PLANNING FOR THE FUTURE: THE GEOGRAPHIC SUPPORT SYSTEM INITIATIVE FOR THE U.S. CENSUS BUREAU

Tim Trainor
Geography Division, U.S. Census Bureau, Washington, DC. 20233 – timothy.f.trainor@census.gov

KEY WORDS: U.S. Census Bureau, GIS, addresses, maps, MAF, TIGER, geospatial

ABSTRACT:

In the years between the 2000 and 2010 censuses, the Geography Division at the U.S. Census Bureau implemented a program to improve the positional accuracy of the Topologically Integrated Geographic Encoding and Reference (TIGER) file. The MAF/TIGER Accuracy Improvement Project (MTAIP) was a huge success that the Census Bureau intends to build upon.
Beginning in Fiscal Year (FY) 2011 the Geography Division of the U.S. Census Bureau has put forth an expanded Geographic Support System (GSS) Initiative. The GSS Initiative will focus on improving the Census Bureau's address coverage and quality while maintaining the significant spatial data in the MAF/TIGER database.
The GSS Initiative calls for the following: in FY 2013 and beyond, expanded use of the Community Address Update System (CAUS) for rural addresses; partner and commercial address lists as source data in all areas; first-time use of the USPS DSF in Puerto Rico; increased use of imagery; use of change detection techniques; continuous street and other feature improvements through maintenance programs and change detection techniques using Geographic Information System (GIS) files and imagery provided by partners; and quality measurement and validation of address and spatial data from field work and the use of imagery and Global Positioning System (GPS) control points.

1. BACKGROUND

The U.S. Census Bureau's TIGER File has served as a valuable spatial data source since the late 1980's. The original TIGER System was built by a small army of dedicated developers who designed and programmed a new type of spatial database and all of the applications necessary to access, update, and use TIGER spatial data and associated attributes. Commercial databases were not designed for spatial data nor were early GIS commercial software programs able to recognize the internal structure of the TIGER database. Applications were written in computer code for all geographic functions including feature, boundary, and attribute updating, geocoding, mapping, geographic processing, database management, and production. The 1980 and 1990 censuses used a temporary national address list that was a separate database and was linked by common geographic codes. Unlike the earlier address control files, which were not maintained, the development of a Master Address File (MAF) for Census 2000 was continually updated to support the annual American Community Survey throughout the recent decade leading to support for the 2010 Census.

The geography of the earth is constantly changing. These changes, resulting from both natural and man-made events, are continuous, irregular, local, and most often unpredictable, particularly from a national perspective. For instance, the impact of man-made activities such as the expansion of the boundaries of a town or city or the construction of a new housing subdivision have ancillary effects on addresses and other geographic features depending on the type of change. The effect of physical changes caused by unpredictable weather on features affects a census or statistical survey and is not an ordinary activity, which also is generally a local event. To successfully conduct a national census or survey these changes must be taken into consideration.

Changes on the landscape often results in new addresses for new housing that are added to the Census Bureau's MAF. Addresses take on various forms and are used for different purposes. For Census Bureau use, at least four basic requirements are needed. The address must be mailable, deliverable, locatable, and geocodable. For each address, knowing if it is mailable, that is, it meets U.S. Postal Service (USPS) minimum requirements as a mailable,address is paramount. Over 92% of housing units in the U.S. have a mailable address in the MAF. A mailable address does not assure that the mail is delivered. Having a record of whether each address is deliverable according to USPS rules and local post office delivery practice is needed to assure that a questionnaire will reach its intended housing unit destination. Where mail is not delivered using a house number, street name, and town name convention, a different approach to addressing is required. In these cases, a descriptive address is needed to visually locate the housing unit by a field lister and/or census enumerator. This descriptive address type is needed for non-deliverable and deliverable addresses. The descriptive address meets a third requirement that each address is locatable.

Lastly, an address on its own does not provide its location on the ground. Said another way, where is "101 Main Street"? Address location is accomplished in TIGER by assigning a geographic code to each address that associates that designation with a census block. In the example "101 Main Street", there could be more than one instance of such an address, even with the same town name. By assigning a specific geocode comprised of the state, county, census tract, and census block, each address is distinguished from another by such geocodes.

Another way to uniquely identify a housing unit is through the unit's specific geographic location using latitude/longitude. During the 2010 Census Address Canvassing operation, approximately 140,000 census listers

checked addresses that were provided from the MAF in preparation for the questionnaire mailout. While conducting the address check, the listers captured geographic locations of the housing units using a specially designed mobile device that was equipped with GPS capability. The use of geographic coordinates to locate housing units, independently or in combination with the mailing address, in subsequent census and survey field operations requires further study and planning.

Updates to TIGER are also the result of changes driven by technology. With the advent of GPS, enumerators acquired housing unit locations as they canvassed the nation. In future censuses and surveys, enumerators can correctly locate housing units that did not respond to the census or survey through the use of GPS technology and geographic coordinates stored in TIGER combined with an address list, both of which are now stored in a relational database called MAF/TIGER (MTDB). In rural areas described above, the GPS is an accurate aid over previously used physical descriptions (white house with green shutters on corner) and hand drawn map spots of house locations on Census Bureau provided maps.

Use of precise locations of housing units requires that other spatial database features have similar location quality. Since 2003, the U.S. Census Bureau has been improving the locations of street centerlines in TIGER to an accuracy of 7.6 meters CE95 or better. This national effort ensures correct relationships for the assignment of housing units to the smallest level of census geography so that the GPS quality coordinates are on the correct side of each street. Each questionnaire can then be assigned the correct geographic area codes based on its coordinate location.

2. THE NEED FOR AN EXPANDED U.S. CENSUS BUREAU GEOGRAPHIC SUPPORT SYSTEM (GSS)

The GSS Initiative is needed to improve the coverage and quality of addresses for future censuses and surveys conducted by the U.S. Census Bureau. Oversight groups and research bodies such as the General Accountability Office (GAO), Office of Inspector General (OIG), and National Research Council, have repeatedly reported on the need to continually update and maintain the Master Address File (MAF) (2004, 2006, 2008).

The annual American Community Survey (ACS) is a household level survey. Housing without an address in the MAF is not considered for inclusion in the ACS. New construction, including entire housing subdivisions in this circumstance, is not included in a survey that offers detailed characteristics about the nation's population at a local level. Therefore to ensure a comprehensive address list to sample from, addresses without a city style address (house number, street name), particularly evident in rural areas, need to be updated throughout the decade.

Addressing in Puerto Rico poses unique challenges based on the characteristics of the address and the way in which addresses are assigned. Closer collaboration with the USPS is planned supplemented with improvements in understanding the characteristics of addressing and mail delivery in Puerto Rico.

Efforts at assuring a more complete and accurate census or survey have lasting impacts on statistical data users. The commitment to provide high quality geographic products, which includes addresses, to Census Bureau statistical programs is of paramount importance. Examples of benefits to public data users include:

- Relevant and accurate reporting of demographic and economic statistics
- A more accurate picture of the United States and its population for decision-makers (at all levels of government) and commercial enterprises
- Resources for state and local government planners in their analysis of population growth and change for identifying the need for types of services, new amenities, and schools
- Ensures data currency by maintaining a national address list and capturing new street growth and halts the degradation of the existing street network's positional accuracy and attributes

3. WHAT IS INCLUDED IN THE GSS INITIATIVE?

The GSS Initiative is an integrated program of improved address coverage, continual spatial feature updates, and enhanced quality assessment and measurement. It supports a targeted, rather than a full address canvassing operation during 2019 in preparation for the 2020 Census. The three tenets of the initiative focus on address coverage, maintenance of the feature network, and improvements in quality.

Improved address coverage includes exploring methodologies to achieve complete and current address coverage, concentrating on rural areas, Puerto Rico, and Group Quarters, and improving geocoding of all addresses to their location and higher levels of geography. Planned activities include:

- Targeting specific areas of weakness in the MAF.
- Initiating programs with partners to continuously receive addresses throughout the decade; includes developing tracking, processing and reporting systems to facilitate address receipt and processing
- Identifying commercial address list sources; conduct quality reviews
- Upgrading the Community Address Update System (CAUS) software to facilitate the future expansion of rural address collection and enable hardware for Global Positioning System (GPS) rather than manual collection of housing unit locations

Feature improvement includes continuous updates of the street network to facilitate improving the successful assignment of addresses to their correct geography. By maintaining an accurate and complete street network the cost to the taxpayer is reduced by making census and survey operations more efficient by enabling census takers to find and enumerate assignment areas quickly. Accurate streets also improve census data by helping to ensure that housing units locations are within the correct census block for tabulation thereby assuring accurate allocation of the population. Planned activities include:

- Broadening participation in existing programs for receiving partner Geographic Information System (GIS) files and imagery (both critical as source data); includes

expanding functioning management, acquisition, distribution, and use systems
- Researching change detection techniques; includes review/analysis of existing commercial software [both street network and address update applicability]

Quality improvement includes broadening quality assessments and providing quantitative measures. This also includes geographic program evaluations that measure data quality and the effectiveness of systems to sustain the quality of geographic data throughout the decade. Planned activities include:

- Continuing to build the inventory of GPS control points to assess the positional accuracy of street locations
- Improving the quality of the data, IT processes and geographic data products.

With each decade in recent history beginning in 1960, mailing questionnaires to households has been the preferred method of data collection to save on the high cost of door-to-door enumeration. In assuring a complete census, a check of the address list is conducted during address canvassing where, in effect, the nation is divided into assignment areas and temporary listers check the address list. Listers check if the address is correct, make necessary changes where it is warranted, add new addresses, or delete non-existent addresses. While they are canvassing their assignment and, in effect, walking every road in the nation, listers add new roads and related information such as road names.

During the 2010 Census, for the first time, listers used mobile devices that were, in effect, their office. Many listers did not regularly visit one of the 494 Local Census Offices around the nation. Rather, they received their assignment area maps and address list usually wirelessly and spatial and address data were transmitted and read by the visualization and listing software. Field staff conducted their assignment and wirelessly transmitted the results to centralized processing offices that packaged data and transmitted the data back to headquarters, sometimes within the same day.

The 2010 Census plan included the use of GPS to capture the housing unit locations. In order to assure the association of highly accurate locations of each housing unit to its correct census block, the road network that is used to delineate census block boundaries had to be accurate as well. This goal shifted the traditional TIGER data reality of one of relative accuracy to one where features were positionally accurate.

As part of their set of tasks, listers walked to the front door of each housing unit and with the GPS enabled device, captured the house location. New streets also were added with GPS. The transmissions were added to the MTDB in preparation for the next set of census field operations.

It is estimated that a targeted address canvassing will realize significant cost savings. A goal of the initiative is a complete and accurate address list for each legally established community. This approach changes the focus from a nationwide canvassing organized by assignment areas to targeted areas outside of areas where the legal governments, agree with the census address list. The initiative will address the challenge of how governments will accept the coverage and quality of the census address list for their area of jurisdiction.

All initiative activities contribute to Master Address File (MAF) and TIGER database improvement. These activities build on the accomplishments of this past decade's MAF/TIGER Enhancement Program (MTEP) which improved the positional accuracy of TIGER spatial features, redesigned the MAF/TIGER database(s), established a partnership network, enacted the community address update system (CAUS) and emphasized quality measurement.

While a major focus is on rural areas, the GSS initiative also includes improvements to areas with city style addresses. For example, where addresses do not geocode, research will be conducted to understand the condition and reasons for the failure and will explore methods for improving the probability of geocoding similar address situations. Different addresses within a single unit sometimes pose challenges. Hidden housing typically is not obvious, for example, a basement apartment or a garage apartment, which are very difficult to identify and locate.

The 2010 Census included a concerted effort to include housing referred to as group quarters. Examples are numerous and include nursing homes, prisons, group homes, and college dormitories. Group quarters were not maintained in MAF maintenance programs because only the address was available on the primary source for the MAF. Special procedures are required for group quarters from assuring a sufficient number of questionnaires to knowing specific locations of each facility for field operations.

Topological principles are applied with the improved data the same way it was required previously. Persistent topology guarantees that correct geographic relationships are maintained in the MTDB. Geographic relationship principles are an important component of address and spatial data quality.

4. PLANNING FOR THE GSS INITIATIVE?

Planning for the GSS initiative began by forming ten working groups (WGs) on topics that focus on areas that require some level of research and development. The WGs include:

1. Policy
2. Project and Contract Management
3. Address Coverage and Sources
4. Feature Coverage and Sources
5. Quality, Assessments and Evaluations
6. Partnerships
7. Research and Development
8. Geocoding
9. MAF/TIGER Integration/Linkage
10. GPS

As geography impacts all censuses and surveys, members of each WG include staff from the Geography Division as well as staff from various Divisions all across the agency. In addition, the U.S. Geological Survey is actively participating on several WGs as well as the U.S. Postal Service. The National Geodetic Service within NOAA participates on the GPS WG. Without offering an exhaustive list of activities, at least one example topic of interest is presented for each WG.

4.1 Policy Working Group

The policy WG serves as a conduit between the other GSS WGs and Census Bureau management, including the agency's Policy Office. This group determines if issues presented to them are related to policy as well as whether or not an issue has legal implications. An example of a question the policy WG discusses is what are the acceptable uses of addresses as practices currently apply to constraints imposed under Title 13 of the U.S. Code. In this case, the policy WG evaluates the merit as a policy question and then discusses and recommends a course of action.

4.2 Project and Contract Management Working Group

The project and contract management WG provides guidelines for following best practices of project management, offers tools that enable communication among WGs and other staff members, establishes a governance process for reporting and decision-making, and supports and governs approaches for contract support for each of the WGs.

4.3 Address Coverage and Sources Working Group

The address coverage and sources WG focuses on improving address coverage on a continuous basis, through analysis and improvement of the Census Bureau's Master Address File. They work closely with other WGs to improve the MAF using partner-supplied and/or commercial address lists. Examples include state, county, local and tribal government supplied address lists, utility records, and other sources. Managing source, quality, vintage and other metadata characteristics at the address level requires a tracking system that not only documents the information, but also allows for its use in meeting the goals of the WG.

Complete address coverage includes the addition of addresses in high growth areas. Identifying these areas is accomplished using various approaches including harvesting information from sources that meet acceptable criteria. Change detection is another approach that is considered by the WG in identifying new housing and their address information.

There are occasions where fieldwork is needed to acquire, verify and evaluate address information. To pursue these cases, the WG develops criteria for selecting census blocks for consideration for fieldwork by staff in twelve decentralized offices across the nation. Areas for consideration generally are outside of good address coverage offered by the USPS delivery sequence file (a source that indicates the order in which mail is delivered) or areas that have addresses that are historically problematic.

4.4 Feature Coverage and Sources Working Group

The feature coverage and sources WG focuses on the analysis and improvement of features such as roads, railroads and hydrography in the Census Bureau's TIGER database. Current features are important for several reasons. Roads serve as the basis for geocoding an address. They are required for orientation in order to conduct fieldwork. They also serve as an anchor for delineation of boundaries of geographic areas. The WG collaborates with other WGs to identify high growth areas and to identify census blocks with new roads and new housing.

The WG determines criteria for evaluating sources of spatial data and works with other groups to identify potential spatial data sources from tribal, state and local governments, and commercial and other organizational data sources. With other WGs, they propose methodologies for working with partner supplied spatial data to update TIGER. This WG also identifies methods and sources to evaluate existing feature data in MAF/TIGER.

As the landscape of the nation changes, this WG looks for new ways of reporting change so that new feature data are identified and existing feature data are updated and improved as a result of new information and improved technology tools. Similar to the task of the address WG, managing source, quality, and vintage and other metadata characteristics at the feature level requires a tracking system that not only documents the information, but also allows for its use in meeting the goals of the WG.

4.5 Quality, Assessments, and Evaluations Working Group

The quality, assessments and evaluations WG is responsible for assessing the quality of the MTDB and of potential sources used to update the database. The WG utilizes data quality elements, measures, and procedures to assess the quality of the MTDB as well as potential sources for use in updating the database. Working with other WGs, the quality, assessments and evaluations WG identifies areas in need of updating and helps to determine address characteristics that best meet the needs for census and survey use. The WG uses the inventory of GPS control points to assess the positional accuracy of streets within the MTDB and the quality of potential spatial data sources. They also expand and implement spatial data source evaluation methods. The WG develops evaluation criteria and standards and specifies software applications and produces reports of summary assessments and evaluation data.

4.6 Partnerships Working Group

The partnership WG is responsible for developing relationships with national, state, county, local and tribal governments to foster the goals and progress of the GSS initiative. The WG initiates programs with partners to seek out the best address and spatial sources. They continuously receive address lists (address updates) and GIS files (feature updates) throughout the decade. They invite tribal, state and local governments to participate in partnering with the required primary source information. Where the best government sources are not available, the WG seeks sources from commercial enterprises and other organizations. In working with partners, the WG determines requirements and preferred options for effectively receiving partner supplied data and for communicating results back to the partners. To effectively manage information about each address and individual feature, the WG also develops requirements for systems that are used to track, process and report on partner-supplied address lists and GIS files.

4.7 Research and Development Working Group

The research and development WG conducts research and development to identify the best methods for updating the MTDB and furthering the goals of the GSS Initiative. The WG also researches emerging technologies such as change detection, crowd-sourcing, and volunteered geographic information. Where requirements warrant special technical considerations such as unique addressing conditions in Puerto Rico and Spanish language use on functional mobile GIS, the WG pursues technical solutions. As most of the WGs have a research and development objective, this WG coordinates, collaborates and complements their efforts by seeking technical solutions to create new capabilities or to improve existing processes. Examples include developing and implementing web-based interfaces for receiving partner-supplied address and spatial data; enabling partners to geocode, reformat, and upload their address lists; developing extensions to current data capture functionality that include collection of additional street attributes that are identified as integral to geocoding (for example, including low/high address ranges and zip code information). The WG prepares proposals for consideration, develops an approach for development, outlines a process for prototyping and testing, evaluates methodologies to determine success and proposes an implementation strategy.

4.8 Geocoding Working Group

The geocoding WG researches and identifies the best methods for linking housing units and their addresses to the correct geographic location. The Working Group develops requirements for a new improved geocoding system that assigns an accurate block code to every MAF unit. The WG develops in-office operations that support the use of GPS coordinates as preferred structure points for new housing units and assigns metadata and quality ranking values to each preferred structure point in the MAF. Where use of GPS is not feasible, the WG develops and implements in-office operations that support the continued collection of manual, relative-positioned coordinates. The WG researches existing ungeocoded addresses and determines root causes for determining the next course of action. The WG researches address standards and address standardization software and makes appropriate recommendations.

4.9 MAF/TIGER Integration/Linkage Working Group

The MAF/TIGER Integration/Linkage WG studies the consistency between address information contained in the MAF with similar and related information contained in TIGER. The WG determines methodology for ensuring consistency between the MAF and TIGER such as ensuring that street names within the MAF are consistent with those in TIGER and ensuring that the assignment of map spots to census block is the same in both the MAF and TIGER.

4.10 GPS Working Group

The GPS WG is responsible for the evaluation of GPS technology and how it applies to the goals of the GSS initiative. The WG monitors advancements and evaluates various devices/instruments used to capture GPS data.

5. WHAT IS THE SCHEDULE FOR THE GSS INITIATIVE?

The GSS initiative is a 10-year program that increases geographic support for the annual American Community Survey and other current surveys, ultimately leading up to the 2020 Census. Each census cycle requires significant effort to plan, prototype, test, and ultimately implement many nationwide operations. There is no room for failure once the census process is underway.

The first year of the initiative focuses on discovery of ideas for new requirements, improvements, and exploration of new capabilities. During the fiscal year, these ideas are pursued among the various initiative working groups. Proposals are made through an efficient governance process followed by initiation of tasks to test an idea, learn new technical solutions, prototype an approach or improve a current geographic support activity.

Activities in subsequent years focus on implementing developments that occur in FY2011. Priority is assigned to the greatest need and/or largest impact first followed by measurable achievements each year to reach the goal of nationwide coverage of activities. The mid-decade date represents a focal point for deciding if a targeted address canvassing approach is feasible with partnering tribal, state, and local governments.

The remaining fiscal years will continue production activities leading up to the various census geographic operations that will likely begin in 2017 in preparation for the 2020 Census.

6. CONCLUSION

The goal of the GSS Initiative is a continuously maintained and highly accurate address list that includes every household in the nation supports efficient and effective statistical data collection for censuses and surveys. The annual American Community Survey, which offers the detailed characteristics of the population, and on-going current surveys realize the benefits of a continually maintained address list. Partnering with governments across the nation to not only maintain the address list but to verify its credibility year-to-year obviates the need for a costly full address canvassing operation leading up to the 2020 Census.

The goals of the Census Bureau's GSS initiative are challenging but attainable. The initiative focuses on improvement to current approaches to maintaining geographic support while exploring new methodologies to efficiently work with partners by collecting and using important information that supports a national fabric of geographic data that builds on previous investments to conduct censuses and surveys.

7. REFERENCES

National Research Council (U.S.). Panel on Research on Future Census Methods.(2004*). Reengineering the 2010 Census: Risks and Challenges.* http://www.nap.edu/catalog.php?record_id=10959 (accessed Sept 13, 2010).

U.S. Department of Commerce, , Office of Inspector General. 2008. *Census 2010: Dress Rehearsal of Address*

Canvassing Revealed Persistent Deficiencies in Approach to Updating the Master Address File . http://www.oig.doc.gov/oig/reports/2008/OSE-18599.pdf (accessed Sept 13, 2010).

United States Government Accountability Office. 2006. *2010 CENSUS Census Bureau Needs to Take Prompt Actions to Resolve Long-standing and Emerging Address and Mapping Challenges* http://www.gao.gov/new.items/d06272.pdf (accessed Sept 13, 2010).

8. ACKNOWLEDGEMENTS

This paper reports the results of work undertaken by U.S. Census Bureau staff. This paper is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.