Big Microdata for Population Research

Steven Ruggles

Abstract This article describes an explosion in the availability of individual-level population data. By 2018, demographic researchers will have access to over 2 billion records of accessible microdata from over 100 countries, dating from 1703 to the present. Another 2 to 4 billion records will be available through restricted-access data enclaves. These new resources represent a new kind of data that will enable transformative research on demographic and economic change and the spatial organization of society.

Keywords Microdata · Big data · Censuses

Introduction

The quantity of microdata available for population research is exploding. In 2000, about 100 million individual microdata records were readily accessible to the research community. We now have over 750 million records describing individuals, and the number will exceed 2 billion by 2018 (Fig. 1). Billions of additional records will become available in coming years through restricted data enclaves. This vast new trove of microdata—in concert with new technologies—has the potential to transform the spatiotemporal analysis of demographic behavior and economic activity. Most of the data consist of high-density samples or complete census enumerations, and they usually provide rich geographic detail. Some data extend as far back as 1703, and they describe more than 100 countries representing over 80 % of the world's population. The Integrated Public Use Microdata Series (IPUMS) will make the new data easily interoperable over time and between countries.

In this article, I describe the original development of microdata by the U.S. Census Bureau, new microdata from international statistical agencies and historical sources, and restricted-access microdata. I conclude by suggesting some new research opportunities and broader implications of big data for population research.

S. Ruggles (✉)
Minnesota Population Center, University of Minnesota, 225 19th Ave South, Minneapolis, MN 55455, USA
e-mail: ruggles@umn.edu

The Origins of Microdata

The U.S. Census Bureau invented microdata a half-century ago. The expansion of social science research in the late 1950s and early 1960s had led to growing demand for special tabulations, which the Census Bureau prepared on a reimbursable basis (Kraus 2011). To meet the demand, in 1962 the Census Bureau drew a 1-in-1,000 sample of the long-form records from the 1960 census, removed identifying information, and made the data available to researchers on computer tapes for $1,500 (Brunsman 1963; Duncan and Shelton 1978; Hauser 1960; U.S. Census Bureau 1964). The demographic community was highly enthusiastic about the new kind of data. Otis Dudley Duncan stated that "the importance of this innovation can hardly be overestimated" (Duncan 1974:5097). Mason et al. (1977:3) concurred, explaining that the 1960 census microdata sample was a "development of profound significance to social research" because it gave the research community "freedom to retabulate or manipulate without the constraints imposed by a fixed set of printed volumes."

The costs of data processing and storage fell dramatically in the 1960s and early 1970s, and the scale of microdata increased proportionately. In 1972, the Census Bureau released 68 times the quantity of microdata from the 1970 census as it had for the 1960 census. A year later, the Census Bureau issued an expanded version of the 1960 microdata designed to be fully compatible with the 1970 data with respect to record layout and coding. Beginning in 1973, the Census Bureau also created public-use microdata for the Current Population Survey (CPS) (Berg 1973). Census Bureau microdata from both the decennial census and the CPS quickly became basic tools for population research, and by the mid-1980s they were the most widely used sources in the pages of *Demography* (Ruggles 2005; Ruggles et al. 2012). In the following decades, the Census Bureau continued to release microdata from the CPS and from each successive decennial census, and since 2000 has added large annual microdata samples from the American Community Survey (ACS). The Census Bureau has now
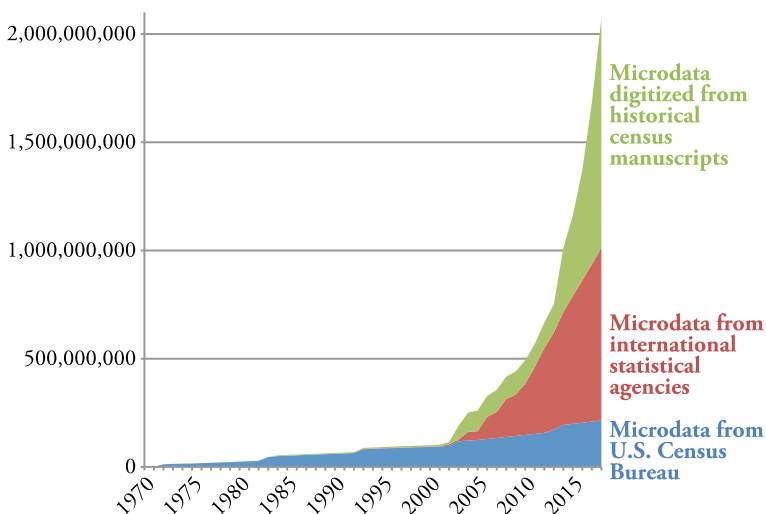

Fig. 1 Number of individual cases in microdata readily accessible for population research

released 157 million records describing individuals, and the number will rise to about 206 million by 2018.

Microdata From International Statistical Agencies

Despite the success of large-scale microdata disseminated by the Census Bureau, statistical agencies in other countries were slow to create similar resources.[1] Before 2000, most countries had no systematic program for preservation or reuse of census microdata once the statistical agency had published summary tables. As a result, most machine-readable census microdata from the 1960s and 1970s had already disappeared by the mid-1990s. Much surviving microdata were at immediate risk of destruction because of deterioration of the storage media or retirement of technical staff needed to locate and interpret the files.

Historical demographer Robert McCaa decided to take action. He believed that international microdata should be easily accessible to all researchers under a uniform set of nondisclosure rules, and so he embarked on a 15-year campaign to liberate and preserve the world's statistical heritage. This effort has been amazingly successful. From 1998 to 2011, McCaa convinced 100 national statistical agencies around the world to collaborate with the IPUMS project; perpetual agreements with each country ensure long-run preservation and free access for the academic community (McCaa and Ruggles 2002; Sobek et al. 2011).

As a direct result of McCaa's efforts, IPUMS has released anonymized integrated microdata samples for 238 censuses of 74 countries taken between 1960 and 2011. Most countries outside Western Europe and North America provide good geographic precision, identifying places with as few as 20,000 residents (Fig. 2). With a few exceptions, individuals are nested within families and households, and the data contain information about the interrelationships of all members of each residential group. The data also include information on economic activities, ethnicity, educational attainment, fertility, migration and place of former residence, marital status, and consensual unions. Many developing countries provide information about mortality and disabilities, as well as extensive housing characteristics, usually including water supply, sewage, and physical characteristics of the dwelling such as floor and roof materials and number of rooms. For 63 countries, IPUMS provides microdata from multiple census years (3.6 census years per country, on average). The samples are usually large: almost two-thirds include 10 % or more of the population, and 85 % of the samples include at least 5 %. Taken together, the 238 samples currently available include 545 million observations. By 2018, the IPUMS project expects to have released data for almost 800 million observations drawn from 300 censuses of about 100 countries.

---

[1] A few national statistical offices—including those of Australia, Brazil, China, Colombia, Mexico, Norway, and South Africa—made internal microdata available to selected academic researchers by special arrangement. Statistics Canada began producing Public Use Microdata Files (PUMF) in 1974, and the United Kingdom created Samples of Anonymized Records (SARS) in 1993.

	Period Covered	No. of Censuses and Surveys 2013	No. of Censuses and Surveys 2018	Millions of Person-Records 2013	Millions of Person-Records 2018	Population of Smallest Identified Places (thousands)	Name/Other Identifier	Utilities	Dwelling Characteristics	Basic Demographic	Fertility/Mortality	Birthplace and Nativity	Ethnicity/Language	Work	Education	Income	Migration	Disability
1. U.S. Census Bureau																		
Censuses	1960–2010	6	7	67	85	100	○	○	●	●	∘	●	●	●	●	●	●	●
Surveys	1962–present	481	546	91	121	100–500	○	○	●	●	∘	●	●	●	●	●	●	●
2. IPUMS-International																		
Africa	1970–present	38	63	68	135	20		●	●	●	●	●	●	●	●	∘	●	●
Asia	1970–present	60	84	198	333	20–90		●	●	●	●	●	○	●	●	∘	●	•
Latin America	1960–present	77	87	145	197	20		●	●	●	●	●	∘	●	●	•	●	•
Europe/Canada	1960–present	56	67	86	123	20–16,000		○	●	●	∘	●	∘	●	●	∘	●	○
3. Historical																		
Britain	1851–1911	2	6	30	206	No minimum	○			●	∘	●		○	●	●	●	●
United States	1790–1950	10	16	67	798	No minimum	○		○	●	∘	●	○	•	•	∘	○	•
Other	1703–1930	21	53	33	68	No minimum	○			●	∘	●	●	○				
4. Restricted																		
United States	1960–present	17	23	1,146	1,482	No minimum	∘	●	●	●	∘	●	●	●	●	●	●	●
Africa	1970–present	0	50	0	453	No minimum		●	●	●	●	●	●	●	●	∘	●	●
Asia	1970–present	0	15	0	603	No minimum		●	●	●	●	●	○	●	●	•	●	•
Latin America	1960–present	0	58	0	583	No minimum		●	●	●	●	●	∘	●	●	∘	●	●
Total Accessible Through IPUMS		751	929	785	2,066													
Total Including Restricted Access		751	929	1,773	4,817													

Key

●Many variables available in most/all sources ○Some variables available in most/all sources •Many variables available in some sources ∘Some variables available in some sources

Fig. 2 Characteristics of big microdata collections

Microdata From Historical Sources

The fastest-growing category of big microdata is based on digital transcriptions of historical census enumeration forms dating from 1703 to 1950. Historians were the first scholars to use census microdata outside of statistical agencies: in the late 1930s, Owsley and Owsley (1940) transcribed U.S. census enumeration schedules to punch cards and used an electric sorting machine to analyze the social structure of the antebellum South. Census microdata were a mainstay of the "new" social and economic histories of the 1950s through the 1970s, but the resulting historical data sets were generally proprietary and typically covered only one or two localities.[2]

Soon after the Census Bureau released consistently coded microdata for 1960 and 1970, demographers Samuel Preston and Halliman Winsborough independently arrived at the idea of extending the series backward by digitizing national samples of the historical census schedules and making them available as a general resource for the research community. Preston produced samples for the 1900 and 1910 censuses, and Winsborough and colleagues made samples for 1940 and 1950. In the late 1980s, Russell Menard and I took over the historical digitization project, and we constructed samples to fill in the rest of the U.S. censuses back to 1850 (Ruggles 2005).

The rapid growth of historical census data after 2000 has resulted primarily from the adaptation of genealogical data to meet scientific purposes. Starting in 1982, FamilySearch—the genealogical arm of the Church of Jesus Christ of Latter-day Saints—organized an army of 25,000 volunteers to transcribe information stored on microfilm images of the 1880 U.S. census and the 1881 censuses of Britain and Canada. When the work was completed in the late 1990s, FamilySearch provided copies of digital census transcriptions describing 84 million persons for academic use. In 1999, these data became the centerpiece of a new international collaboration—the North Atlantic Population Project (NAPP)—to develop complete census enumerations for comparative historical research (Roberts et al. 2003). In addition to Britain, Canada, and the United States, NAPP now includes participants from Denmark, Egypt, Iceland, Ireland, Mexico, Norway, and Sweden. Each year, more censuses are added. NAPP presently disseminates data on 130 million persons drawn from 33 censuses of nine countries enumerated between 1703 and 1930 (Ruggles et al. 2011a).

Over the coming five years, the quantity of accessible historical microdata will increase by an order of magnitude, to 1.1 billion individual records. About 95 % of this increase will result from three major new collaborations with genealogical organizations. Kevin Schürer, one of the original NAPP collaborators, is working with findmypast.com to create a complete British microdata series that includes every enumerated person and variable in every census from 1851 to 1911, a net addition of 175 million records. The British series will have especially precise geographic coding, with fine-grained, consistent geography describing both place of birth and place of residence. In a second major new project, the Minnesota Population Center is collaborating with Ancestry.com to digitize all variables from the 1940 census of the United States and outlying territories, for a total of 134 million persons and 70 variables, including wage and salary income, educational

---

[2] Early national historical samples were created for Argentina and Canada, but they did not become broadly accessible until much later (Darroch and Ornstein 1979; Somoza and Lattes 1967).

attainment, migration, detailed employment information, and street address. We plan to use the street address information to geocode the location of individual households.

The third and biggest new microdata collection will capitalize on a donation of unprecedented scale of census data digitized by both Ancestry.com and FamilySearch. Over the past decade, the two genealogical organizations independently digitized information about all persons enumerated in the U.S. censuses of 1790 through 1930. In July 2008, they agreed to merge their databases and reconcile the discrepancies to improve accuracy (Ancestry.com 2008); the reconciliation was completed in 2012. Ancestry.com and FamilySearch devoted approximately 22 million hours to the transcription of information from 650 million individual records, the equivalent of over 10,000 person-years of effort. The data-entry cost to replicate the collection in the United States would be about $420 million.[3] Ancestry.com has now donated this extraordinary data collection for academic research and education.

The Ancestry/FamilySearch database for 1790–1930 includes a core set of variables for every census year, including geographic location, age, sex, and race, as well as name. Birthplace information is available in all but a few of the early years, and from 1880 forward the data include marital status, the relationship of each individual to the household head, and the birthplace of each individual's mother and father, allowing the identification of second-generation Americans. Other key variables—such as year of immigration, duration of marriage, literacy, occupation, children ever born, children surviving, and disability—are available sporadically.

With a few exceptions, the historical data are transcribed from public sources, and there are no confidentiality restrictions, so the historical data sets ordinarily provide full geographic information. The historical data sets also ordinarily include names, although to protect commercial interests, there are restrictions on the dissemination of names for the data sets donated by genealogical organizations. Nine countries have multiple complete enumerations with full identification, allowing researchers to trace individuals and households over the life course and across several generations. NAPP has already produced preliminary linked samples for Britain, Canada, Norway, and the United States, and the group plans to link the entire populations of all nine countries. The identified data can also be linked to other sources. For example, plans are already underway to link the 1940 census to five current longitudinal surveys to assess the impact of early-life conditions on later outcomes and to link the 1940 census to later administrative and vital records.

Restricted Microdata

Large-scale microdata encompassing entire populations or very large samples are becoming available for recent censuses as well as historical ones. Because of confidentiality risks, there are significant access restrictions for these data.

The Census Research Data Centers (RDCs), operated by the U.S. Census Bureau's Center for Economic Studies, house complete decennial census microdata for 1970

---

[3] This estimate covers the costs of dual keying only; data cleaning, checking, and reconciling two copies would incur additional expense. The cost estimate assumes the average Ancestry.com keying rate and the U.S. average salary for data-entry keyers according to the Bureau of Labor Statistics (2011).

through 2000, including both short-form and long-form records. Thanks to a recently completed Minnesota Population Center project to restore the 1960 census, complete long-form data covering 25 % of the 1960 population will soon be available through the RDCs (Ruggles et al. 2011b). The RDCs also house American Community Survey data covering 47 million persons, with about 5.4 million persons added per year. In all, the microdata housed in the RDCs currently include 1.1 billion person-records with full geographic identification down to the block level. The Minnesota Population Center is currently converting these data to IPUMS format.

The Minnesota Population Center plans to develop an *international* restricted data enclave modeled on the U.S. Census RDCs. Many of the international censuses provided to IPUMS by statistical agencies arrive as complete enumerations; IPUMS staff then draw samples and anonymize the data to ensure confidentiality. To date, IPUMS has archived 123 complete enumerations from Africa, Asia, and Latin America totaling about 1.6 billion person-records. The project expects to receive a few more complete censuses from the 2000 round and many more from the 2010 round. The International Restricted Data Center (IRDC) will provide access to these data in a secure environment. Like the Census RDC, no restricted data will ever leave the secure servers; instead, remote access will be available through virtualization. All results will be assessed for potential confidentiality risks, and the IRDC staff will email the output to the researchers only after review. The IRDC is a collaboration of IPUMS with many statistical agencies, and the details of the configuration are still being negotiated. In addition to housing complete enumerations, we expect that the IRDC will disseminate hundreds of sample data sets with more geographic detail than is shared under the regular IPUMS microdata dissemination rules.

Why Big Microdata?

Consistent large-scale microdata that extend over many decades and span national boundaries with fine geographic detail provide a unique laboratory for studying demographic processes and for testing social and economic models. These data will enable new kinds of research to understand the impact of geographic context at all levels, from immediate neighbors to continents, and to see how those effects are conditioned by economic and demographic transformations. Promising topics of investigation include residential segregation; migration and migrant settlement patterns; urban sprawl; the economic and ideational context of declines in fertility, mortality, and intergenerational coresidence; rural depopulation and agricultural consolidation; the identification of concentrated poverty; causes and levels of change in ecosystems as a function of human-environment interactions; comparative cross-national policy analysis; and multilevel analysis of the impact of community characteristics on individual behavior.

Consider the case of residential segregation. In the past, investigators studying residential segregation were forced to base their analyses on small-area summary statistics: they could study only places, not people. With complete microdata and full geographic information, researchers will be able to develop new measures of segregation that can control for individual-level and family-level characteristics. In some cases—including the 1880 and 1940 censuses—each household in urban areas will be geocoded, making it possible to assess proximity through individual-level measurement, rather than relying on summary statistics for small areas (Lee et al.

2008; Logan and Zhang 2012). This information, together with the 1940 migration question on place of residence in 1935, will permit powerful multivariate analysis of the determinants of residence decisions at the individual level.

Linked microdata also have the potential for transformative research. Early versions of linked historical microdata have already shifted our perception of life-course change in the past. They have revealed that occupational mobility was far higher in the nineteenth century than it is today, migration was far more frequent, and the formation of intergenerational families was most common among the rich rather than the poor (Ferrie 2005; Long and Ferrie 2007; Ruggles 2011). The next generation of linked microdata will be far more powerful, with 100 times the number of records, more reliable links, and coverage across entire lives and across multiple generations, allowing multilevel analysis of the demographic and economic context of mobility and family transitions.

The new data will complement survey research. Researchers frequently use census summary characteristics to uncover neighborhood characteristics of respondents to demographic surveys, but they are limited to the basic tabulations provided by the published census. The availability of complete microdata for small geographic units will allow creation of more subtle, focused, and consistent measures of neighborhood and community context, and will make it easier to measure neighborhood change. As noted, U.S. surveys with older respondents will be able to identify these persons in the 1940 census, providing measures of early-life conditions at the levels of family and neighborhood.

The large scale of the microdata allows demographers to study particular communities and small, dispersed populations, but it also enables big studies that span many countries and large time spans. Economic development and cultural change across the globe have been highly uneven. This great variation creates opportunities to assess the impact of economic and cultural characteristics on individual behavior, thereby offering the potential for understanding the consequences of both structural and ideational change. To take just one example, Esteve et al. (2012) examined the rise of cohabitation for 350 regions of 13 Latin American countries over four decades and assessed the impact of national and regional social and economic conditions on the pace of change. With big microdata and detailed geographic information, researchers can assess the effects of changing local context on individual behavior at multiple scales. Data that allow investigators to simultaneously examine a broad sweep of time and detail of spatial organization have the potential to yield new insight into the processes of change that are transforming demographic behavior.

To maximize the utility of big microdata, the Minnesota Population Center is developing new infrastructure. Central to this effort is *Terra Populus* (TerraPop), part of the National Science Foundation DataNet initiative to ensure long-term access and preservation of scientific data. TerraPop will bring all the large-scale microdata under one interoperable umbrella, using column-store database technology to accommodate the scale of the data (Abadi et al. 2008). A central goal of the project is to make microdata easily interoperable with other kinds of spatiotemporally referenced data, including raster data sets derived from satellite imagery and climate models, economic indicators, and policy and legal data. For example, the prototype TerraPop system now available allows users to extract characteristics of Malawi farmers and append measures of the crops grown in their local area and the level of agricultural

productivity, along with indicators of temperature and precipitation. In the future, the system will incorporate information about a wide range of characteristics of places, including environmental policy, social insurance, biodiversity, and unemployment. The system will also make it simple for users to convert microdata into small-area or raster format for visualization or spatial modeling.

Big microdata differ from other forms of big data that have recently drawn attention. Former Census Bureau Director Groves (2011) draws a distinction between "designed data," such as the census, and "organic data" that are the byproduct of automatically digitized transactions. There is much excitement about using organic big data from social networks and commercial transactions to better understand society (Giannotti et al. 2013; Keller et al. 2012; King 2011). Data generated solely as a byproduct of social or commercial interactions, however, have limitations as sources for population research. Organic data are voluminous but shallow: they often have no clearly defined universe, are unrepresentative of the general population, and do not systematically provide information about most of the things demographers care about, such as demographic behavior, education, work, and living conditions.

Large-scale microdata have none of these liabilities. The universe is the entire population. Although data quality varies, nonresponse rates for even the worst censuses compare favorably with other sources, including sample surveys routinely used by demographers. The microdata are highly structured, providing consistent information about individuals nested within families, which are in turn nested within neighborhoods, communities, regions, and nations. Finally, the microdata focus directly on the subjects of central interest to population research.

TerraPop provides a spatial framework that can provide context for the organic data described by Groves (2011). Even if organic data tend to be nonrepresentative and shallow, data sources such as cell phone and Internet traffic, nighttime lights from satellite imagery, and even social networking content can be invaluable to social science if they are calibrated to specific populations and places. By combining big microdata with organic big data, we can enrich the microdata and frame the organic data.

Big microdata are a new kind of source material. We will soon have individual-level information about entire populations or large samples covering most of the world's population with multiple observations at high geographic resolution. The data will cover the last two centuries for several North Atlantic countries and the last two to five decades for the rest of the world, allowing us to observe directly the demographic and economic transformations that are reshaping society. We will enrich the microdata with information describing characteristics of the places in which people live, including land use, land cover, climate, and social policies, as well as organic data sources.

Big microdata are just as novel—and I believe just as important—as the original release 50 years ago of the first microdata for the U.S. census. We need new analytic techniques to take advantage of these new opportunities. For example, inferential statistics developed for small sample surveys are inappropriate for analyzing entire populations and billions of records. We need new research strategies, modeling methods, and data mining techniques to capitalize on the scale and scope of the sources. Most important, we need compelling research ideas that can transform our newfound digital abundance into better understanding of the shifting spatial organization of society, the processes of demographic and economic change, and the interactions of human activity and natural systems.

References

Abadi, D .J., Madden, S. R., & Hachem, N. (2008). Column-stores vs. row-stores: How different are they really? In *SIGMOD'08, Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data* (pp. 967–980). New York: Association for Computing Machinery.

Ancestry.com. (2008). *FamilySearch and Ancestry.com team to publish new images and enhanced indexes to the U.S. censuses* [Press release]. Retrieved from http://corporate.ancestry.com/press/press-releases/2008/07/familysearch-and-ancestry.com-team-to-publish-new-images-and-enhanced-indexes-to-the-u.s.-censuses/

Berg, S. V. (1973). The CPS viewed from the outside. *Annals of Economic and Social Measurement, 2,* 99–105.

Brunsman, H. G. (1963). Letter to Dr. Joshua Lederberg, April 22, 1963. Joshua Lederburg Papers, National Library of Medicine. Retrieved from http://profiles.nlm.nih.gov/BB/G/K/F/C/_/bbgkfc.pdf

Bureau of Labor Statistics. (2011). *Occupational employment and wages, May 2011*. 43–9021 Data Entry Keyers. Retrieved from http://www.bls.gov/oes/current/oes439021.htm

Census Bureau, U. S. (1964). *Censuses of population and housing, 1960: Two national samples of the population of the United States, 1/1,000, 1/10,000: Description and technical documentation*. Washington, DC: U.S. Government Printing Office.

Darroch, G., & Ornstein, M. (1979). *Canadian historical social mobility project. National sample of the 1871 census of Canada* [computer file]. Toronto, Ontario: York Institute for Social Research and Department of Sociology, York University.

Duncan, O. D. (1974). Developing social indicators. *Proceedings of the National Academy of Sciences, 71,* 5096–5102.

Duncan, J. W., & Shelton, W. C. (1978). *Revolution in United States government statistics: 1926–1976. U.S. Department of Commerce Office of Statistical Policy and Standards*. Washington, DC: U.S. Government Printing Office.

Esteve, A., Lesthaeghe, R., & López, A. (2012). The Latin American cohabitation boom, 1970–2007. *Population and Development Review, 18,* 55–82.

Ferrie, J. P. (2005). History lessons: The end of American exceptionalism? Mobility in the United States since 1850. *Journal of Economic Perspectives, 19,* 199–215.

Giannotti, F., Pedreschi, D., Pentland, A., Lukowicz, P., Kossmann, D., Crowley, J., & Helbing, D. (2013). A planetary nervous system for social mining and collective awareness. *European Physical Journal, 214,* 49–75.

Groves, R. (2011). "Designed data" and "organic data." Director's Blog, U.S. Census Bureau. Retrieved from http://directorsblog.blogs.census.gov/2011/05/31/designed-data-and-organic-data/

Hauser, P. M. (1960). The 1960 census as an instrument for demographic research. *Population Index, 26,* 199–211.

Keller, S. A., Koonin, S. E., & Shipp, S. (2012). Big data and city living: What can it do for us? *Significance, 9,* 4–7.

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science, 331,* 719–721.

Kraus, R. S. (2011). *Statistical déjà vu: The National Data Center Proposal of 1965 and its descendants* (Working paper). Washington, DC: U.S. Census Bureau Retrieved from http://www.census.gov/history/www/reference/publications/working_papers.html

Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review, 73,* 766–791.

Logan, J., & Zhang, W. (2012). White ethnic residential segregation in historical perspective: US cities in 1880. *Social Science Research, 41,* 1292–1306.

Long, J., & Ferrie, J. P. (2007). The path to convergence: Intergenerational occupational mobility in Britain and the U.S. in three eras. *The Economic Journal, 117,* C61–C71.

Mason, W. M., Taeuber, K. E., & Winsborough, H. (1977). *Old data for new research* (CDE Working Paper 77–3). Madison, WI: Center for Demography and Ecology.

McCaa, R., & Ruggles, S. (2002). The census in global perspective and the coming microdata revolution. *Scandinavian Population Studies, 13,* 7–30.

Owsley, F. L., & Owsley, H. C. (1940). The economic basis of society in the late ante-bellum South. *Journal of Southern History, 6,* 24–25.

Roberts, E., Ruggles, S., Dillon, L. Y., Gardarsdottir, O., Oldervoll, J., Thorvaldsen, G., & Woollard, M. (2003). The North Atlantic Population Project: An overview. *Historical Methods, 36,* 80–88.

Ruggles, S. (2005). The Minnesota Population Center data integration projects: Challenges of harmonizing census microdata across time and place. In *2005 Proceedings of the American Statistical Association, Government Statistics Section* (pp. 1405–1415). Alexandria, VA: American Statistical Association.

Ruggles, S. (2011). Intergenerational coresidence and family transitions in the United States, 1850–1880. *Journal of Marriage and Family, 73,* 136–148.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M., & Sobek, M. (2012). *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota.

Ruggles, S., Roberts, E., Sarkar, S., & Sobek, M. (2011a). The North Atlantic Population Project: Progress and prospects. *Historical Methods, 44,* 1–7.

Ruggles, S., Schroeder, M., Rivers, N., Alexander, J. T., & Gardner, T. K. (2011b). Frozen film and FOSDIC forms: Restoring the 1960 census of population. *Historical Methods, 44,* 69–78.

Sobek, M., Cleveland, L., Flood, S., Ruggles, S., & Schroeder, M. (2011). Big data: Large-scale historical infrastructure from the Minnesota Population Center. *Historical Methods, 44,* 61–68.

Somoza, J., & Lattes, A. (1967). *Muestras de los dos primeros censos nacionales de población, 1869 y 1895* [Samples of the first two national censuses of population, 1869 and 1895] (Documento de Trabajo No. 46). Buenos Aires, Argentina: Instituto Torcuato Di Tella, Centro de Investigaciones Sociales.

U.S. Census Bureau. (1964). *Censuses of population and housing, 1960: Two national samples of the population of the United States, 1/1,000, 1/10,000: Description and technical documentation.* Washington, DC: U.S. Government Printing Office.