

# **Conflation and the National Historical Geographic Information System**

## **A Case Study**

David Van Riper  
Research Fellow  
Minnesota Population Center  
University of Minnesota

November, 2008

The National Historical Geographic Information System (NHGIS), housed at the University of Minnesota's Minnesota Population Center, is a National Science Foundation-funded project to digitize and disseminate historical United States census data. There were two main parts of this project. First, the Minnesota Population Center collected as much of the historical census data as possible and created a website to disseminate the information. Second, we created shapefiles that depict historic census tracts and county boundaries for the entire country. Scholars may combine the attribute data with the boundaries and analyze spatial and temporal trends in migration, ethnicity, health, income and other topics at multiple scales (e.g., national, state, county, neighborhood) over 200 years. More information on NHGIS can be found at [www.nhgis.org](http://www.nhgis.org).

The historical census tract and county boundaries that are a product of NHGIS were created from the Census Bureau's TIGER/Line Census 2000 files. Since those files were released, the Census Bureau started the Master Address File/Topologically Integrated Geographic Encoding and Referencing Accuracy Improvement Project (MAF/TIGER AIP). This project will improve the accuracy of spatial features (roads, water, etc.) used to create census tracts and other census units. Unfortunately, this accuracy improvement has serious implications for the NHGIS boundary data. NHGIS leveraged features in the 2000 TIGER/Line files to construct historic census tract and county boundaries. We re-used boundaries if they remained constant between decades and used existing TIGER/Line features—such as rivers and roads—when they formed historic boundaries. If these features move because of the improved accuracy, the historic boundaries will not overlap with entities from the 2010 census or subsequent censuses. The realigned NHGIS boundaries will help people extend their studies of spatio-temporal trends and patterns to future decennial censuses and future releases of yearly American Community Survey data. A scholar who has analyzed changes in neighborhood composition between 1990 and 2000 can extend that analysis to 2010 because the historic boundaries will be realigned with 2010 boundaries.

In light of these more accurate TIGER/Line features, we developed a plan to conflate the historic census tract and county boundaries to the new features. ESEA's MapMerger ([www.esea.com](http://www.esea.com)) was chosen as the software to assist with the conflation process. As an ArcMap extension, MapMerger allows us to work within a familiar environment and to work with data geodatabases. It also allows us to work at the county or state level. MapMerger's tools automate much of the boundary alignment process, enabling the project to be completed in less time than would otherwise be required. The scope of the project is as follows:

### **Census Tracts**

For 1990 and 2000, the entire country is subdivided into census tracts. There were approximately 64,866 census tracts in 2000 and 60,513 census tracts in 1990.

For the 1910-1980 censuses (tracts were first used in 1910), only counties that participated in Metropolitan Statistical Areas were tracted. As we go back in time, we have fewer and fewer census tracts to conflate.

### **Counties**

The Census Bureau has tabulated data for counties since the first census in 1790. Thus, we will conflate county boundaries for all censuses from 1790 to 2000. Luckily, there are many fewer counties than census tracts! There were approximately 3,100 counties in 2000.

Numerous benefits will accrue from realigning the NHGIS county and tract boundaries with the new TIGER/Line features. Accurate spatial datasets are becoming increasingly important as research begins to exploit new technologies. For example, without the proposed realignment, highly precise spatial data collected from handheld GPS units or satellite imagery may be incompatible with existing NHGIS boundaries. The realigned boundaries will help users integrate new data with the historic census tracts and counties.

Thus far, we have developed and implemented a methodology for conflating census tract boundaries. First, we unioned all census tract polygons for all decades in the United States together and converted the results to a line feature class. Second, to each line, we assigned the decades in which it served as a census tract boundary. Third, we obtained TIGER/Line shapefiles for all counties that have finished the Accuracy Improvement Project from the US Census Bureau's website. Fourth, we created a file geodatabase for each state and imported the new TIGER/Line shapefiles into the appropriate geodatabase. Fifth, we split apart the lines from steps one and two into county tiles and imported the tiles into the appropriate state geodatabase. Sixth, we used ESEA's MapMerger tool in ArcMap to conflate the census tract boundaries we originally created to the more accurate spatial features from the new TIGER/Line shapefiles. Seventh, we created polygons from the MapMerger output and determined whether the number of output polygons matched the number of expected polygons. If the counts failed to match, we would return to ArcMap and fix the problems. Eighth, after conflating all census tracts for a given state, we merged the conflated census tracts together and used ArcGIS topology to check for holes and overlaps. If there were any topology errors, we would return to ArcMap and fix the problems.

We started the conflation processing in July 2008, and the project is scheduled to finish in 2011.

### **Contact Information**

ESEA Corporation  
280 Second St., Suite 270  
Los Altos, CA 94022  
Tel: (650) 941-4175  
Email: [esea@esea.com](mailto:esea@esea.com)  
Web: [www.esea.com](http://www.esea.com)