Using Location-Based Information to Improve Cultural and Demographic Statistical Data

Leslie Godwin
Assistant Division Chief, Geographic Program Management
United States Census Bureau, Geography Division

The United States Census Bureau first utilized location-based services in support of its 2010 Decennial census. Field Representatives carrying Hand Held Computers equipped with Global Positioning System technology captured a coordinate for each residential structure in the United States, facilitating the assignment of residential structures to their correct geography and improving the quality of address based census cultural and demographic statistics.

The Census Bureau collects and reports information through a wide range of programs, including the Economic Census and yearly current censuses and surveys. The 2010 Decennial Census Hand Held Computer was designed for a singular data capture operation. Based on lessons learned from the 2010 decennial experience, this decade the Census Bureau plans to expand the use of location-based technology to improve the quality of field data collection and ultimately data reporting. This includes a consistent approach, regardless of the type of census or survey. Requirements are being gathered and prototypes developed for a corporate listing device with improved technology that will meet the needs of all field data collection and data reporting activities.

United States Census Bureau Mission Statement
“To serve as the leading source of quality data about the nation’s people and economy.”

Data collection, processing, tabulation, and dissemination are central to accomplishing the United States (U.S.) Census Bureau’s mission. The U.S. Constitution requires a census, defined as being the complete count of the population, be conducted every ten years. With the U.S. population estimated at over 310 million by the end of last decade, the decennial census has grown to be the largest peacetime activity undertaken by the federal government, providing the nation with a wealth of cultural and demographic statistical data. Examples of decennial data use include determining how many representatives each state will have in the U.S. House of Representatives, creation of districts by local government entities for elections, schools and utilities, and determining the allocation of more than $400 billion in federal funds yearly.

Of equal importance are the Census Bureau’s Economic Census and current censuses and surveys. The Economic Census and a Census of Governments is conducted every five years and measures and profiles businesses and government organizations in the U.S. Current surveys yearly provide communities with information needed to plan investments and services. The American Community Survey (ACS) now offers information annually about the social and economic characteristics of communities, providing information about income, educational attainment, housing, occupation, place-of-work, and more. The nationwide implementation of the ACS in 2005 changed the method for collecting detailed demographic characteristics, replacing the decennial census “long-form questionnaire,” which was distributed to a sample of households, with a yearly sample survey. This approach to data collection has heightened the need for a high quality and continuously updated geographic framework to support data collection, processing, tabulation, and dissemination.
The Master Address File (MAF) and Topologically Integrated Geographic Encoding and Referencing (TIGER) System is the Census Bureau’s geographic framework. The MAF/TIGER database contains an address or physical description and a coordinate location (referred to as a map spot) for every residential structure known to the Census Bureau; spatial features such as streets and hydrography; and the legal and statistical geographic areas and their boundaries. The MAF/TIGER database provides the infrastructure for assigning each residential structure (and population) to its correct geographic location; this activity (known as geocoding) is essential to correctly reporting and using statistical data.

The Census Bureau relies on several data sources to update the MAF/TIGER database. Mailing addresses are captured twice yearly from the U.S. Postal Service’s Delivery Sequence File; a computerized file containing all addresses serviced by the U.S. Postal Service. Additional address information (including in some limited instances map spots) and features (primarily street centerlines) are obtained from Geographic Information System (GIS) files contributed by the Census Bureau’s tribal, state, county, and local government partners. A third data source is imagery; in-house staff extracts new streets and identify areas of new construction.

An effective, though more costly, method for updating the MAF/TIGER database is through ongoing field data collection. The Census Bureau relies on its Field Representatives (FRs) when conducting interviews for various censuses and surveys to provide address and feature data updates throughout the decade. National field data collection operations are conducted approaching, and in support of, the decennial census.

**The Use of Location-Based Services**

In 2009, more than 140,000 Census Bureau temporary FRs fanned out across the U.S. to complete an operation known as address canvassing. Address canvassing was the first major 2010 Census operation; its purpose being to update and verify the more than 145 million residential addresses contained in the MAF against “ground truth” – the actual living quarters found during the canvassing of almost the entire country (the exceptions being remote areas of two states, Alaska and Maine). Address canvassing was undertaken to ensure a complete address list for the delivery of the decennial census questionnaires to be mailed in March 2010.

For the first time in the history of Census Bureau field data collection, FRs collected accurate map spots using Hand Held Computers (HHCs) equipped with Global Positioning System (GPS) technology. Use of the specially outfitted HHC was a key innovation for the 2010 Decennial Census, and collection of positionally accurate map spots was identified as critical to reducing geocoding and mapping errors, ensuring a more accurate address list and, ultimately, a more accurate count of the population. Accurate map spots were one way of identifying residential structures that appeared on the Address List two or more times with differing addresses (referred to as “duplicates”). Duplicate addresses had become a growing concern with increasingly diversified methods and programs contributing addresses to the MAF. For example, a group quarter structure located on King Street might have entered the MAF by an FR collecting data for a current survey who identified a structure with a basic street address of “107 King Street”, while a data call for group quarters information from local governments might have
added “Shady Knoll Home” to the MAF. Associating a geographic coordinate with each address, and comparing and identifying the latitude/longitude of the two address records as the same solves the confusion of duplicates. Additional goals of location-based services use were improving the efficiency of field data collection by ensuring FRs could easily locate the correct residential structure during required follow-up operations, and reducing costs through improved address locatability and decreased travel time with the latter being dependent upon offering advanced routing and navigation functionality.

**Preparatory Work**

The Census Bureau began preparing for incorporating location-based services into census operations early in the decade.

Improving the positional accuracy of the MAF/TIGER database to ensure map spots acquired using GPS technology could be assigned successfully to their correct census block was a necessity. The MAF/TIGER Accuracy Improvement Project (MTAIP) updated TIGER between 2003 and 2008 with the assistance of a contractor who aligned all street centerlines and added new streets to a positional accuracy of 7.6 meter or better with a Circular Error of 95 percent (refer to Figure 1).

The MTAIP itself utilized location-based services and operated as a GPS technology test bed. Source data for improving street features included partner-provided GIS files and National Agricultural Imagery Program aerial imagery. The contractor conducted centerline field data collection to build street centerline files using GPS technology in geographic areas where neither of these two desired and less expensive source data was available. A sample of GPS street intersections collected by an independent contractor (110 intersections per county) was compared to the improved TIGER street intersection coordinate and differences in meters calculated to ensure the quality of realigned TIGER streets prior to improved file acceptance.

Concurrent with the MTAIP, the Census Bureau extensively evaluated the use of GPS technology, exploring the potential benefits of more efficient methods for data collection and navigation. Projects

---

1 TIGER began life from a patchwork quilt of data sources. Prior to the MTAIP, the Census Bureau described the positional accuracy of TIGER as at best meeting the established National Map Accuracy standards (approximately +/- 167 feet) where 1:100,000-scale maps from the U.S. Geological Survey were the original source. The Census Bureau could not specific the accuracy of features derived from other initial data sources or subsequently contributed by its FRs through field data collection. The positional accuracy was described as highly variable.
2 The Census Bureau’s goal was to improve the accuracy of TIGER to support the placement of map spots having a positional accuracy of 3.0 meters or higher in the correct census block 99.6% of the time.
3 The narrowest road widths encountered in urban areas where structures are located curbside as well as the error associated with collecting both street centerlines and map spots were taken into consideration in determining the TIGER positional accuracy requirement.
4 A Circular Error of 95 percent (CE 95) is the confidence that 95% of the time any street point within TIGER will fall within, and meet, the 7.6 meter positional accuracy requirement.
5 The contractor’s field technicians used a vehicle equipped with a GPS, inertial navigation unit (INU), and distance measurement instrument (DMI) to track vehicle location. Multiple imaging systems attached to the roof of the vehicle imaged the roadways and road signage, and personal computers recorded the information that was then used in final, post-processing steps to produce a centerline street file.
6 The MTAIP unit of work was the county.
focused on improving the efficiency of data collection rather than providing navigation functionality to FRs as the MAF/TIGER database lacked (and continues to lack) specific road attributions such as turn indications and one way road flags necessary to provide an advanced version of this location-based service. Early testing highlighted the importance of developing comprehensive procedures for GPS field operations. Standard data collection requirements for GPS activities were refined, procedures sharpened, and training developed to introduce FRs to GPS technology. Field Representatives were trained in GPS signal considerations that included allowing the GPS receiver to get a satellite fix before beginning work; remaining stationary while collecting map spots to collect data that is not skewed; positioning the body to avoid satellite signal blockage; collecting a GPS coordinate in a place of minimal overhead and surrounding obstructions to achieve maximum reception; and avoiding large metal objects (trailers, storage sheds, silos) and bare rock faces that can reflect the GPS signal in a phenomenon called “multi-pathing.”

Simultaneous with the above-described activities, the HHC was being developed, prototyped, and tested under a Census Bureau contract. The HHC ultimately would display a GPS-powered “You Are Here Indicator” (YAHI) to provide FRs with limited location and routing functionality. The YAHI traced the location and movement of the FR along TIGER streets (the MAF/TIGER database was the source data displayed on the HHC), enabling the FR to see their current location and by selecting an address on the address list to see where they needed to be and plan their route accordingly. The decision was made that collecting map spots using GPS technology would remain a behind-the-scenes function with the FR

---

7 Temporary FRs employed for short term operations oftentimes are retirees; a large percentage of FRs in 2009 were not well versed with location-based services and the use of GPS technology.
unaware of the collection. The GPS collection was initiated unbeknownst to the FR when the FR began a manual map spot collection. The Census Bureau also worked with the National Geodetic Survey (NGS) regarding the feasibility of planning for and completing a differential correction for GPS captured map spots. The Census Bureau/NGS partnership led to the development of post-processing software. The HHCs had Wide Area Augmentation System (WAAS) capability providing a three-meter or less radial accuracy (95 percent) in an unobstructed environment. The three-meter accuracy coverage was extended by post processing map spots collected using GPS technology with Continuously Operating Reference Station (CORS) after the fact if the WAAS signal was unavailable during collection.

One feature of the HHC allowed FRs to collect street centerlines using GPS technology. Early testing identified the difficulty of successfully integrating newly collected GPS streets with existing TIGER streets due to gaps or overlaps at start/end points. Procedures were established to ensure FRs began the GPS coordinate collection process before driving onto the street to be collected, driving the entire length of the street before stopping at a safe location to end coordinate collection, and once stopped clipping the newly-added street to its proper start and end points. Field Representatives also were instructed to drive as close to the center of a street as possible so that the GPS feature that was captured most closely represented the “centerline” of all lanes of traffic.

Address Canvassing a Success with Significant Address and Street Updates

During address canvassing the FRs assigned GPS coordinates to more than 98 percent of residential structures in the MAF. There was a significant reduction in duplicate and incorrect addresses: 144,890,809 addresses were sent to address canvassing while 134,171,391 addresses went forward to enumeration following the operation. Field Representatives took 2,756,444 actions on streets: 623,544 roads were added with 97% of the additions occurring using GPS technology, 758,166 erroneous streets were deleted, and 1,189,403 road names were updated. A result of employing the HHC was that, for the first time, new streets and corrections to existing streets in TIGER were completed in time for use in the next field operation, as there was no longer the need to transfer inaccurate updates received via a paper based format.

Lessons Learned

Address canvassing, as the first nationwide Census Bureau field data collection operation employing GPS technology, offered myriad lessons for improving the future use and efficiencies of location-based services.

HHC performance issues

The 2010 Decennial Census HHC was designed for a single data capture operation. The HHC enabled the use of GPS technology; however, performance issues affected the ability of the FRs to successfully and continuously utilize location-based services. Hardware design and capability is integral. Limiting factors reported by FRs included HCC shut-down in extreme temperature conditions (severe cold in Alaska and the northern tier of states and severe heat in southern and tropical states, such as Hawaii). Field Representatives had difficulty viewing screen displays in the sunlight and complained that too
small screen size required time-consuming and extensive zooming in and out of geographic areas. Individual FR workload scheduling was affected by HHC power requirements (battery life), although early tests did lead to an eventual battery recharge time requirement of less than eight hours; insufficient storage capacity to handle data in large census blocks containing 1,000 or more residences; and data transmittal time. Data transmittal to and from the HHC was not real time; rather, the requirement was for a 500K byte download in less than 24 hours with less than 48 hours for a complete software download to update the application when needed. Field Representatives transmitted data upon completing each assignment area, or if the transmittal failed, attempted a nightly transmittal of collected data to headquarters.

**GPS technology limitations**

GPS technology is not a panacea and the Census Bureau experienced its known limitations in address canvassing. Poor GPS signal reception, loss of GPS signal integrity, and limited positional accuracy were encountered to varying degrees throughout the country in differing environments. In the inner city streets of urban areas lined with skyscrapers the visibility of GPS satellites was limited and signals were obstructed for extended periods. In other areas, dense vegetative cover caused interference. Signal distortion resulted in both environments from large buildings, large metal objects, and bare rock faces.

Characteristics of the data being captured also were limiting factors. Residential structure overhanging roofs, awnings, porches, and landscaping (tall trees close to houses) interfered with GPS reception at the preferred map spot collection point: the structure’s front door. Field Representatives were instructed to stand at the front door step for 20 seconds while manually map spotting the point to achieve maximum GPS signal reception and accuracy but FRs did not understand the need for this given the hidden nature of the GPS collection and so would ignore the instruction. Another problem associated with collecting the residential structure point at the front door is long term in nature and related to the desired future move from field data collection to a reliance on partner-supplied or commercially produced GIS files as primary sources of continuous MAF/TIGER updates. The Census Bureau’s collection instructions differ from common practices for locating address points, which are either the center of the structure rooftop, the center of the structure’s “parcel,” or at the street. This difference in capture point may affect the Census Bureau’s ability to ingest and reconcile data from partner-supplied or commercially produced GIS files as primary sources of continuous MAF/TIGER updates.

It should be noted that special considerations accounted for multi-units within a single structure given the inability to capture positionally accurate GPS locations indoors. One map spot was collected for the structure at the main entrance and associated with each of the apartments within the structure. Semi-attached housing units and row houses or townhomes posed difficulties for accurate GPS map spot collection when front doors were located less than three meters apart.

**Manual vs. GPS map spots**

Possibly the most valuable lesson learned was related to the hidden aspect of GPS map spot collection. As previously mentioned, FRs were unaware of GPS map spot collection as it was a background activity invoked by manually placing a map spot when standing at the front door of a residential structure. The
result of an FR failing to follow procedures and manually inserting all map spots for visible residential structures from a single location is a group of GPS coordinates representing the FRs actual location while recording the many manual structure points, or a “curbstoning cluster.” (Refer to Figure 2) The Census Bureau preferred map spot was GPS over manual IF the GPS map spot fell within the correct collection unit (the census block) after being post-processed. Curbstoning clusters required introduction of an in-house processing operation that compared distances between a structure’s GPS and manual map spots followed by an interactive effort to identify the most preferred map spot.

![Figure 2: Curbstoning cluster example. Red pins are manual structure points recorded by the FR; the yellow pin is the GPS location recorded for all displayed manually collected structure points.](image)

**Next steps: moving forward with location-based services**

This decade the Census Bureau plans to expand the use of location-based services to improve the quality of field data collection and ultimately data reporting, incorporating lessons learned from the 2010 decennial address canvassing experience. The Census Bureau is developing a consistent approach to field data collection, regardless of the type of census or survey. The HHC was designed to meet the needs of a single operation to support the 2010 Decennial Census – address canvassing. A second device, the Automated Listing and Mapping Instrument (ALMI), is a personal computer that has been retrofitted with GPS and is being used in the short term for limited current surveys field data collection. In the meantime, the Census Bureau is identifying requirements and developing prototypes for a corporate listing device with improved technology that will meet the needs of all field data collection and data reporting activities.

---

8 “Curbstoning” is census terminology for FRs sitting on the curbstone and falsifying information about a household rather than obtaining it. The term dates to the era when decennial census data were collected through the traditional door-to-door process.
In 2010 an interdivisional “Improving Operational Efficiency” development team was formed to identify the high level requirements that will define the functions of a Census Bureau corporate Listing and Mapping Application (LiMA). The LiMA will be platform independent and executable on multiple devices. Concurrent with the requirements development phase, work has begun on resolving hardware and operating system considerations. Potential real-time data transmission is being investigated. GPS Field Data Collection Guidelines and Technical Guidelines are being developed to take advantage of improving GPS technology and to obtain the hardware required to do so.

**The importance of consensus requirements**

The past decade identified the need for developing consensus requirements early in a decade and applying the same requirements to all field data collection operations. The LiMA requirements team applied a “from scratch” technique, conducting their work without reliance on previously written requirements to ensure a focus on root needs as opposed to modifying existing processes. A use case diagram assisted in identifying functions. Team members submitted their specific operational requirements for consideration and a listing of agreed upon high level requirements along with a matrix of operation applicability to support a modular LiMA development approach were produced. The requirements capture “what to achieve” (i.e., collect a GPS map spot for a residential structure) rather than the “how to achieve” (i.e., actual steps for successfully collecting a GPS map spot for a residential structure), though some elemental decisions were agreed to (i.e., the FR must be aware and actually invoke collection of a GPS map spot for a residential structure) based on lessons learned.

**Identifying requirements expanding the use of location-based services**

Additional functionality unavailable in the HHC and ALMI were identified as desirable following the realization of previously available, and now considered base, functionality. Examples of requirements for additional functionality are:

- **The system (i.e., the LiMA) will support Global Positioning System/Global Navigation Satellite (GPS/GNSS) functionality.** This requirement acknowledges the importance of technology advancements and that the future desired carrier phase (Russian – GLONASS, USA L1 and L1 carrier – GPS, or European – Galileo to name a few) may be different from those used in the past.

- **The system will be able to suggest to the user and display on a map a prioritization or ordering of their assignment areas for the day for all surveys assigned to the user.** This requirement identifies the importance of FRs being assigned multiple surveys and collecting data in the same manner for the entirety of their daily workload rather than being trained in field data collection for a single survey having unique requirements. It also identifies the importance of improved workload prioritization and routing to increase efficiencies.

- **The system will use Geography Division approved imagery in the map display area, sufficient for the user to do his or her work, when there is connectivity AND the system will support displaying additional imagery (from on-line sources such as Google or Bing), when there is connectivity.** This requirement documents that imagery as a backdrop will improve the use of location-based
services as well as the desire to use available, on-line resources such as Google or Bing rather than to duplicate applications and sources.

- The user can view previously collected user path imagery (i.e. bread crumbing) that is associated with a specific residential structure. This requirement offers the user in a follow-on operation additional location and navigational aid to successfully arrive at the desired structure.

- The user can clearly distinguish a map spot’s capture methodology (manual or using GPS/GNSS technology) from the map spot symbology. This requirement underscores the need for collection methodology transparency, providing the FR instant clues as to the origin of information.

- The system will inform an FR accomplishing quality control tasks if map spot manual and GPS/GNSS coordinate collection were not within a to-be-specified distance of one another. This requirement moves quality control from after-the-fact processing forward to the data collection processes. It supplies the FR with additional information to ensure early in the process that GPS technology limitations or failing to follow procedures has not resulted in poor data quality.

Development of high-level requirements is complete and a 2020 Automated Field Listing team is moving forward with developing lower-level requirements. Instruments are being prototyped and an early version of the LiMA is being created for use in the first 2020 decennial tests scheduled for late 2013 and early 2014.

Conclusion

The 2010 Census address canvassing operation illustrates the importance of location-based services in updating the Census Bureau’s address list and contributing to the essential activity of assigning each residential structure (and population) to its correct geographic location with the resultant improved reporting and use of statistical data. The Census Bureau is committed to location-based services as a method of continued address and street update; understanding the need for a single set of requirements that can be applied to all field data collection activities. This decade’s work has begun with an understanding of the expanded role of location-based services.