Conversation with Timothy F. Trainor[1]

Nancy K. Torrieri
Interview Editor, SJIOS
E-mail: nktorrieri@gmail.com

Geography provides the framework for statistics, survey design, sample selection, data collection, data tabulation, and data dissemination at statistical offices worldwide. Timothy Trainor, Chief Geospatial Scientist for the U.S. Census Bureau, provides a look at how geographers at that agency have helped shape the Census Bureau's mission to serve as the leading source of quality data about the people, places, and economy of the U.S. by providing cutting-edge innovations, expertise, a commitment to customers, and a global outlook.

---

[1]The views and opinions expressed in the conversation are those of the interviewee and do not necessarily reflect the policy or position of the U.S. Census Bureau, the *Statistical Journal of the International Association for Official Statistics,* nor IOS Press.

Tim forecasts that in the future, GIS (initially, an abbreviation for Geographic Information Systems, and, more recently, Geographic Information Science), will focus on analyzing geospatial and statistical data to understand their core meaning and get the most value from such rich data sources. Data quality, while oftentimes discussed, will be expected and demanded of data providers. Higher resolution and greater precision of even more information require new methods of applying and assessing data quality, which provide new opportunities for those involved in geospatial data management. The foundation provided in this conversation helps set the stage for advances in the future.

Interviewer: Please tell us about yourself, your academic background, and your career. What was your path to geography and the U.S. Census Bureau?

My undergraduate degree is in history from Rutgers University, and when I went to obtain a teaching certificate, I started taking geography classes at Glassboro State College (now Rowan University). Folks at the school where I was taking classes were so interested and so passionate about what they were doing, and I think their enthusiasm transferred to me. After teaching for two years, I applied for a Rotary Foundation fellowship and decided to do a post-graduate degree in cartography at Glasgow University in Scotland. I came back, taught another year, and needed to make a decision on whether I would pursue a career in cartography. So I accepted a position with a small company, Vernon Graphics in Elmsford, NY, where I worked on an interesting orthophoto mapping project

for the whole state of Vermont at very large scales. At that time, the best career opportunities were in government. I began working for the Defense Mapping Agency in May 1980 but then was contacted by the Census Bureau as they were in the middle of the 1980 Census and needed help. The second day on the job I was sent to Jeffersonville, Indiana, where much of the Census Bureau's large-scale map production activities took place, to help manage a massive mapping operation. At that time we had over 1,000 people drafting paper maps. Amazing – but mapmaking was all done by hand in those days because the world of digital mapping had not yet taken off.

I worked in the Census Bureau's Geography Division (GEO) Cartographic Methods Branch initially, then advanced to management positions where I led a highly skilled team of cartographers in producing more maps than most other organizations. In 2003, I led the geospatial standards side of the GEO and managed an area there for about a year before being promoted to the assistant division chief over geographic areas and mapping. At the end of 2008, I became Chief of the GEO, and I was in charge of the GEO during the 2010 Census and during the years that followed, until May 2016. During that time, I was responsible for staff who performed all of the tasks needed to create and manage a national address list with locations of housing units, all of the legal, statistical, and administrative boundaries for the Nation, and a digital representation of all of the roads in the country with their associated names.

The Census Bureau put in place methods and procedures to continually update the address list through partnerships with governments that would have not only significant benefits to the 2020 Census, but that would also support ongoing current surveys like the American Community Survey (ACS). That step reflected a recognition of the importance of geospatial information to the goals of the Bureau and especially to its leadership role among statistical organizations in the U.S. and around the world. By 2015, the Census Bureau accepted the need for a special position that carried with it a set of responsibilities and duties that would reflect the agency's leadership position in geospatial data activities at the Census Bureau, while representing the agency internationally to share knowledge and best practices in geospatial science, standards, policies, and activities. So the Bureau created a new position, Chief Geospatial Scientist, in 2016. I was appointed as the first Chief Geospatial Scientist, and that is the position I currently hold. I think the Census Bureau's creation of that position says a lot about the Bureau and its recognition of the importance of geospatial information as it relates to the goals of a world-renowned statistical organization. Overall, it is a position that is more outward looking than inward looking, and it dovetails with the Census Bureau's need to meet and interact with other external organizations such as professional groups that support research in geography, statistics, demography, and social science.

I am also involved in planning associated with the 2020 Census. In that position I'm spending my time in three primary areas. The first focuses on leadership guidance for the 2020 Census. Having been through four prior censuses, I have learned quite a bit about what's required to conduct a successful national census of population and housing. Having had the opportunity to be in a senior leadership position for the 2010 Census, I am sharing my experience with the current leadership team on planning and implementation for the 2020 Census. The second area of responsibility centers on the Federal Geographic Data Committee (FGDC), which coordinates all geospatial activities within the federal government. I represent the Department of Commerce (DOC) for the work associated with that Committee. Within the DOC, contributions to that Committee occur chiefly by the National Oceanic and Atmospheric Administration (NOAA) and the Census Bureau. I am very active on both the FGDC Executive and Steering Committees, and, in addition, provide expertise to the National Geospatial Advisory Committee. The FGDC serves as the focal point for geographic coordination activities within the federal government as well as geospatial efforts that are national in scope.

Aside from work to support the 2020 Census and the FGDC, I spend significant time representing the U.S. on geospatial activities coordinated by the United Nations (U.N.). As the head of the U.S. Delegation to the U.N. Committee of Experts on Geospatial Information (UN-GGIM) I was elected by the Member States as Co-Chair of this effort to play a leading role in setting the agenda for the development of global geospatial information and to promote its use to address key global challenges. In 2011, the U.N. created a committee of experts on geospatial management under the U.N. Economic and Social Council. In addition to the Member States, the UN-GGIM has active participation by three networks. One network is made up of representatives of international professional societies like the International Society for Photogrammetry and Remote Sensing, the International Cartographic Association, the Global Spatial Data Infrastructure Association, the International Hydrographic Organization, the Interna-

tional Geographical Union, and others. The academic network was created as an avenue for researchers and interested academics to get involved in the work of the UN-GGIM. The private sector network offers an opportunity for companies and trade groups to engage in geospatial information projects that support UN activities.

The goal of this Committee is to try to convince high-level leaders in different countries of the importance of geospatial information, why it is critical to a nation's economy and well-being, and why it should be funded. Sometimes it is a struggle for many countries to fund the creation and use of geospatial data in view of the many other urgent priorities they face. We promote the identification of specific kinds of goals and data that would be most useful. For example, the Committee is currently supporting a recommendation to develop a global geodetic reference frame to document accurate positions on the earth based on common measurement methodologies and techniques. This is helpful to address issues associated with such diverse applications as precise positioning, sea level changes affecting land administration, and the integration of statistical and geospatial information. Another recommendation of the Committee concerns the need for a common unit of geography to report the collection of data to analyze and study sustainable development and related goals. For now, small geographic areas and geometric grids are both accepted as data reporting units. As part of the 2030 Agenda for Sustainable Development, the Member States identified 17 goals, 169 targets for achieving them, and 232 indicators of progress. Of the 232 indicators of progress, 84 are judged as significantly difficult to achieve in the sense that a methodology for collecting and/or using these data is spotty or nonexistent. In some cases, geospatial information may contribute to elevating their condition for use.

Over the last few years, the Committee has worked with the chief statisticians of the world through the U.N. Statistical Commission to stress the importance of the geospatial information available to them for their national statistical organizations and their programs. This information adds value to the geospatial knowledge that is critical to sustainable development. It's clear that geography and those who understand and use geographic knowledge have a strong leadership role. Geospatial data combined with statistics help support analysis of the planet and the people on it.

Interviewer: Tell us about activities leading to the development and evolution of TIGER, starting when you arrived at the Census Bureau in 1980.

For the 1980 Census, everything related to geography – and many other activities at the Census Bureau – were done by hand. The maps for data collection and data dissemination showed the geographic areas and features that were needed for conducting a census as well as tabulating the data that was collected. For data collection, every area of the country was assigned a type of enumeration area that either involved visiting a household or mailing a questionnaire. We had limited use of address information at that time given that most areas within the country did not have city-style addresses (house number and street name). Where we did use addresses, we had purchased address lists from vendors but we were not maintaining our own address file because of the high number of changes that occur during a 10-year time period. For each decennial census until the 2000 Census, we used files we purchased from vendors then put them aside and replaced them with the next set of files 10 years later.

We had some basic geographic information that included, for example, the boundaries and map features like roads and railroads that we needed to help us delineate the census geographic areas and put them onto the maps. We had a map base source either from state departments of transportation or from USGS topographic quadrangles for urban areas. We added census geography to these map bases and made new maps. We also had organized a simple, yet fairly sophisticated, system of geographic codes reflecting the earliest efforts of geographic standardization. These geographic codes (and by definition, the coding system) became part of the Federal Information and Processing Standard (FIPS) codes that we used to maintain existing geography, delineate geography for new geographic areas, and include that information not only on maps but also as part of computer programs that tabulated census results to the correct geography. The FIPS codes made it possible to relate a geographic area to locations on the ground. The National Institutes of Standards and Technology (NIST) maintained the FIPS codes, and the geography that the Census Bureau maintained supported those codes. So, for example, the nation was divided into either enumeration districts (EDs) for most non-urban areas or block numbering areas (BNAs) for the urban portions of the country, and these two geographic areas were the smallest units of geography we maintained on a national basis before and during the 1980 Census. We

delineated EDs based on criteria we had, but a lot of it was simply educated geographic guess work in first determining the geographic extent of the urban areas for BNAs. What areas were left in the non-urban areas became the EDs. That meant that for rural areas, the published data could never be reported for an area smaller than an ED, and most published data was at higher levels of geography such as a county. In the urban areas, census blocks formed the foundation of each BNA and block statistics were made available for basic counts for the urban areas while more detailed population and housing characteristics were published at higher geographic levels to conform with confidentiality requirements.

Legal boundaries for towns or cities can periodically change. Let's say a town annexed new territory, and it extended out into a farmer's field, and the town started adding streets. We had to put all of that new information on the base map. In doing that, you're changing the codes for some parts of the newly annexed area. For example, in moving boundaries, you might be calling areas cities that were previously the rural area of a town or county. If you added new roads and if for some reason you did not make changes to the boundary network, you had a possible geographic error which could lead to misallocation of data collected from a census or survey. Checking and editing had to be done by hand.

The impact of these issues became clearer as a result of the 1980 Census. Inconsistencies across geographies and between different kinds of geographic information became recognized as a difficult and serious problem. The thought at that time to minimize this challenge was, if we were able to create a file or source from which all information flowed, even if you had an error, the error would be consistent through different geographies, you could correct it once, and the correction would be implemented for other geographies at the same time. If you're doing that for a small community, that's a big job in itself, but if you're doing it for the entire U.S., that's a huge job! And because there are variations in geography from state to state, county to county, and city to city, different rules apply to different situations. That led us to recognize that a database was needed, and we should try to make it happen. One of the first jobs I had at the Census Bureau was to write a white paper on the need for information in a geographic database. I then began developing specifications for digitizing the roads and hydrography from the USGS 1:2400 scale map sheets. The U.S. Geological Survey (USGS) learned of the Census Bureau's plans and told us about their own new map series, which al-

lowed them to scan new data. We didn't need the same information that the USGS did. But the scale of the map series they were developing, 1:100,000, appealed to us because it was manageable (56,000 map sheets vs. 1,800 map sheets). So we adopted that standard for our own needs.

Robert (Bob) Marx was the genius behind the notion and development of the Topologically Integrated Geographic Encoding and Referencing (TIGER) system, an automated geographic database covering the United States, Puerto Rico, and the Island Areas: American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, and the United States Virgin Islands.[2] Marx devised the notion of TIGER at a time when the databases of the day were principally relational databases comprised of rows and columns. TIGER made it possible to create the first nationwide digital map of geographic boundaries, roads, hydrography, and other features. TIGER/Line products, public derivatives of the TIGER System, helped power the GIS industry by providing a common geospatial framework that could link census and other data with GIS through topology, which assured the maintenance of correct geographic relationships.

Interviewer: Discuss some of the challenges in developing TIGER with the USGS.

The Census Bureau and the USGS forged a cooperative agreement to work together to develop TIGER. The USGS had data on a number of features we needed, and some we did not, but both agencies shared responsibilities for developing TIGER. The DOC and the Department of the Interior, where the USGS is housed, had a successful collaboration, not just of ideas and concepts, but of real work. The USGS viewed the development of TIGER as a priority, and that view made it possible for both agencies to succeed. Leaders in the GIS industry who were active in the 1980s would say this relationship served as a catalyst for the development of GIS for the world. The most difficult and expensive process was converting map data to digital form and this project made that possible for the U.S., which also served as a successful model for other parts of the world.

---

[2]Reference information on the geographic terms and concepts used by the U.S. Census Bureau is available at https://www.census. gov/geo/reference/.

We took the feature information pertaining to roads from the USGS and prepared nearly 56,000 1: 24,000 scale map sheets. We decided that the USGS feature coding scheme could not be used directly. The USGS coding structure was additive in nature. Each feature could have up to seven different multi-numeric codes describing its characteristics. This would have meant a lot of tagging, and a lot of interpretation that was crucial to the USGS but for which we neither had the same need nor the same feature interpretation skill level. We developed our own feature coding scheme, which was a simpler coding structure with one and only one code for each feature in our file. This was one of my contributions to the process. We used an alphanumeric coding structure that had fewer digits and that seemed to work very well. A variant of that structure is still used today. We recoded all the roads in the US while the USGS coded the hydrography. We manually digitized railroads and the miscellaneous transportation features (pipelines, powerlines, etc.) that were not as important to us as other features provided by the USGS.

Interviewer: Discuss the operational challenges that accompanied the development of TIGER. Were there any development benchmarks that you recall as being especially significant?

Bob Marx put aside the organizational structure of the GEO and created new teams to provide for a more functional approach to getting the work done. Remember – we started with nothing in terms of a design – nothing we could copy, and we did not know if we could actually develop and make this new idea work. We didn't know what the end would look like. We had to figure out how to get started, then move forward. There were a few efforts in the origin of GIS on the environmental side of things, but nothing in a general geographic environment that had the kind of data we needed. This was a time in the origins of personal computers when many new ideas were being tested. But there was no geospatial data structure that could handle the diversity and scale of data that would become TIGER. Anyone heading down the road of automation came up against the immense cost of digitizing data, the most time-consuming thing to do, and it was very complicated to translate the meaning of map data into digits. We had to develop the notion of topology to assure that the geographic relationships between features and boundaries (and eventually addresses) were accurate to avoid the troubles we experienced in the 1980 Census. But there was one special notable benchmark – something of a victory for us as I recall. In 1986 we made available for evaluation a TIGER extract for Boone County, Missouri that included the city of Columbia. We made it available on 9-track computer tapes (long before CD-ROMs). The reaction was overwhelming. People were so impressed and so happy to see something. Soon we were inundated with requests from GIS professionals for TIGER extracts for other counties. We set up a process to meet the demand, charging only for the cost of putting the files onto the media. It was a major operation, but totally worth the time and effort, because we were aware that those requesting the files knew their value and were deeply appreciative. For some folks who were beginning to write software, this was the beginning of a new GIS business for them – this made it possible for some GIS companies to get started. Today, downloads of TIGER shape files are free while the data can be accessed and viewed via TIGERweb (https://tigerweb.geo.census.gov/tigerwebmain/TIGER web_main.html). Considering that TIGER is still maintained and updated on a regular basis and serves a vast audience of users, that's quite a contribution and a bargain.

Staff in the GEO worked very hard to meet a difficult and demanding schedule for developing TIGER. Initially, we used the UNIVAC, which provided a sequential processing environment. At some point we shifted to a Digital Equipment Corporation (DEC) environment for virtual processing to enable greater throughput and then used Silicon Graphics workstations to process the extensive amounts of data. In those early days, if we needed to draw a line on the screen, we had to write code to make the line. We had to write code to create and scale fonts where we had to digitize letters and numbers in order to have 2–3 different fonts. These did not exist when we began. We were making maps for use by field enumerators but had no idea of whether it was possible for enumerators to use large digital maps in the field. We had to write software to develop and plot the maps needed for the 1990 Census where the plot files were unique to the output device.

In the original TIGER data, we wanted to preserve much of the work we had done for the metropolitan areas of the country. For many urban areas, we had created something called the Geographic Base File-Dual Independent Map Encoding (GBF-DIME) file. This was a file structure of nodes (points) that allowed for geocoding of addresses. To exclude this source meant that much more work would be required for ar-

eas where most of the population lived. We decided to carry forward these areas within the TIGER system so that we could maintain the geographic relationships that already existed. The downside is that when these urban areas were mapped, the point data was limited only to street intersections. Many of the early maps failed to show the curvature of a road, which sometimes affected their usability in the field. But there were opportunities for changes, too. For example, we knew that there was interest in the Census Bureau and elsewhere of having statistical data at geographic levels smaller than the ED. This was driven principally by the redistricting needs. States and locales were drawing boundaries and being challenged about the boundaries.

We had made a decision as part of TIGER development to create block-level geography for the entire country. For the 1980 census, we had some block-level data for urban areas and a couple of states that had contracted to have block-level data, but for most of the land area, the ED was the smallest unit and those were pretty big. We also had census tracts for urban areas, but not for rural areas, so we created a collection block numbering area to serve the needs of the nonurban part of the census tract program. Once we created blocks, we created block groups that followed rules for nesting geography, so everything nested in a consistent state/county/tract/block configuration. We were able to offer statistical data and the block-level counts for every state to do their redistricting for the first time after the 1990 Census.

Interviewer: What challenges lie ahead for geographic operations for the 2020 Census? Are there plans to do things differently from previous censuses?

In conducting a national census for somewhere as diverse and large as the U.S., it is an amazing event for those that have lived through it, and they never forget it. The upcoming census will be the fifth census I have worked on.

We do have some planned changes for many operations for the next census. The current operational plan, available online at https://www.census.gov/2020 census, gives a detailed description of these planned changes. They include for the first time an internet self-response option. The overall goal of reducing costs affects every one of the 35 census operations, and many of those operations include a geographic component.

As far as changes that directly involve geographic operations, the most critical is a reengineered address

canvassing operation. Address canvassing for previous censuses involved traveling down every street to verify and update addresses in the Master Address File (MAF) in preparation for data collection. It is one of the two most expensive decennial census operations. For the 2010 Census, in order to assure we had a good address list, we hired 140,000 people to drive or walk every street in the country and check their assignment area – defined by a list of addresses for each census block – against what they saw on the ground. This time around we are conducting an address canvassing operation through "in-office" procedures, first, by updating the address list based on new information from the USPS and data from tribal, state, and local governments, and information from third parties (such as commercial vendors). Clerks are reviewing satellite imagery to determine where changes in addresses are occurring. That makes it possible to target areas of known development for further research on adds, deletes, moves, or other changes in the address list. While 100 percent of the addresses will be reviewed in the office, address canvassing in the field is planned for those areas where change is frequent, difficult, or requires further research, which accounts for only about 25 percent of the total number of addresses for the 2020 Census. This design change will reduce costs dramatically compared to the last census. This operation also positively impacts another planned 2020 Census improvement that will simplify the field management structure to reflect newer capabilities like case management, Global Positioning System (GPS) technology, and GIS tools.

Interviewer: Coverage is an important issue related to the quality of the address information. How is the Census Bureau working to improve coverage and avoid having duplicate addresses?

The address canvassing procedures are designed to provide updates, but a practice long used at the Census Bureau relates to the fact that we never throw anything away. If we are made aware of a new address, for example, as a result of a field operation, we put the address into the MAF. If in a subsequent operation we match it against the USPS Delivery Sequence File (a primary source of address information delivered to the Census Bureau on a biannual basis) and get a mismatch, we investigate, but we don't throw away the address that we think is invalid. Keeping address variants minimizes the amount of research required to determine the cor-

rect address. That's not the best solution, since we need to minimize the need to keep all data records over time. A better solution would be to use criteria-based decision making to learn which address is right. The best source of address information is from the local governments where addresses are created. During this decade, we have engaged with local and state governments to acquire their data to reduce the level of effort required just before the census. We are working closely with governments and organizations to share best practices in effective address data management. This has been one of the best geospatial data management improvements between the last census and the 2020 Census.

Interviewer: Tell us about the Local Update of Census Addresses (LUCA) program. Why was that program created?

When we went through the 1990 Census, we used an address control file (ACF) as a master address list. That was principally a commercial product. We bought address files from a commercial vendor. But there were problems with some of the data, serious problems that affected redistricting (undercoverage for some communities), and the states wanted improvements in the master address list (known today as the Master Address File) that the Census Bureau used. That need spurred the passage in 1994 of the Census Address List Improvement Act (Public Law 103–430), which changed the Census Bureau's decennial census address list development procedures. The Act established a working relationship with the USPS and expanded the methods the Census Bureau could use to exchange address information with tribal, state, and local governments in preparation for the census in order to support its overall residential address list development and improvement efforts. The Census Bureau devised the LUCA program to implement address-sharing efforts with governments, provided they signed a confidentiality agreement to protect the data.

The 1994 Act enabled the working arrangement with the U.S. Postal Service that continues today. We receive biannual updates of the address list used by the USPS. Those updates are critical to improving the quality of address information for the Census, and for statistical surveys such as the ACS, which replaced the decennial census long form survey after Census 2000. The MAF in 2000 served as a benchmark for future use and was improved with each delivery sequence file (DSF) delivery from the USPS leading up to the 2010

Census. The LUCA program is the one opportunity for local governments to see addresses managed by the Census Bureau that serve as the frame for the decennial census. Local, state, and tribal governments that choose to participate in this program have an opportunity to add or change addresses and to appeal differences prior to the creation of the final address frame leading up to the census. This commitment continues again for the 2020 Census.

Interviewer: How can the Census Bureau meet the expectations of data users and stakeholders in the future?

For starters, data are the ingredients for planning and decision making. No matter how successful innovations like TIGER are, you still need complete and accurate data and that is also true for those of us who work as geographers. It's incumbent on agencies like the Census Bureau to maintain, improve, and expand what they do as far as data goes so that its value is realized through its use, that the data are clearly understood, and that a continuous improvement process becomes a priority. There is a lot of data being created, much of it geospatial, and people need to be comfortable understanding how to use it. That suggests the Census Bureau needs to continue its role in data management and support continuing education in the use of its data across the agency and with the user community. International collaboration and research are also relevant so that best practices, techniques, and methods can be shared.

Geography should be the foundation for this education and serve as the catalyst for cross-disciplinary perspectives needed to make this education successful. The GEO should help make that possible, and the Census Bureau should continue to make that a priority.

Interviewer: Tim-thank you for your time. It's clear that that the Census Bureau has a leadership role among geographic and statistical organizations around the world. Its programs and products benefit geographers, statisticians, and other data users in government, business, academia, and the private sector.