Big Census Microdata: IPUMS in the Federal Statistical Research Data Centers

Catherine Fitch, Steven Ruggles and Erin Meyer
IPUMS and the Minnesota Population Center University of Minnesota

Todd Gardner
U.S. Census Bureau

For the past decade IPUMS and the Census Bureau have been collaborating to recover, preserve, document, harmonize, and disseminate all surviving machine-readable population census microdata from 1960 to the present. The fruits of this collaboration are now available. With 1.3 billion person records and 530 million household records, IPUMS-FSRDC is the largest structured individual-level data collection describing the demographic characteristics of a population (Table 1). These big microdata will open important new research opportunities and stimulate the development of new methods of demographic analysis.

Table 1. IPUMS Census Microdata in the FSRDC

	Short Form			Long Form		
	Persons	Density	Fields	Persons	Density	Fields
2010	308,745,538	100	8	NA		
2000	281,421,906	100	7	45,024,000	16	88
1990	248,709,873	100	13	40,000,000	16	87
1980	226,545,805	100	22	43,073,000	19	102
1970	203,184,772	100	20	40,600,000	20	77
1960	NA			44,500,000	25	76
Total	1,268,607,894			213,197,000		

The internal versions of the decennial census data provide high-density samples with low-level geographic information and other additional detail that is not available in the public use files. The IPUMS research team, working with the Census Bureau, has created harmonized versions of these files; harmonized long-form decennial census datasets for 1970 through 2000 are now accessible for use in the FSRDC. These IPUMS-format files provide standardized variable coding over time, making it far easier for researchers to compare census years. The data also includes IPUMS-specific variables, such as family interrelationship variables which identify spouses and children present in the household and help researchers leverage complex household information. The availability of IPUMS-format data allows researchers to develop and test their analyses and data transformations using freely-available public data and to make the most efficient use of their time working inside the FSRDC. We are now planning to create IPUMS-format versions of the American Community Survey (ACS) microdata available in the FSRDC.

One of the key strengths of the restricted-access census microdata files is the fine geographic detail available to study spatial patterns. Unfortunately, the available geographic units—such as

census tracts—are inconsistent from one census to the next. Presently, for example, it is very difficult to measure the change in poverty between 1990 and 2000 for Minneapolis neighborhoods, since there were dozens of changes in tract boundaries between the two census years. We built on the work of our National Historical Geographic Information System (NHGIS) project, which has constructed boundary files for census tracts and counties in all periods. We constructed consistent tracts spanning the entire period from 1960 to the present. In many communities, tract boundaries have been relatively stable, and the consistent tracts will provide an effective and efficient identifier of small area geography. Researchers can use these files to define consistent geographic units in the restricted microdata.

These harmonized complete-count data will allow innovative analyses of spatial change, including racial and ethnic residential segregation, urban and suburban spatial dynamics, environmental justice, and commuting patterns. The data will support consistent analyses of the impact of neighborhood context on individual behavior and will permit studies of the smallest subpopulations, such as individual American Indian tribes, specific occupation groups, or particular immigrant groups.This use of these files will greatly enhance the usability and visibility of a data series that is destined to have a profound impact on the practice of social science research.

This paper will describe the challenges of creating IPUMS-FSRDC, including recovery of parts of the 1960 microdata. We will describe some of the exciting new research opportunities made possible by this new data resource. Finally, we will provide details about the data, including differences between the public use data and the internal census files; description of constructed IPUMS variables, such as family interrelationships; an overview and explanation of the consistent geography variables; and our plans to create IPUMS-format versions of ACS data in the FSRDC.