occupational income scores (Sobek 1995, 1997), the Registrar-General's classification, and the Cambridge Scale Score (Prandy 1992).

Documentation

The microdata are of little use without adequate metadata to interpret them. The design of an integrated documentation system is central to the project. We will provide comprehensive documentation on each of the censuses included in the database. The process of writing this documentation is demanding intellectual labor, but it is critical to ensure the intelligent use of the database. The development of these materials will be a collaborative enterprise, requiring close coordination of the national projects.

Detailed variable discussions are not sufficient in themselves. We will provide a wide collection of supporting information to aid in the interpretation of the data. Users will often require access to information from the original census collection, so we plan to include facsimiles of census forms and enumerator instructions, and procedural histories of each census. We will provide images of census forms, maps and any other documentary elements not readily presentable in text format. Where the original documentation is in another language, we will translate the most essential material into English. Where foreign-language material is extensive, however, we will provide English-language summaries as well as the full text in the original language. The documentation system will also describe all procedures undertaken to generate the integrated database. This documentation will include the actual computer code, the data dictionaries and a textual description of the data manipulation process.

Since the amount of material will be large, we will implement advanced automated search features in the documentation system. Users will be able to search by keywords and concepts across variable descriptions and all of the various elements of the metadata, including census forms, enumerator instructions, programming descriptions, and topical essays. Rich hypertext links throughout the documentation system will allow nonlinear access to information based on the user's needs and interests. All documentation will be available on the web and on CD-ROM or DVD-ROM, and we will make a self-extracting downloadable version so users can easily install the documentation system on their desktop computer.

All documentation will be compliant with the Data Documentation Initiative (DDI) metadata standard. The DDI is a non-proprietary, hardware independent, neutral standard that preserves the content and relational structure of the full documentation. The machine-understandable structure of the DDI allows for automated processing by data access software. The DDI standard, completed in March 2000, was developed by an international committee that represented a range of stakeholders in social science data dissemination, including the Census Bureau, the Bureau of Labor Statistics and the national data archives of Great Britain, Norway, and Canada.

Dissemination and Preservation

Few users will be interested in analyzing the entire dataset of 90 million cases. Therefore, we will disseminate the database through an automated DDI-based data extraction tool. The data access system for the North Atlantic database will be based on the software we are currently developing as part of the Integrated International Microdata Access System (IIMAS) and will build on our experience with the IPUMS data access tool (<http://www.ipums.org>). Briefly, the IIMAS extraction system integrates access to metadata and microdata and allows users to carry out substantial manipulation of the data without resorting to programming. Owing to differences between the late-twentieth century sample data used in the IIMAS project and the nineteenth-century complete-count North Atlantic data we will need to modify

the system, but this will be far less expensive than starting from scratch. The data access system will be mirrored in Britain, Canada, Norway and the United States, and will provide unrestricted access for academic researchers. Without this collaboration, it is likely that access to the British, Norwegian, and Icelandic data would be far more restrictive, especially for researchers based outside those countries.

Long-run survival of the database beyond the project period is critical. The University of Minnesota Population Center, the Norwegian Historical Data Center, and the UK Data Archive all guarantee to maintain the system for a period of at least 25 years beyond the end of the project. In addition, we will deposit the database and access software with the Inter-University Consortium for Political and Social Research as well as the UK Data Archive to ensure permanent preservation.

Collaborators

The investigators will work closely together, with weekly on-line meetings and daily interaction. If a consensus cannot be reached on particular design issues or coding decisions, the collaborators have all agreed to abide by the opinion of the majority. Steven Ruggles, Director of the Minnesota Population Center, is the Principal Investigator. He will be in charge of overall project coordination, will oversee software development and will be in charge of the U.S. coding operation, which is funded by NICHD.

Kevin Schürer, Director of the UK Data Archive, will lead the British component of the project, assisted by Matthew Woollard, an expert on historical occupational structure. The British data, comprising 30 million cases, represents a formidable challenge. Coding operations are already well under way, funded primarily by the Economic and Social Research Council and the UK data archive. The funds requested in this grant for Britain represent only the additional costs that will be incurred to make the British coding compatible with the other countries.

Lisa Y. Dillon, President of the International Microdata Access Group and Chad Gaffield, Director of the Institute for Canadian Studies, will direct the Canadian effort. The Canadian census project is comparatively underfunded, and we are requesting modest assistance for both data management and translation of French-language occupational titles and family relationships.

The Norwegian project will be jointly headed by Gunnar Thorvaldsen, Director of the Norwegian Historical Data Centre and Jan Oldervoll, Director of the Digital Archive of the Norwegian National Censuses. These investigators will obtain funding for the project from the University of Tromsø, the University of Bergen, the Norwegian Research Council and other Norwegian sources. Professor Thorvaldsen will spend the first year of the project on sabbatical leave at Minnesota to participate in the design of the web-based collaboratory.

The Icelandic dataset is somewhat different: because the Icelandic population was much smaller than that of the other North Atlantic countries, census coding operations are substantially less expensive. The coding and manipulation of the Icelandic data will be shared between the Norwegian Historical Data Centre and the Minnesota Population Center. In addition, we will hire Ólöf Garðarsdóttir to consult on Icelandic occupations, birthplaces and other variables. Garðarsdóttir, an Icelandic national affiliated with the Department of Historical Demography at the University of Umeå, has published sixteen articles based on historical census research.

Schedule of work and deliverables

In the first year, programmers at Minnesota will develop on-line dictionary management system. In addition, the national teams will prepare and exchange HISCO-coded excerpts of dictionaries. As data cleaning is completed, we will standardize the format of the raw transcriptions and will load the data into a common database management system. We will begin the production phase of the data dictionary work

in the second year. During the third year of the project, we will refine the web-based dictionary software and commence work on design and implementation of constructed variables. In the final year, we will complete the final version of software for producing the dataset with all constructed variables, write the documentation, and finalize the data dissemination software. We will release the database simultaneously on servers in Britain, Norway and the United States in September 2005.

Results of Prior NSF Research

The Principal Investigator has directed several major NSF-funded projects without which the present initiative would not be feasible. The first of these was the Integrated Public Use Microdata Series (IPUMS). Among several NSF awards for the IPUMS project, the most important was the first: "Integrated Public Use Microdata Series," SBR-9118299, \$464,913, 4/1992-10/1995. Before the IPUMS, it was difficult to use U.S. census microdata in time series because of variations in classification systems, file formats and documentation. The IPUMS transformed a diverse collection of census microdata files into a coherent series of individual-level U.S. census data drawn from thirteen census years between 1850 and 1990. By putting all the census samples in a compatible format with consistent variable codes and integrating their documentation, the IPUMS has greatly simplified the use of multiple census years. Just as important, the IPUMS project pioneered methods of electronic dissemination that have democratized access to these resources. In the five years since the first general release of the IPUMS, it has served as the basis for ten books, 23 completed Ph.D. dissertations, 143 published articles, at least 20 dissertations in progress and hundreds of working papers, conference presentations and research reports (<http://www.ipums.org/usa/research.html>).

The second major project was "International Integrated Microdata Access System," SBR-9907416, \$3.5 million, 10/1/99-9/30/04. The IIMAS project is collecting and preserving late-twentieth century census microdata samples from around the world, creating and disseminating an integrated international census database incorporating data from seven countries, producing integrated documentation for the database, and developing an electronic data access system. The first phase of the project will include census microdata from Brazil, China, Colombia, France, Hungary, Ghana, Kenya, Mexico, Spain, the United States and Vietnam. The IIMAS project differs sharply from the present project, since it involves late-twentieth century pre-coded census samples that were created as a byproduct of computerized census processing. The North Atlantic project, by contrast, must digest raw transcriptions of the characteristics of entire nineteenth-century populations. Nevertheless, the two projects are highly complementary. Most important, the North Atlantic project will take advantage of the enormous investment of the IIMAS project in dissemination software. Moreover, we plan to draw samples from the North Atlantic database and convert them to IIMAS-compatible form, thus enriching that database.

The third and most closely connected NSF project was "Population Database of the United States in 1880," SES-9910961, \$200,000, 5/1/01-4/30/01. The goal of this project was to acquire and clean the census data for 1880 created by the Church of Latter-Day Saints, in exchange for permission to freely disseminate the data for academic research. That task is now nearing completion, and we will soon embark on coding the data with funding from the National Institute of Child Health and Human Development (HD 39327). Without this prior support, the present project would be impossible.

Many publications have resulted from this work; for examples, see Hall, McCaa and Thorvaldsen (2000); Fitch and Ruggles 2000; Gardner 1995, 1998, 1999, forthcoming; Gardner, Sobek and Ruggles 1999; Ruggles 1993, 1994a, 1994b, 1995a, 1995b, 1996a, 1996b, 1997a, 1997b, 2000; Ruggles and Brower forthcoming; Ruggles, Hacker and Sobek 1995; Ruggles and Menard 1995; Ruggles and Sobek 1998; Ruggles, Gardner, and Sobek 1996; Ruggles, Sobek, and Gardner 1996; Sobek 1996, 1997; Sobek and Ruggles 1999; Block and Star 1995.

References Cited

- Anderson, Michael. (1972). *Family Structure in Nineteenth Century Lancashire*. Cambridge, England: Cambridge University Press.
- Block, William C. and Dianne L. Star. (1995). "Data entry and verification." *Historical Methods* 28: 63-65.
- Coale, Ansley J. and Susan C. Watkins, eds. (1986). *The Decline of Fertility in Europe*. Princeton: Princeton University Press.
- Cho, Lee-Jay; Retherford, Robert D.; Choe, Minja Kim. (1986). *The Own-Children Method of Fertility Estimation*. Honolulu: University of Hawaii Press.
- Dillon, Lisa Y. (forthcoming). "Elderly women in late Victorian Canada." In Bob Hesketh and Chris Hackett, eds. *Canada: Confederation to Present*. (CDRom), University of Alberta.
- Dillon, Lisa Y. (2000). "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 4: (4) 447-483.
- Dillon, Lisa Y. (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.
- Dillon, Lisa Y. (1998). "Between generations and across borders: Living arrangements of the elderly and their children in Victorian Canada and the United States," Ph.D. Dissertation, Minneapolis: University of Minnesota.
- Duncan, Otis Dudley. (1961). "A socioeconomic index for all occupations." In A. Reiss et al., *Occupations and Social Status*. Chicago: Free Press.
- Edvinsson, Sören and Johnny Karlsson. (1998). "Recoding occupations in the Demographic Data Base into HISCO." In *HISMA Occasional Papers & Documents Series*, Berlin. No. 3/1998.
- Elman Cheryl. (1998). "Intergenerational household structure and economic change at the turn of the twentieth century." *Journal of Family History* (4) 417-440.
- Ferrie, Joseph P. (1996). "A new sample of males linked from the Public Use Microdata Sample of the 1850 U.S. federal census of population to the 1860 U.S. federal census manuscripts." *Historical Methods* 29: 141-156.
- Fitch, Catherine A. and Steven Ruggles. (2000). "Historical trends in marriage formation." In Linda Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thomson and Arland Thornton, eds., *Ties that Bind: Perspectives on Marriage and Cohabitation*. Hawthorne: Aldine de Gruyter, 59-88.
- Foster, John O. (1974). *Class Struggle in the Industrial Revolution: Early Industrial Capitalism in Three English Towns*. London.
- Gardner, Todd. (1995). "Software development for the Public Use Microdata Samples." *Historical Methods* 28: 59-62.

- Gardner, Todd. (1998). "The metropolitan fringe: Suburbanization in the United States before World War II." Ph.D. Dissertation, University of Minnesota.
- Gardner, Todd. (1999). "Metropolitan classification for census years before World War II." *Historical Methods* 32: 139-50.
- Gardner, Todd. (forthcoming). "The slow wave: The changing residential status of cities and suburbs in the United States, 1850-1940." *Journal of Urban History*.
- Gardner, Todd, Matthew Sobek and Steven Ruggles. (1999). "The IPUMS data extraction system." *Historical Methods* 32: 119-24.
- Gjerde, Jon. (1992). *From Peasants to Farmers: the Migration from Balestrand, Norway to the Upper Middle West*. Cambridge: Cambridge University Press.
- Guest, Avery. (1987). "Notes from the National Panel Study: linkage and migration in the late nineteenth century." *Historical Methods* 20: 63-77.
- Gunnlaugsson, Gísli Ágúst and Ólöf Garðarsdóttir. (1996). "Transition into widowhood: life course perspective on the household position of icelandic widows at the turn of the century." *Continuity and Change* 11(3), 435-458.
- Hacker, J. David. (1999). "The Human Cost of War: White Population in the United States, 1850-1880." Ph.D. dissertation, University of Minnesota.
- Hacker, J. David. (2000a). "Rethinking the 'Early' Decline of American Fertility." Paper presented at the Minnesota Population Seminar, October 2000.
- Hacker, J. David. (2000b). "New estimates of census underenumeration in the United States, 1850-1880." Paper presented at the annual meeting of the Population Association of America, Los Angeles.
- Hall, Patricia Kelly, Robert McCaa and Gunnar Thorvaldsen (2000). *Handbook of International Historical Microdata*. Minneapolis: Minnesota Population Center.
- Hareven, Tamara K. (1978). "The dynamics of kin in an industrial community." In *Turning Points: Historical and Sociological Essays on the Family* John Demos and S.S. Boocock, eds. Chicago: University of Chicago Press.
- Hareven, Tamara K. (1982). *Family Time and Industrial Time: The Relationship between the Family and Work in a New England Industrial Community*. New York: Cambridge University Press.
- Katz, Michael B. (1975). *The people of Hamilton, Canada West: family and class in a mid-nineteenth-century city*. Cambridge: Harvard University Press.
- King, Miriam L. and Diana L. Magnuson. (1995). "Perspectives on historical U.S. census undercounts." *Social Science History* 19(4): 455-66.
- Kramarow, Ellen. (1995). "The elderly who live alone in the United States: historical perspectives on household change." *Demography* 32(3): 335-52.

- Knights, Peter R. (1991). *Yankee Destinies: the Lives of Ordinary Nineteenth-Century Bostonians*. Chapel Hill: University of North Carolina Press.
- Magnuson, Diana L. (1995). *The Making of a Modern Census: The United States Population Census, 1790-1950*. Ph.D. Dissertation, University of Minnesota.
- Modell, John (1978). Patterns of consumption, acculturation, and family income strategies late nineteenth-century America. In T.K. Hareven and M. Vinovskis, eds. *Family and Population in Nineteenth Century America*. Princeton: Princeton University Press.
- Nugent, Walter (1992): *Crossings: the Great Transatlantic Migrations, 1870-1914*. Bloomington, Indiana: Indiana University Press.
- Prandy, K. (1992). "Cambridge Scale Scores for CASOC Groupings." Working Paper No. 11, Sociological Research Group, Cambridge University.
- Runblom, Harald and Hans Norman, eds. (1976). *From Sweden to America: a history of the migration*. Minneapolis, Minnesota: University of Minnesota Press.
- Ruggles, Steven. (1993). "Historical demography from the census: applications of the American census microdata files." In Roger Schofield and David Reher, eds. *Old and New Methods in Historical Demography*. Oxford: Oxford University Press, 383-93.
- Ruggles, Steven. (1994a). "The origins of African-American family structure." *American Sociological Review* 59: 136-51.
- Ruggles, Steven. (1994b). "The transformation of American family structure." *American Historical Review* 99: 103-28.
- Ruggles, Steven. (1995a). "Sample designs and sampling errors in the Public Use Microdata Samples." *Historical Methods* 28: 40-6.
- Ruggles, Steven. (1995b). "Family interrelationship coding in the Integrated Public Use Microdata Series." *Historical Methods* 28: 52-8.
- Ruggles, Steven. (1996a). "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 démographiques du passé*. Lyons: Centre Jacques Cartier, 21-40.
- Ruggles, Steven. (1996b). "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*. New York: Aldine de Gruyter, 254-71.
- Ruggles, Steven. (1997a). "The rise of divorce and separation in the United States, 1880-1980." *Demography* 34: 455-66.
- Ruggles, Steven. (1997b). "The effects of AFDC on American family structure, 1940-1990." *Journal of Family History* 22: 307-25.
- Ruggles, Steven. (forthcoming). "Living arrangements and economic well-being of the aged in the past." *Population Bulletin of the United Nations*.

- Ruggles, Steven, J. David Hacker and Matthew Sobek. (1995). "Order out of chaos: General design of the Integrated Public Use Microdata Series." *Historical Methods* 28: 33-9.
- Ruggles, Steven and Susan Brower. (forthcoming). "New estimates of household composition in the United States, 1850-1990." In Michael Haines, Richard Sutch and Susan Carter, eds., *Historical Statistics of the United States: Millennial Edition*. Cambridge: Cambridge University Press.
- Ruggles, Steven, Todd Gardner and Matthew Sobek. (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 and Matthew Sobek. (1998). *IPUMS-98: Integrated Public Use Microdata Series* (five volumes). Minnesota Population Center.
- Ruggles, Steven, Matthew Sobek and Todd Gardner. (1996). "Distributing large historical census samples on the Internet." *History and Computing* 9: 145-59.
- Sobek, Matthew. (1995). "Occupation and Income Scores." *Historical Methods* 28, 47-51.
- Sobek, Matthew. (1996). "Work, status, and income: Men in the American occupational structure since the Nineteenth Century." *Social Science History* 20: 169-207.
- Sobek, Matthew. (1997). "Occupational structure and the labor force in the United States, 1880-1990." Ph.D. Dissertation, University of Minnesota.
- Sobek, Matthew and Steven Ruggles. (1999). "The IPUMS project: An update." *Historical Methods* 28: 102-10.
- Sogner, Sølvi (1990). *Far sjøl i stua og familien hans : trekk fra norsk familiehistorie før og nå* ("The father and his family. Aspects of Norwegian Family History") Oslo.
- Sogner, Sølvi (1998). "The Norwegian stem family: myth or reality?" In A.Fauve-Chamoux & E. Ochiai eds., *Household and the stem family in EurAsian Perspective*, Proceedings of the Twelfth International Economic History Congress, August 1998: 95-114.
- Thernstrom, Stephan. (1964). *Poverty and progress; social mobility in a nineteenth century city* Cambridge, Massachusetts: Harvard University Press.
- Thorvaldsen, Gunnar. (1995). *Migration in the province of Troms 1865-1900. A Study based on the Censuses*. (Dr. philos avhandling: *Migrasjon i Troms i annen halvdel av 1800-tallet. En kvantitativ analyse av folketellingene 1865, 1875 og 1900*).
- Thorvaldsen, Gunnar. (1997). "Marriage and names among immigrants to Minnesota, *Electronic Journal of the the American Association for History and Computing* <http://mcel.pacificu.edu/history/JAHC/Thorvaldsen/ThorIndex.HTML>
- van Leeuwen, H.D., Ineke Maas and Andrew Miles. (1997). Creating a historical international standard coding of occupations (HISCO), Max Planck Institute, Berlin.

Wall, Richard. (1995). "Elderly persons and members of their households in England and Wales from preindustrial times to the present." In David I. Kertzer and Peter Laslett, eds. *Aging in the Past: Demography, Society and Old Age*. Berkeley: University of California Press, 81-106.