

# Reports on National Historical GIS Projects

## The U.S. National Historical Geographic Information System

Over the past twenty years, there has been a rapid growth in the area of geographic information systems (GIS) and the related creation of spatially addressable data sets. Much spatial data are related to population statistics, including U.S. Census Bureau's TIGER files (geocoded street and enumeration-unit files that allow for spatial analysis and mapping). GIS has significantly broadened the scope of questions that can be asked with geospatial data, and has popularized the use of mapping techniques for the display of spatial information. GIS has facilitated the increased use of geodemographic analysis, including geomarketing and many forms of population analysis, by providing powerful means of analyzing the wealth of population data that are now spatially referenced. Examples of these methods include the assessment of environmental justice and racism at multiple spatial scales (regional, urban, community); the calculation of segregation indices, and identification of areas of concentrated poverty; and the development of neighborhood indicators, including a multitude of economic and social measures based on population data. All of these kinds of analysis are based on applying GIS methods to census data.

Increasingly, researchers are attempting to use the census geographic base files for geodemographic analyses. For instance, after the 1990 U.S. census, it was possible to document the changes in geodemographics between 1980 and 1990 using the 1990 TIGER files. A common application was mapping the change in minority populations between the two periods. However, researchers are constrained mostly to two or three decades of temporal analysis with the availability of only post-1970 digital files. The development of digital geographic base files for the period 1940 to 1990 would allow a more detailed analysis of population change, at much finer levels of resolution (especially tract level), for most urban areas. Many potential research projects/application areas would benefit from the availability of such boundary files. Unfortunately, there is no complete set of digital census boundary files for the United States. For instance, comprehensive tract-level files simply do not exist other than for a very short time period.

The National Historical Geographic Information System (NHGIS), a five-year project funded by the National Science Foundation, will result in a comprehensive U.S. census database—at the census tract and county

levels—for both geographical and attribute data from 1790 to 2000. Technological change presents an unprecedented opportunity to make these data readily available for social science research. Bringing the complete census within reach of social scientists will unlock the potential of two centuries of data collection, and will stimulate research in economics, history, sociology, geography, and other fields.

In addition to creating a comprehensive spatio-temporal database, the NHGIS project will enable scholars to perform robust spatio-temporal analysis of census data, such as comparing census data with different enumeration boundaries through areal interpolation procedures that will be provided. The project consists of three major components: data and documentation, mapping, and data access.

1. The data and documentation component gathers all extant machine-readable census summary data; fills holes in the surviving machine-readable data through data entry of paper census tabulations; harmonizes the formats and documentation of all files; and produces standardized electronic documentation according to the recently developed Data Documentation Initiative (DDI) specification for metadata standards.
2. The mapping component creates consistent historical electronic boundary files for tracts, counties, and larger geographic units such as states.
3. The data-access component creates a powerful, but user-friendly, Web-based browser and extraction system, based on the new DDI metadata standard. The system provides public access free of charge to both documentation and data, and presents results in the form of tables or maps.

The process of reconstructing census-tract boundaries begins with the Census 2000 TIGER files as the base. From these files, tracts for both 2000 and 1990 can be directly obtained. The 1992 TIGER files (updated after the 1990 census) provide the 1980 tract boundaries. For earlier decades, the tract-level boundaries are created backwards in time from scanned paper census maps to the first decade that a city was tracted which, for many cities in the eastern United States, dates to 1910. The enumeration unit of analysis for this tract work is the county, that is, tracts are scanned and processed together at the county level. When complete, boundary files and a significant amount of census statistical data will be processed for all census tracts and county units. As of the 2000 census, the entire United States had been tracted. County boundaries are based on tract boundaries where possible. For rural areas after 1910 and all areas before 1910, county boundaries are created backward in time, reflecting changes recorded in historical maps and other sources. NHGIS plans to create county boundaries for every county that was enumerated in a U.S. census. More than one-fifth of present-day states will have county data available from the first census, 1790, to the most recent census in 2000.

A significant part of the project involves the creation of several versions of the database corresponding to the different scales included in the NHGIS. This requires developing systematic and robust procedures for geographic generalization. The first component of this work involves determining which scales will best meet the needs of potential users. Will users require detailed metropolitan-level data, or more generalized county-level data for a multistate analysis? Our preliminary work has identified three scales as potentially most useful: one at 1:150,000 for detailed tract-level mapping; one at 1:400,000 for county-level mapping, and one at 1:1,000,000 for regional and national mapping. Current research is focusing on the exact methods of generalization to be applied, the potential measurements that are needed to ascertain both the selection of the algorithms, and the determination of tolerance levels. A final consideration is the Web-based delivery of multiscale geographical data.

The NHGIS is being developed at the University of Minnesota's Population Center. By 2005, researchers expect to have completed almost all census-tract editing. This work includes producing county-level boundary files containing the census tracts for each of the years for which tracts existed. Focus has now turned to comparing the completed files with the census data and making any necessary modifications prior to their release. The project began releasing test versions of the census data in late 2004, with modifications to the test files continuing as necessary. The final release of all tract and county-level data is scheduled for July 2006.

### *NHGIS Resources Online*

- National Historical GIS online: <http://www.nhgis.org>

—*Robert B. McMaster and Pétra Noble*  
*University of Minnesota*



## The Great Britain Historical GIS

Work on what was to become the Great Britain Historical GIS (GBHGIS) began in 1994 when Ian Gregory teamed up with Humphrey Southall at Queen Mary, University of London. In previous research with David Gilbert, Southall had created a large database of nineteenth-century demographic statistics. This consisted of census data published every ten years, vital registration statistics from the *Registrar General's Annual Reports* and *Decennial Supplements*, Poor Law statistics from the Returns of the Poor Law and Local Government Board usually published bi-annually, and an assortment of other data published at a variety of dates. Their initial aim was to create a GIS holding the changing boundaries of the administrative units to which these data referred.