Planning for the Future, Building on Decades of Success: The Role of Geospatial Data at the Census Bureau[1]

Michael R. Ratcliffe and Dierdre Bevington-Attardi
Geography Division
U.S. Census Bureau
Washington, DC 20233 USA
Michael.r.ratcliffe@census.gov
Dierdre.bevington.attardi@census.gov

Introduction

The task of collecting, tabulating, and presenting demographic and economic census and survey data requires accurate and current geographic information. Census and survey operations, frame development, and sample design require address and locational information for housing units and businesses to ensure questionnaire delivery and accurate geocoding of results. Geographic areas provide the framework for tabulation and presentation of statistical data as well as the management of data collection activities. Roads and other landscape features provide the underlying spatial framework for address and address range maintenance and delineation of geographic areas. Geospatial databases and geographic information systems (GIS) are critical to effective management of the spatial data needed to support data collection, tabulation, and dissemination activities.

The Census Bureau has been a leader in the collection, production, and dissemination of geospatial data, particularly since the development of the Topologically Integrated Geographic Encoding and Referencing (TIGER) database for the 1990 census. Over the past 25 years, the use of, and need for, geospatial data has expanded to all levels of government as well as throughout the commercial and non-governmental sectors. Within the Census Bureau, use of GIS and geospatial data for statistical data analysis and review, planning and coordination of data collection activities, and mapping and data visualization products, also has expanded substantially. These expanded uses and needs have raised several challenges:

1) A need for greater positional accuracy to support the use of global positioning systems (GPS) in field data collection operations and to ensure accurate geocoding of data.
2) Frequent and timely updates of critical geospatial data, particularly in areas experiencing change.
3) Varied and flexible geospatial data products to meet diverse user needs and to support access to geospatial data through web services.

[1] This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the authors and not necessarily those of either the U.S. Census Bureau.

In this paper, we discuss the various activities at the Census Bureau to meet these challenges. We provide background information about the development of the Master Address File (MAF)/TIGER database to provide context. We discuss initiatives to collect and process geospatial data to ensure that data in the MAF/TIGER database are as current and accurate as possible. We also discuss the uses of geospatial data throughout the Census Bureau, and the decentralization of development of mapping and data visualization products.

MAF/TIGER Database

The Census Bureau's MAF/TIGER database contains all the address and spatial data necessary to support census and survey programs, including mapping and production of geographic information products to facilitate review and analysis of statistical data. There are few statistical data collection and tabulation programs at the Census Bureau that do not rely on MAF/TIGER in some manner.

The TIGER database developed out of the joint Census Bureau and the US Geological Survey (USGS) effort in the early 1980s to develop a "computer-driven" system to support large volume production of high-quality maps to support each agency's map publication programs. This mutual exchange between TIGER and the USGS National Mapping Program in the early 1980s was designed to produce more accurate Digital Line Graphs (DLGs), topographic maps, and census mapping products for survey operations, boundary update programs, and cartographic reference publications. "A digital map of the nation" containing transportation and hydrographic features for the conterminous United States was created by scanning, digitizing and classifying (or naming) features derived primarily from 1:100,000 scale USGS map sheets (1,823 total) as well as the Census Bureau's 1980 GBF/DIME files for the urbanized portions of metropolitan areas. A wide variety of sources of varying accuracy and quality also were used as supplemental information to record the "relative" location of features, with varied, undocumented "locational" accuracy.

Differences in needs and timelines for implementation led to divergence in the development of a single digital map of the nation. The Census Bureau needed TIGER to support activities for the 1990 Census, putting development on a much faster timeline than was needed by USGS. Further, at that time, Census Bureau field data collection and data tabulation programs needed only relative accuracy of geographic information, whereas USGS's geoscience needs required positional accuracy. In other words, from the Census Bureau's operational standpoint, as long as housing units were geocoded to the correct side of the street, it did not matter whether the street was represented in its positionally accurate location on the face of the earth. Relative accuracy was sufficient for the 1990 and 2000 decennial censuses, but would become increasingly insufficient as local and state governments and other data users relied upon TIGER data for use in their own geodatabases. It also was insufficient as the Census Bureau began to make plans for use of GPS in the collection and recording of housing unit locations for the 2010 Census.

When it was released for use in the 1990 census TIGER contained a latitude/longitude-coordinate value for each of more than 30 million feature intersections and more than 42 million feature segments that outlined more than 12 million polygons. This database was the first digital seamless map of the United States, Puerto Rico, and the Island Areas generated with such unprecedented detail. In addition, an automated address matching capability was added to the TIGER System to geocode residential addresses to the correct collection geography. TIGER/Line extract files were available for the first time for the public as ASCII files. The geographic codes and names in the TIGER database enabled some very ambitious statistical mapping programs and spatial analysis using GIS applications. For that census, and each subsequent decennial and economic census, large amounts of socioeconomic data were visualized instantaneously for every geographic entity delineated and stored in the TIGER database.

In addition to the recent database merger of over 150 million addresses from the Census Bureau's MAF, there has been a corresponding rise in the number of geospatial data products and online mapping tools and web services that utilize MAF/TIGER database (note that addresses and address points are available only to users and analysts within the Census Bureau). These products, tools, and services have brought more changes—demands for greater accuracy, frequent and timely updates, and increasingly varied and flexible digital platforms. The Census Bureau's 25-year old database, now over 25 terabytes in size, has an ongoing and evolving history.

Given its size, accuracy improvement programs and quality assurance for the MAF/TIGER database is ongoing. Because this is built on an inherently dynamic system, improving quality standards have remained the main focus for all geospatial initiatives at the Census Bureau. The decision to use Global Positioning System (GPS) software on handheld devices in field data collection operations and imagery analysis to detect and evaluate feature change necessitated a shift to greater positional accuracy, and less tolerance for inconsistency and error. This use has fueled the Geography Division's need to improve address coverage, spatial feature updates, and enhanced quality assessment and measurement through partnerships with state, county, and local governments.

TIGER was a homegrown system of data retrieval, interagency collaboration, and publicly accessibility that was to become a powerful business model and database source for burgeoning GIS businesses. The GIS community that was beginning to debut on a global stage was able to incorporate a massive stockpile of geospatial data that was low cost, unrestricted and public domain, and built systems designed for analysis of environmental, economic, infrastructure, and homeland security issues. TIGER/Line files have served as a primary source of geospatial information within the GIS community, and in turn, census operations designed to update and maintain TIGER data have adapted to changes in GIS technology. The demand for additional refinements and modernization of this database grew into a strategic plan that was formalized with a vision in October 2000 for a redesigned MAF/TIGER system.

In a relatively short period, the MAF/TIGER database was transformed into a commercial off-the-shelf (COTS) database environment that allowed the Census Bureau to manage its geospatial information as an object-relational Oracle database and develop new applications with GIS software and hardware (including GPS mobile computing devices). The MAF/TIGER Enhancement Program (MTEP), conducted between 2000 and 2010, vastly improved field data collection and MAF/TIGER update capabilities in this new database environment, where street features were brought into alignment with GPS coordinates or locations based on imagery. This involved a comprehensive plan to migrate the data, assign coordinates to the spatial (TIGER) database through the MAF/TIGER Accuracy Improvement Project (MTAIP) and develop core applications to support partnership programs with state, local, and tribal governments. The simultaneous creation of a streamlined exchange of partnership shapefiles from these new applications eliminated the time-consuming two-step conversion process from ASCII files. Public access to these files has been an added bonus of this transformation that, in turn, created new opportunities for individual mapping and programming functions.

Building on the work of the previous decade, the Geography Division has implemented the Geographic Support System (GSS) Initiative, an integrated program of improved address coverage, continual spatial feature updates, and enhanced quality assessment and measurement through partnerships. State, local, tribal, and contractual partners have readily provided the means for maintaining and updating the MAF/TIGER database and further improved the quality and potential uses in federal, commercial, and non-profit sectors.

The MAF/TIGER database today is over 25 terabytes in size. In addition to base geospatial data, it stores boundary and attribute information for legal and statistical geographic entities for the US, Puerto Rico, and the Island Areas including:

- 6.7 million collection blocks (used by decennial census operations)
- Over 730,000 assignment areas for decennial census field workers
- 3,143 counties and equivalent entities
- 40,000 local governmental entities (i.e., cities, towns, townships, villages)
- Nearly 10,000 census designated places (unincorporated places)
- 78,000 census tracts
- 220,000 block groups
- 11.1 million census blocks

Uses of GIS at the Census Bureau

The development of TIGER, and in particular, the public availability of TIGER data via TIGER/Line files (since 2007 in shapefile format), helped spark the growth of the GIS industry in the United States. The transformation from manual methods to on-demand digital mapping had far-reaching impacts across disciplines. Geographers, statisticians, economists, and demographers at the Census Bureau use GIS to review statistical data prior to dissemination and analyze demographic and economic patterns and trends. In addition to data analysis, the availability of GIS has significantly enhanced census operations, data analysis, dissemination, and map production. In this section of the paper we outline ways in which GIS is used at the Census Bureau and discuss a few examples of such uses.

GIS is used at the Census Bureau in the following ways:

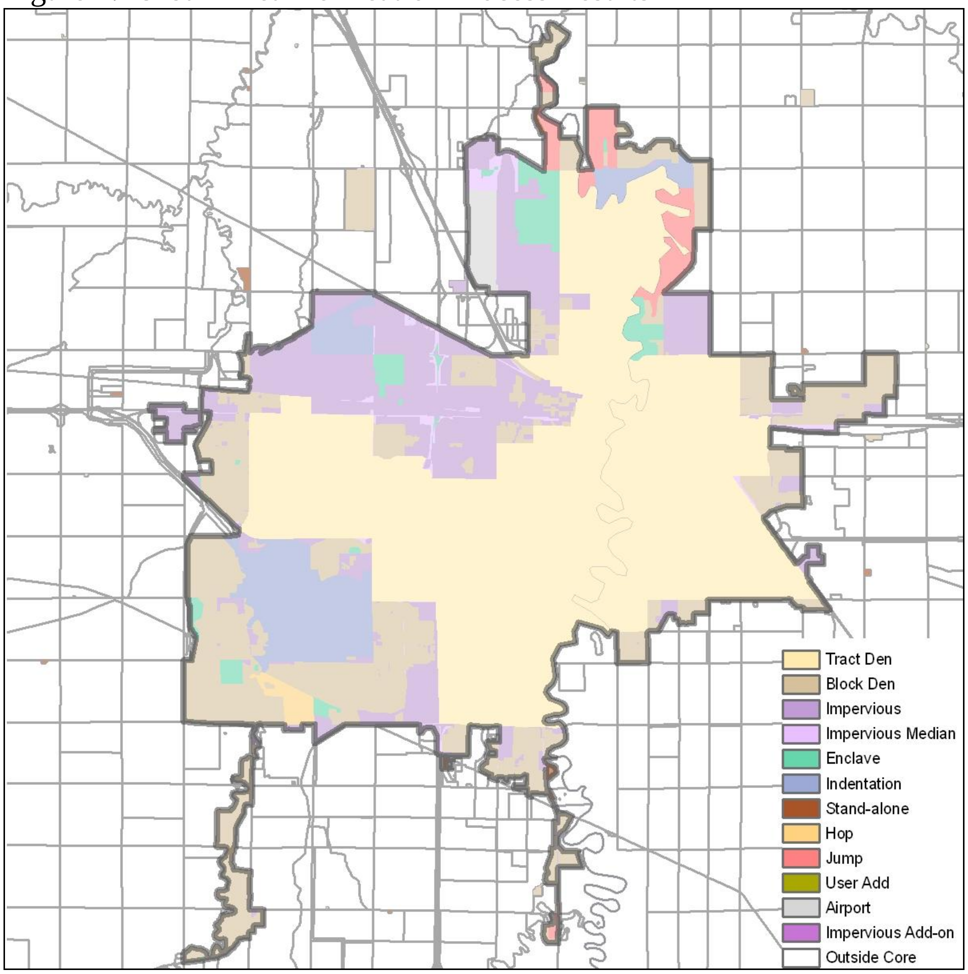
- Management of field data collection operations
- Interactive or automated delineation of geographic areas
- Review of address, spatial data, and boundary updates submitted by state and local government and other partners
- Review of statistical data prior to dissemination
- Analysis of statistical data for use in reports
- Map production, including development of interactive on-line mapping applications

Field data collection operations at the Census Bureau make use of GIS to 1) ensure efficient and effective assignment of field workers to data collection tasks and 2) delineate geographic areas used to manage, assign workloads, and monitor progress of a census or survey. The Census Bureau's Field Division uses GIS to relate residential locations of field workers to demographic survey sample cases to assign tasks and workloads. For example, staff planning for an upcoming 2020 Census test correlated the location of potential sample blocks identified in several different sample designs to residential locations of current field staff in order to determine whether the test could be carried out using existing staff. Census Bureau staff used GIS to map the locations of potential sample blocks within 30, 45, and 60 mile buffers around census blocks where field workers reside. In addition, staff identified the field worker nearest each potential sample block in order to determine the number of blocks assigned each worker to avoid excessive workloads.

Geography Division staff use GIS daily to review and edit address, feature, and boundary updates prior to insertion into the MAF/TIGER database. To facilitate this work, Geography Division often utilizes customized software written specifically for use in commercial off-the-shelf GIS packages. For example, to facilitate the acquisition and review of updates to boundaries for American Indian areas for the 2010 Census, Geography Division staff developed extensions to ArcGIS that could be used by GIS-enabled tribes to review, update, and submit boundary changes as part of the Tribal Statistical Areas Program. Other geographic area update programs prior to the 2010 Census, such as the Boundary and Annexation Survey, the Participant Statistical Areas Program, and the School District Review Program, utilized the MAF/TIGER Partnership Software (MTPS)—a purpose-built, self-contained GIS for use by program partners with varying GIS access and capabilities. The MTPS contained all of the information needed to review and update geographic area boundaries and attributes within the respective update program, and did not require the user to have specialized GIS training or experience.

For the 2010 Census, Geography Division also used GIS to automate delineation of urban areas and ZIP Code tabulation areas (ZCTAs). Automated geographic area delineation software also was used to produce initial boundaries for geographic areas by which to manage census operations; final boundaries were produced through interactive review in a GIS. For the 2010 Census Urban Area Delineation Program, geographers used GIS to compile and integrate various input data sets, such as 2010 Census population counts at the census block and census tract levels, selected layers from the National Land Cover Dataset (NLCD), and boundaries for airports. Geographers used ArcGIS 9.x software, employing Visual Basic for Applications (VBA) script and using ArcObjects to automate the delineation process, applying published criteria to relevant data. Information relating to each stage of the delineation process were stored within the GIS program, both for quality assurance review and for later retrieval to answer questions about the delineation process and results. Figure 1 below depicts the various stages in the urban area delineation process at which census tracts and census blocks qualified for inclusion in an urban area, as stored and utilized in a GIS.

Figure 1. Urban Area Delineation Process Results

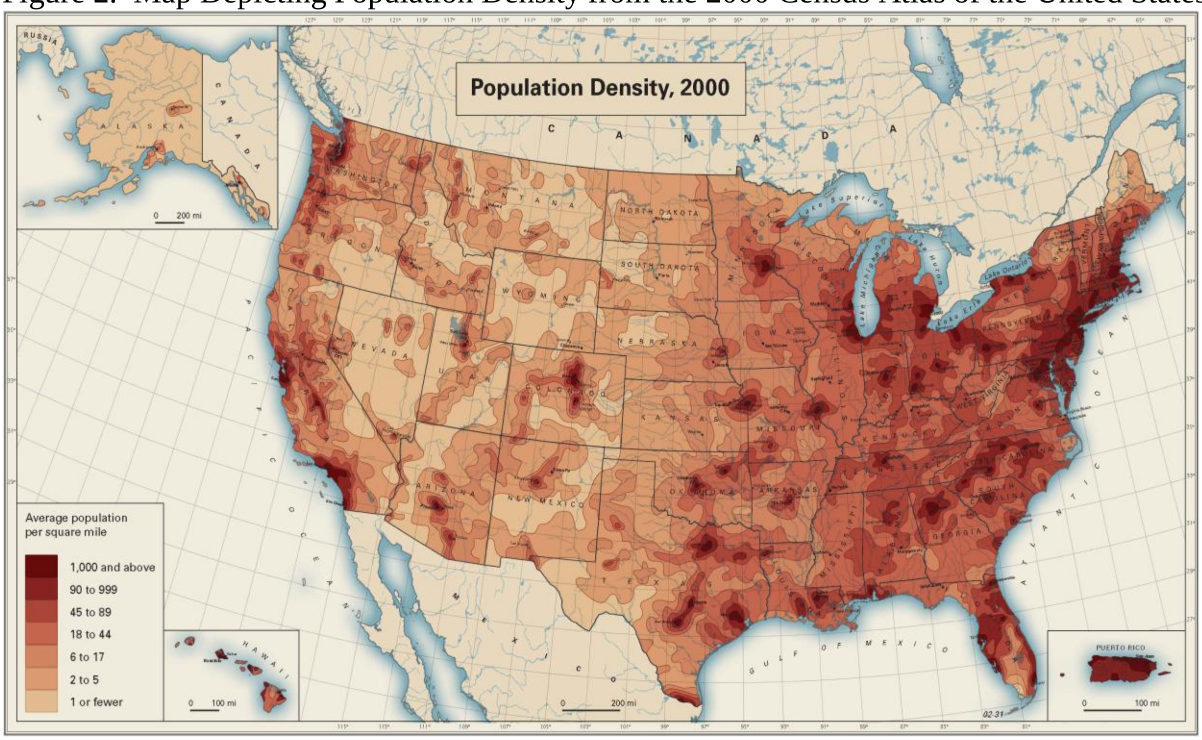


The ZCTA delineation process utilized GIS in two stages. First, in order to ensure a closer relationship between the US Postal Service's ZIP Codes (which exist as linear delivery routes) and census block-based ZCTAs, geographers used GIS to build a network of Thiessen polygons which were then dissolved to create a framework of lines separating clusters of housing units with different ZIP Codes. This process added approximately 36,000 non-visible lines to the TIGER database for use as census block boundaries and then ZCTA boundaries. The second stage of the process utilized GIS to automate ZCTA delineation, enable interactive review, and then produce files for upload into TIGER.

An early example of the use of GIS to review statistical data involved review of annual county-to-county migration patterns as part of the Census Bureau's Population Estimates Program. Demographers in the 1990s had noticed anomalous patterns of high rates of in-migration or out-migration for individual counties in the source data derived from the address information on multiple years of tax returns. Using GIS, they analyzed in-migration and out-migration data and detected spatial patterns that stemmed from filing practices of commercial tax preparers that created the impression that individuals had moved from one county to another. After identifying the problem, demographers were able to develop methods to correct the anomalies within the population estimation process. Geographers and analysts within the Population Estimates program routinely use GIS to review changes in boundaries in relation to base population and housing unit data. Census Bureau researchers use GIS to prepare reports and other analytical materials following data tabulation and production.

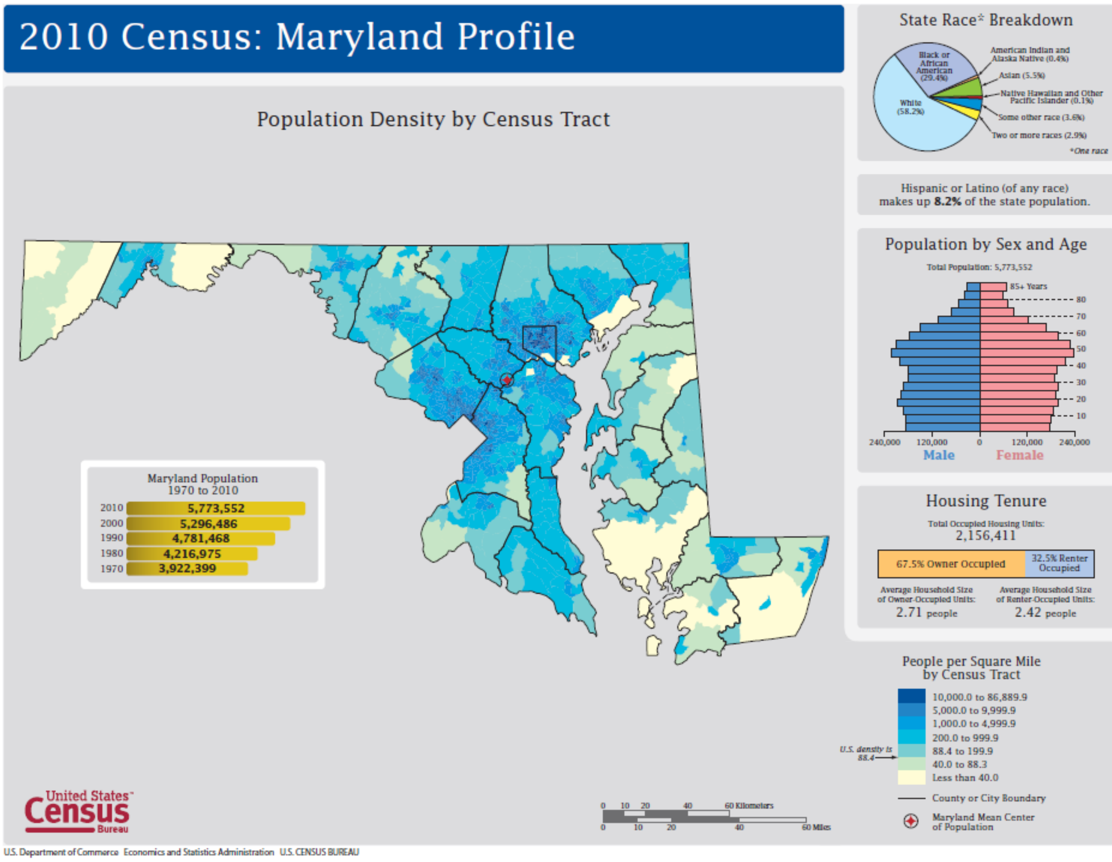
One of the more exciting map ventures from Census 2000 was the return of the statistical atlas with publication of *Mapping Census 2000* and *The Census Atlas of the United States*. Each atlas was produced primarily by the Census Bureau's Population Division with support from the Geography Division and Housing and Household Economics Division. As themes or "story lines" were developed for these products, the use of GIS featured prominently in the data analysis and map production for each atlas. Figure 1 below depicts population density in the United States and Puerto Rico for Census 2000. The map design was modeled after similar designs appearing in Census Bureau statistical atlases and reports from the late-19th century and was presented alongside a corresponding map from the 1880 Census.

Figure 2. Map Depicting Population Density from the 2000 Census Atlas of the United States



The 2010 Census has used GIS in support of thematic map production highlighting data from the American Community Survey and other data tabulation programs. This includes hundreds of thematic maps appearing in Census Bureau press briefs, reports, and other publications or as standalone products. Examples of thematic maps from the 2010 Census produced using GIS include the Geography Division's maps depicting population change from 2000 to 2010; population density by county; and population profiles, which include a combination of graphics and mapped data (Figure 3).

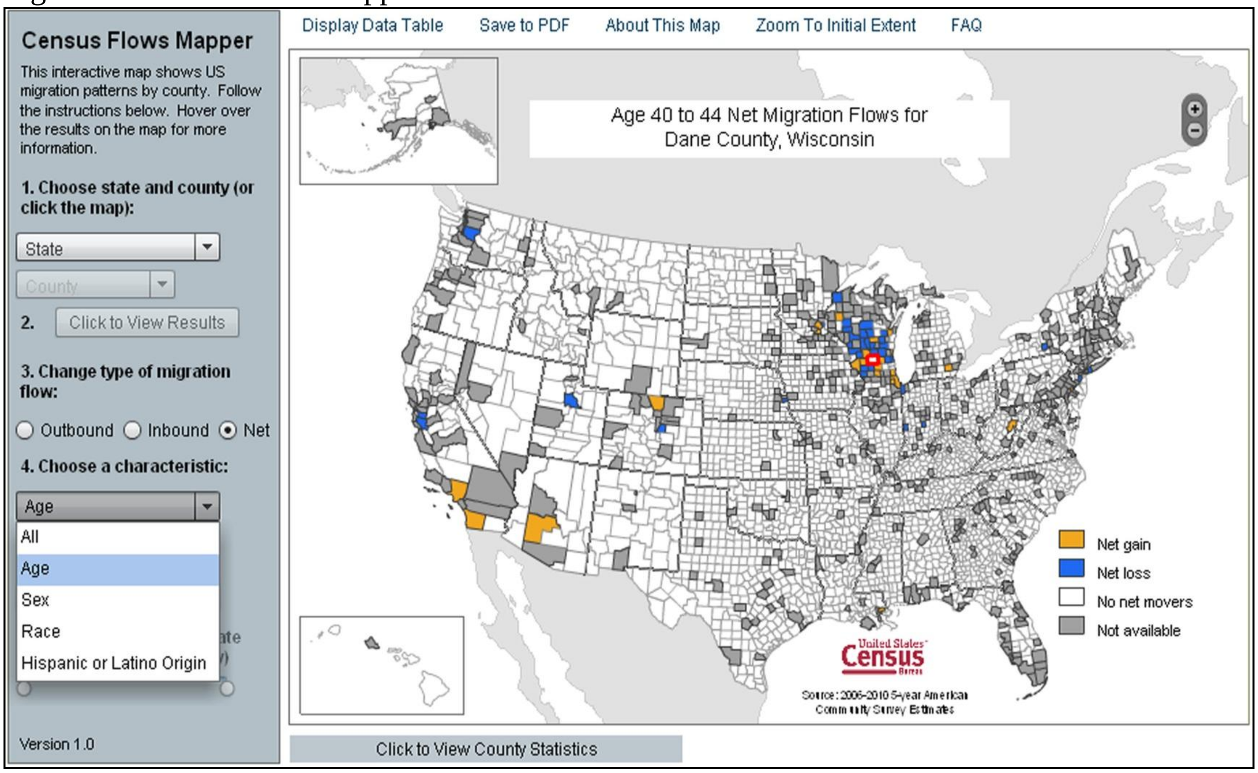
Figure 3. 2010 Census Population Profile: Maryland



GIS has been integral in the production of interactive maps. Online tools offer greater flexibility for the data user, generally providing the ability to access data for a particular variable at multiple levels of geography or multiple variables for a single level of geography. In this regard, GIS has enabled the Census Bureau and other organizations to disseminate data via thematic maps and map interfaces. As with static thematic maps, interactive maps have been produced by various offices and divisions throughout the Census Bureau to support a particular program or partnership.

Census Bureau online tools to visualize statistical data have been developed using mapping tools scripted from computer code (i.e., HTML, Javascript, CSS, Flex) in combination with the Census Bureau's API (containing data from the 2010 Census and American Community Survey). Tools like the *Census Data Mapper*, enable users to create county-based demographic maps of population, race, ethnicity, family, and households from the 2010 Census. The *Census Flows Mapper* provides users with an interface to produce hundreds of county-based migration flow maps based on data from the 2006-2010 American Community Survey (Figure 4). The *Metropolitan/Micropolitan Thematic Map Viewer* visualizes demographic variables by census tract from data tabulated in a 2010 special report on the largest U.S. centers of population and economic activity. *The Small Area Income and Poverty Estimates (SAIPE) Interactive Data Tool* visualizes data for annual income and poverty by age for states, counties, and school districts. The Census Bureau's *OnTheMap for Emergency Management* web application provides first-responders with tools to analyze employment and demographic data in association with hurricanes, wildfires, floods, disaster-declaration areas, and winter storms.

Figure 4. Census Flows Mapper



Challenges for the Future

At the beginning of this paper, we noted several challenges resulting from increased use of GIS both inside and outside the Census Bureau:

1) A need for greater positional accuracy to support the use of GPS in field data collection operations and to ensure accurate geocoding of data.
2) Frequent and timely updates of critical geospatial data, particularly in areas experiencing change.
3) Varied and flexible geospatial data products to meet diverse user needs and to support access to geospatial data through web services.

The Census Bureau's Geography Division, through MTAIP, and working with state and local government and other partners, substantially improved the positional accuracy of roads and other features within the MAF/TIGER database prior to the 2010 Census. This facilitated the use of GPS to collect housing unit locations for the 2010 Census. This work continues through the GSS

Initiative, with an expanded focus on address, geocoding, and feature quality. The Geography Division uses GIS on daily basis to review data inputs as well as to measure and assess the quality of geospatial data contained in the MAF/TIGER database.

In addition to continuing to improve the positional accuracy of data with the MAF/TIGER database, there is a need for more frequent and timely updates of spatial data. This is necessary to support the Census Bureau's annual statistical data collection programs, such as the American Community Survey, the Current Population Survey, and numerous other demographic surveys. The need to achieve greater efficiencies in survey operations requires address information, housing unit locations, and feature data that are as up-to-date and positionally accurate as possible. The need to achieve greater efficiencies and contain costs extends also to planning for the 2020 Census. Research efforts currently underway are focused on efforts to acquire and maintain as complete and accurate an address list as possible in order to reduce the need for address list validation and updating in the field. Other research efforts are investigating the "life cycle" of address information, updates, and changes—essentially tracking when a particular address entered the Census Bureau's MAF and how many times the address may have been reviewed and changed through decennial census operations—as well as methods to efficiently manage non-response follow-up (NRFU) operations. Each of these research efforts are utilizing GIS to map and analyze information and develop spatially efficient assignment of workloads.

With regard to data dissemination, the challenge for the future is to support access and consumption of geospatial data via web services, thus moving away from development and dissemination of a variety of discrete, and sometimes unique, data products. To this end, the Geography Division has developed TIGERWeb, which offers both an online viewer of features and geographic area information within the TIGER database as well as a web mapping service (WMS). TIGERweb eliminates the need for users to download and manipulate TIGER/Line shapefiles in order to view and query the boundaries of census geographic areas and features such as roads, railroads and rivers, lakes and other larger bodies of water. It currently displays the all 2010 Census legal and statistical area boundaries with related names and codes. In addition, it contains population and housing unit counts from the 2010 Census for each of the geographic areas. The TIGERWeb WMS offers the means for developers to directly consume geospatial data in applications. The challenge for the Census Bureau is to modify existing software and databases to create the architecture necessary to support service-oriented approaches to offering and consuming data.

The expansion of GIS throughout the Census Bureau in the 1990s coincided with expanded use of GIS throughout social science disciplines, in particular geography, demography, and sociology. While the Census Bureau actively sought to hire individuals with GIS experience particularly within the Geography Division, new staff in the demographic, economic, and statistical divisions of the Census Bureau also expected to be able to use GIS in the day-to-day work. The impact of GIS on the Census Bureau's mapping program demonstrates the ubiquity of GIS but has also proven to be the greatest challenge. While the ease of use of GIS software and an enterprise license agreement with Esri has enabled mapping on a wider scale throughout the organization it also has made it easier for those without any formal cartographic training to produce maps. In addition, as mapping became decentralized, it became more difficult to know which individuals or offices were producing maps. The program was challenged to avoid duplication of effort, to create standards and best practices for data visualization, and to establish a common “look and feel” to Census Bureau maps.

Map quality has been addressed through the formation of the agency-wide Mapping Coordination Group consisting of individuals engaged in map production. This group exists to develop and share best practices, review draft maps and mapping applications prior to public release, and keep each other informed about map products. To improve the quality of thematic maps cartographers developed a mapping template for use in ArcGIS.and incorporated guidelines in cartographic design, Efforts also are underway to develop an enterprise GIS to further facilitate and ease access and use of GIS and other tools enabling geospatial analysis.

Conclusion

As GIS users transition from enterprise to desktop solutions, individuals and organizations have been empowered by an elaborate system of online services, data retrieval, and mapping that can now be accessed on multiple platforms—a personal computer, smart phone or mobile device. Census Bureau databases have adapted to these developments through the decades and supported a seamless transition to 21st century mapping and analysis. The MAF/TIGER database and its impact on GIS and statistical mapping has fueled the massive deployments of geospatial resources witnessed during census operations and fostered the Census Bureau’s mission. As the Census Bureau continues to serve as the leading source of quality data about the nation’s people and economy through innovations in data dissemination and analysis, the challenge is to maintain and improve the accuracy and currency of geospatial data in the MAF/TIGER database for use in geographical analysis and GIS both inside and outside the Census Bureau. Further, data dissemination strategies must keep current with our customer’s preferences and expectations regarding ways to access and consume data, primarily through delivery of geographic information via web services. Building from its long history of leadership in the development of geospatial databases, data visualization methodologies, and statistical mapping, the Census Bureau continues to innovate, as well as leverage the latest technologies and tools, to produce high quality geographic information.