



INNOVATIONS FOR SUCCESSFUL SOCIETIES

FROM SAVING THE CENSUS TO GOOGLE MAPS: THE U.S. CENSUS BUREAU'S TIGER SYSTEM, 1980–2010

SYNOPSIS

After the 1980 US national census, 53 state and local governments sued to correct alleged errors in the count, and the US Census Bureau found itself at a crossroads. For years, the bureau had integrated information from paper-based sources to create maps for its census takers, and the procedure was slow and unreliable. An overhaul of the cumbersome system would be a complex and difficult task, and there was a deadline: the next national census would take place in 1990. Robert Marx, head of the bureau's geography division, decided to take advantage of new advances in computing technology to improve performance. As part of that initiative—known as Topologically Integrated Geographic Encoding and Referencing, or TIGER—the bureau built interagency cooperation to create a master map, developed a software platform, digitized information, and automated data management. Its efforts generated a nationwide geospatial dataset that fed an emerging geographic information industry and supported the creation of online services such as MapQuest, OpenStreetMap, and Google Maps. This case provides insights on overcoming common obstacles that arise in the collection, digitization, and publication of information in accessible formats, which are challenges that affect many open-data reforms.

Pallavi Nuka drafted this case study based on interviews conducted in August, September, and October 2017. Case published February 2018.

INTRODUCTION

“Today people don't think twice about using Google Maps to look up an address or get directions. But the Census Bureau's TIGER data was what enabled all these mapping companies like OpenStreetMap, MapQuest, and others to get started. It sparked the whole revolution in geographic information systems across the country,” said Jon Sperling, a senior policy analyst in geographic information and analysis at the US Department of Housing and Urban Development and former census geographer. A \$300-million initiative, TIGER (Topologically Integrated

Geographic Encoding and Referencing) created a trove of geospatial data that made it possible to map traffic patterns, pinpoint crime, deliver emergency services faster, and obtain directions—in addition to making census collection, processing, and tabulation faster and less error prone.

At the time of the 1980 census, however, no TIGER was in sight. Unlike some of its European counterparts, the United States had no regularly updated national residency register or address database. Instead, every 10 years, the US Census Bureau counted every person where every person

lived. In 1980, the agency mailed out 88 million questionnaires to the addresses on file with the United States Postal Service, but 1 out of 4 households failed to respond.¹ The need for an accurate count meant going door-to-door to find people who had failed to return a form or who picked up their mail at a different location than where they resided.

To produce the thousands of different maps census takers needed to canvass houses and tag survey forms with geographic locations, the bureau worked from three, poorly linked sources: paper maps, address reference files on magnetic tape, and geographic reference files stored on punch cards. When there was a map update, like

the addition of a new road, staff had to manually copy the change to the other sources. Any inconsistencies had real impact because political boundaries and funding allocations depended on accurate local population counts (See Textbox 1).

Following the 1980 decennial census, state and local governments filed 53 lawsuits against the bureau, alleging error.² In response, Robert Marx, then assistant chief of the bureau’s geography division, invited colleagues from the research and statistical division to work with his team to try to solve the problems underlying the disputed count. Starting in 1981, the Census Bureau’s Geographic Operations Task Force, also known as the coffin twelve—a reference to their windowless meeting

Box 1. How Maps Shape the Count

To accomplish its mission of counting each person where each person lived, the Census Bureau needed two key pieces of information for every person: an address and the location of that address with respect to 2.5 million census blocks, 300 metropolitan areas, 50,000 municipalities, 3,200 counties, 435 congressional districts, and 50 states.

The 1980 census relied on three geographic tools to manage the data collection and tabulation.

1. Traditionally, field-workers using **enumeration maps** walked every census block in their assigned area and listed every address seen along with the block to which the address belonged. By writing down the address and associated block number, the enumerator geocoded the address, thereby transforming it from an address to a location on a map.

2. With the switch to a mail-out/mail-back approach, enumerators visited only those housing units or businesses that had not returned a questionnaire. A computer geocoded all other addresses by using an **address reference file** that cross-referenced addresses on the bureau’s mailing list with a census block. Like maps, those files described the streets, railroads, streams—census blocks in an area. Unlike maps, those files added information about the groups of addresses that belonged on each side of every street.

3. To classify each census block into the correct tabulation units, the census used a **geographic reference file** to assign each block to all of the higher-level geographic units such as city, tract, county, and state.

Not a mapmaker, the Census Bureau derived most of the content for its maps from secondary sources such as municipalities, state highway or transportation departments, private map companies, or federal agencies. All of the field maps produced for the data collection phase of the 1980 census were reproductions of maps from different sources, which the cartographic staff had stitched together. Errors and inconsistencies in the three geographic support components—maps, geographic reference files, and address reference files—added to costs and caused delays, errors, and frustrations in the 1980 census.

Source: “The Census Bureau Needs to Plan Now for a More Automated 1990 Decennial Census.” US General Accounting Office, February 9, 1983; <https://www.gao.gov/products/GGD-83-10>.

room—began to analyze the deficiencies in existing processes and develop a plan of action.

The goal was a single digital geographic database that stored and linked the map data, the address lists, and the geo-referencing information. However, no commercially available software system existed at the time. Instead, leveraging innovations in digital information processing and taking advantage of the growing availability of microcomputers, the bureau crafted a system from scratch. In addition to creating software, it had to encode the information from paper maps and field checks as well as add coordinates, classify and tag map features, define political and statistical boundaries and areas, and geocode addresses.

THE CHALLENGE

There was little debate about the end purposes the system would serve. Every one of the hundreds of thousands of temporary personnel hired to carry out a decennial census required a unique, detailed map for their enumeration area, the area that person canvassed. The maps had to contain recent information on roads, housing units, and political and statistical boundaries. And the capacity had to extend well beyond the simple ability to reproduce maps digitally, as earlier attempts by the US Geological Survey and its United Kingdom counterpart, the Ordnance Survey, had done. The bureau also wanted to integrate geographic information with addresses and with survey data on age, race, and income level. Other end users needed maps at varying scales for a wide range of purposes. For example, city governments wanted to know where to build new schools or health centers, and states wanted to link their information with county and local data to understand migration and commuting patterns, economic growth, and employment across regions or wanted to redraw legislative districts.

Marx's technical development team had to create a geographic information system (GIS)—a tool to capture, store, manipulate, and display spatially referenced data—that would serve the bureau's needs. Timothy Trainor, the Census

Bureau's chief geospatial scientist, said, "The few commercial GIS companies were not interested in urban planning or demographics; they were working on applications for the environment and natural resources like logging or mining."

To build a single, integrated system, Marx and other senior managers had to rethink essential business processes, design and develop new computer applications, create a digital cartographic file for the entire United States, and test the whole system before the next census, which was less than 10 years away.

The bureau had already done some of the work required to build a digital map. In preparation for the 1980 census, the geography division had assembled a series of geographic base files (GBFs), or computerized representations of road networks, by using dual independent map encoding (DIME) to link map features to address ranges. Each of the GBF/DIME files, as they were called, covered one of 338 metropolitan areas that together encompassed more than 60% of the US population but accounted for less than 5% of the nation's land area. The digital files included street names, address ranges, ZIP codes, water bodies, roads, and railroads. The files had to be converted into a format compatible with the TIGER database, amended to render roads as curves rather than just straight lines, and extended to support integration with survey data (See Textbox 2).

Expanding the functionality of this system required additional software to process data and detect input errors. For example, a clerk might mistakenly tag a road in the city of New York with coordinates that would put it in the nearby state of New Jersey. Geography division staff had to be able to enter coordinates, rename streets, and adjust address ranges easily and without error. Staff also wanted to be able to visualize the data and produce clean, corrected maps.

To improve coverage of suburban and rural areas, the bureau first had to obtain an accurate and consistent set of paper maps and then find a way to digitize them efficiently. Doing that meant

Box 2. Another Census Innovation: GBF/DIME

In fall 2017, 50 years after he first joined the Census Bureau, Donald Cooke boarded a flight from Los Angeles to Boston. From 30,000 feet above the earth's surface, Cooke could track his precise location throughout the five-hour flight on his tablet computer by using Google Maps. "I was in airplane mode, and I had Google Maps in cache for the whole route so I could fly along and look out the window and say, 'Oh, that's got to be Dubuque, Iowa'—and sure enough, it was," said Cooke. "There it is—on my Google Maps. The vision of being able to do that was so far beyond anything I think any of us in the digital map world could foresee."

Decades earlier, in 1967, Cooke and other members of a Census Bureau team called the New Haven Census Use Study had invented a technique called Dual Independent Map Encoding (DIME)—a breakthrough in geospatial information processing. Based on work by bureau mathematician James Corbett, DIME was a computerized representation of a street network wherein the record for each street segment included geographic coordinates, address ranges on both sides of the street, census tabulation areas, and ZIP codes. The file format used for storing DIME data was called a Geographic Base File (GBF). The advantage of the GBF/DIME system was that it automated address geocoding. With that invention, New Haven, Connecticut, became the first city in the world with a topologically integrated, geocoded database of its streets.¹

The Census Use Study, which ran from 1966 to 1969, was instrumental in creating and diffusing knowledge about computerized mapping. Spin-off companies founded by several team members—such as Cooke's Urban Data Processing, created in 1968 in Cambridge, Massachusetts, and Cooke's Geographic Data Technology, launched in 1980 in Burlington, Vermont—took advantage of the innovations stemming from the study to build commercial GIS applications. Online mapping platforms such as MapQuest, launched in 1997, and Google Maps, available in 2005, used some of the original algorithms developed by the study.

In preparation for the 1980 decennial census, the bureau adopted the GBF/DIME format as a standard for urban areas and created digital files for 300 metropolitan areas. The development of those files entailed collaboration and data sharing with state and local governments and planning bodies. In some cases, the Census Bureau or other federal agencies offered grants to support the local data collection work and provided computer programs, processing know-how, and a free copy of the completed file for each participating local agency. Local authorities in turn became increasingly aware of the potential for computer mapping, automated address matching, and spatial data analysis to meet local needs.

building partnerships at the federal level with the US Geological Survey—the primary civilian mapmaking agency; figuring out how to coordinate; and agreeing on a division of labor in order to minimize duplication of effort.

Marx's team also had to develop working relationships with other levels of government. To verify that the information on the maps was correct, the bureau had to compare the most current maps from multiple sources such as city

and state governments and private mapping companies. Maps from municipal planning departments usually had the most-up-to-date information on new housing units, recently built roads, and other changes that affected where people lived. However, those local maps had varying scales, formats, and degrees of accuracy. Transferring this geographic information into TIGER required field checks and quality controls.

The anticipated challenges did not end with the production of an accurate digital map, however. The next step was to match the bureau's address list for its mail operation with TIGER. A report by the General Accounting Office (GAO, the audit arm of Congress, now the Government Accountability Office) anticipated that there would be 106 million US households by 1990, meaning that there would be 106 million questionnaires to mail out, geocode, track, and tabulate. In metropolitan areas, the bureau's existing GBF/DIME files already matched addresses to census blocks—the smallest geographic areas used by the census to tally data—and it was only a question of updating information. But for rural areas, no equivalent list existed. The bureau would have to create a matched address list from scratch by canvassing every rural byway and recording the addresses of all housing units.

In addition to those technical issues, the bureau faced basic management challenges. Automation and changes in work flow required internal staff reorganization as well as greater collaboration across the bureau's separate divisions of geography, research, and field operations. The prospect of such changes triggered worker anxiety about possible job losses or the need to learn new ways of working. "When you're adopting new technology, there's a tendency for people to rely on experience rather than change their mind-sets," said Trainor. Ensuring internal buy-in and retraining people on new operating procedures required team building and leadership.

Finally, there was strong political pressure to keep costs down. Cost overruns had pushed the total bill for the 1980 census to over \$1 billion. A 1982 GAO report pegged the total estimated cost for the 1990 census at around \$4 billion and emphasized the need to "rethink census procedures in order to control costs."³ (In constant 1990 dollars, the 1980 census cost \$1.8 billion and the 1990 census cost \$2.6 billion.)

FRAMING A RESPONSE

To address these challenges, the 12-person Geographic Operations Task Force, convened by Marx, drew on lessons from past attempts to digitize map production. Digitization and automation had been on the bureau's agenda and the agendas of a few other federal agencies for more than a decade, but no one had attempted to build a single, comprehensive system for geospatial data processing and analysis.

Marx's 12-person Geographic Operations Task Force, which comprised cartographers, geographers, computer programmers, and statisticians from different divisions, anticipated that the development of TIGER would be a process of experimentation and adaptation. "We had committed to doing a massive thing that we didn't even know was possible," Trainor said. "And because of the enormity of the task, we had to start on part of the job before knowing whether it would be feasible to continue."

With backing from senior management in the bureau and its parent organization, the US Department of Commerce, Marx reached out to counterparts in the Geological Survey, housed in the Department of the Interior, who had worked on similar projects.⁴ Lowell Starr, head of research in the Geological Survey's National Mapping Program, had experimented with digital maps in the 1970s. Marx asked Starr to partner on the project, and in November 1981, the Geological Survey and the Census Bureau formed an interagency group to test the feasibility of building a digital, national map.

The first step was to identify an optimal scale for digitization. The Geological Survey had two national map series: one at a detailed 1:24,000 scale, often used for local planning, and another series at an intermediate scale of 1:100,000. The detailed maps displayed every local street, along with names, whereas the intermediate scale maps showed major roads and city street grids.

The intermediate scale maps were in the early stages of production and most of the country was not yet covered, but digitizing 1,800 of the lower-scale mapsheets rather than 54,000 of the fine-scale ones was a far less daunting prospect. As a test, the Geological Survey team scanned an intermediate scale map into a computer file, and the bureau manually digitized a set of detailed maps for the same area. The census team that examined the results found that the lower resolution map still had sufficient road information to meet census needs. The team therefore now had the base for TIGER. (See Textbox 3)

Clarification of goals was also one of the vital initial steps. As plans for TIGER coalesced, the Geological Survey and the Census Bureau consulted with other parts of federal government—such as the National Aeronautics and Space Administration and the US Forest Service—that, too, were thinking about how to manage geospatial data. The Geological Survey

and the Census Bureau met with experts such as Roger Tomlinson, who had pioneered a digital mapping system for the Canada Land Inventory in the 1960s. Tomlinson had observed the United Kingdom's efforts to digitize national map data, and he cautioned that despite enormous investment, the United Kingdom's computerized maps could not be linked to demographic or population data. TIGER would have to build that bridge. Advice from other agencies and experts made it clear that the final product would have to be useful for both census mapping and analysis.

By the end of 1982, the Geographic Operations Task Force had a broad plan for implementing TIGER, even if many of the details were still to be determined. What was clear was that the initiative would be costly, accounting for much of the bureau's spending during the next eight years.⁵ "It's easier to say *TIGER* than it is to make one," remarked decennial census associate director Roland Moore in 1984, when he and Marx went before a congressional subcommittee to ask

Box 3. What Is TIGER?

TIGER: Topologically Integrated Geographic Encoding and Referencing

Topologically. Topology is the scientific study of how points and lines on a map relate to each other to define a geographic area.

Integrated. The TIGER file is the computer database at the core of the TIGER system that relates, or integrates, the automated description of the earth's surface and boundary features. The topological structure of the TIGER file defines the location and relationship of streets, rivers, and railroads to each other and to the numerous geographic areas for which the Census Bureau tabulates data from its censuses and surveys. This essential geographic information will no longer exist only on separate, independent maps and lists; rather, it will be part of a single computer database.

Geographic. The TIGER file is the repository for storing the automated description of the geographic structure—the political and statistical areas and how they are nested—the Census Bureau uses to conduct its censuses and surveys.

Encoding. Encoding is the process of representing essential US geographic information in computer-readable form. TIGER uses a variety of encoding techniques such as automated map scanning, manual map digitizing, manual data entry and encoding, and sophisticated computer-file matching.

Referencing. Referencing is automated access to and retrieval of census-relevant geographic information for the United States.

Source: Robert W. Marx, "The TIGER System: Automating the Geographic Structure of the United States Census." *Government Publications Review* 13(2); March–April 1986:181–201; [https://doi.org/10.1016/0277-9390\(86\)90003-8](https://doi.org/10.1016/0277-9390(86)90003-8).

for \$194 million in funding. After hearing testimony from bureau staff and others about the potential benefits of TIGER and the drawbacks of existing approaches, the subcommittee agreed, and Congress later authorized the budget.

The schedule was tight. TIGER's operational deadline was in mid 1987—in time for the 1988 test run of the decennial census.

GETTING DOWN TO WORK

The Census Bureau and the Geological Survey, each based outside Washington, D.C., about 40 miles apart, became the twin poles of the TIGER effort. With a plan and a budget in place, the race to complete the project began.

Interagency cooperation

With confirmation that the intermediate-scale 1:100,000 maps were suitable for TIGER, the two agencies next had to determine whether, through a combined effort, their staff could digitize all 1,800 map sheets for the entire country in just a few years. In April 1983, Marx and Starr launched a pilot project to create a digital map database for the state of Florida. “Florida was the guinea pig,” recalled Eric Anderson who led the National Mapping Program’s digitization effort for TIGER. The Geological Survey chose a high-resolution scanner to digitize the 48 Florida map sheets. “The machine was designed for the textile industry to transfer designs to machines that printed fabric or wallpaper, but it had the resolution and speed we needed,” he said. Scanning a single sheet took about two hours. Operators then checked the quality of the output by using an edit station to correct any errors in the image—pixel by pixel. Software converted the output to digital files, but operators still had to verify each for accuracy and manually enter the data in computer-readable form for all features such as roads and waterways. The entire process took 20 or 30 hours for a single sheet, but the pilot provided much-needed evidence that separate map sheets could be digitized separately and then stitched together.

With that proof of concept in place, in December 1983 the Geological Survey and the Census Bureau formally agreed to share the labor required to complete, by mid 1987, a digital database containing transportation and water features for the 48 conterminous states. Both agencies benefited from the agreement. The Census Bureau would have a single, consistent, updatable national map for census takers, and the Geological Survey would get the nationwide digital map it wanted in three and a half years rather than 20—and with all road features checked and updated by Census Bureau staff.

Collaboration required each partner to make trade-offs. By the end of 1983, the Geological Survey had completed only about one-third of the 1,800 mapsheets for the country. And so, to meet the deadlines, it had to accelerate the production of paper maps and make compromises about what information to digitize. Aiming to save time, it agreed to focus on digitizing only map layers that were useful to the Census Bureau, thereby delaying digitization of topographic and land-cover information. For its part, the Census Bureau’s geography division would tag all road features as either *freeway*, *primary US highway*, *city street*, *footpath*, or *alley*.

“For the census, roads were very important in order to delineate [census] blocks and guide enumerators. They wanted much more detail for the roads than the Geological Survey could do,” said Anderson. “But for water and utility features, the Geological Survey wanted more detail than the census really needed. It was easy for them to exclude the detail once it was all digitized.” Indicating actual buildings was beyond the scope of the project, because Title 13 of the US code prevented the Census Bureau from sharing information that could be used to identify where individuals lived.

Both agencies built complementary high-volume digital production systems. As operators scanned paper maps and digitized them, Geological Survey cartographers manually edited

digital files to tag each feature—for example, by identifying which lines represented roads, rivers, railroads, and power lines. The cartographers sent the processed data on computer tape files to the Census Bureau, where geography division staff tagged each road according to type and assigned names.

Weekly meetings between the two agencies' staffs helped guide the ongoing work. At every meeting, Anderson said: "We would give them [the census] sets of data, and they would give us sets of queries. It was a very interactive process." The two sides handed over the data copied onto nine-track magnetic tapes. "Half of a 1:100,000-scale map was considered a 'product' we could share. That was how much could fit on one of those tapes," said Anderson.

Quarterly meetings between Marx and Starr and other senior staff from both divisions helped provide oversight and kept the focus on the end goal. "The leadership [from both agencies] was in sync, very determined, and positive," said Anderson.

Building staff engagement

To create TIGER, both the Geological Survey and the Census Bureau had to reorganize, retrain, and mobilize their existing workforces. Not all staff members were convinced of the need to "go digital," and some feared that automation would replace people with machines. Trainor, then a junior cartographer in the bureau's geography division, said, "Bob Marx realized early in the process that it was hard to get internal buy-in."

During 1983 and 1984, Marx, with backing from senior management in the bureau and its parent organization, the Commerce Department, reorganized the whole division. The human resources department authorized the shuffle as a short-term, temporary restructuring of the division. To help lead TIGER implementation, Marx turned to junior staff members like Trainor, whom he put in charge of a computer mapping team. "There were exceptional opportunities for

assistant geographers and cartographers to participate and contribute," recalled Trainor.

People at both agencies had to adapt to new routines and new equipment. To support the TIGER digitization effort, Anderson supplemented the Geological Survey mapping division's computer infrastructure by adding another high-resolution scanner like the one used in the Florida pilot, two plotters, 40 editing stations, and a variety of minicomputers. The Census Bureau's geography division, too, added new computers and other technology. Accustomed to scribing and composing paper maps, traditional cartographers had to "exchange their engraving tools and artwork cutting knives for computer workstations," said Trainor in a 1990 paper describing the transition to digital mapping.⁶ Instead of relying on visual judgment about how to display information on a map and which information to display, staff had to learn to program quantifiable rules that a computer could use to make those decisions. "It was definitely a retraining process for people," said Anderson. As professionals, most staff members understood the need to adapt. The digital-mapping division sent some of its people to graduate programs at universities that were at the forefront of research on geospatial information processing.

There was a steep learning curve for all involved. "TIGER was such a new thing," recalled Anderson. "We didn't understand the capabilities of the equipment, and we didn't know how to use it at first. We were constantly discovering how to improve the process and finding new techniques. It was not like a routine assembly line. We didn't even know what the assembly line looked like for this."

Given the initial uncertainty about how to proceed and what was achievable, the team set short-term goals and reassessed them every six months. It took two years to optimize the computing environment, develop standard procedures, and ensure interoperability with the census. "Then it was four years of intense and hard work to get through it all," said Anderson.

Initial concerns about job losses dissipated. Both the Geological Survey and the Census Bureau had to hire *additional* personnel to perform software development as well as the data processing required for digitization.

Updating maps

With the conversion of paper maps into computer-readable digital files under way, the next stage of TIGER implementation involved checking and correcting the data. To manage the task of collecting up-to-date geographic information at local levels, Marx established a core geographic support unit in each of the bureau's 12 regional offices. "This was the first time they had dedicated geography staff in every regional census office," said Jon Sperling, who joined the New York office as a geographic coordinator in 1983 and remained with the census until 2001. From 1983 to 1985, Sperling expanded the geographic support team in the New York regional office to 10 full-time staff and another 30 temporary workers on multiple shifts as the decennial deadline loomed.

The geographic support units collected current information on streets, boundaries, and other features to check and update the Geological Survey maps. "We had to call local governments, utilities, and other sources to get recent street maps," said Sperling. In some areas, "we were using aerial photos and tracing the streets." A few commercial mapping companies also had current maps the geographers could use, but all sources had errors and had to be checked. Some sources were good just for the road networks, and others for street names or addresses only. To assess quality, Sperling's staff did spot checks on each source by driving to a random road or address.

Much of the data in the Census Bureau's GBF/DIME files for urban areas also required updating. The 338 files contained street names, address ranges, postal codes, and administrative and statistical boundaries. Accounting for more than 60% of the population, the massive amount of information in the files was neither current nor

complete. The GBF/DIME files were often 6 to 10 years old. Local planning departments had maintained some of them and adapted them to their needs.⁷ However, because local practices varied, everything had to be checked and reentered manually by geographic support staff in the 12 regional offices. "We needed current information from local and state governments for streets and names," said Sperling. "We collected address information from paper maps, tax assessor files on microfiche, and other sources." Staff in the regional geographic support units drove streets in communities across the nation to check and validate the quality of the address data. Using the locally collected maps, workers entered missing streets or street names, fixed misaligned features, and standardized the names of all roads.

Clerical staff in the New York regional office worked in two shifts to edit the digital files and enter the current geographic information for the region. To transfer data between headquarters and regional offices, staff used computer tape files, leased telephone lines, and, as technology evolved, floppy disks. "It required a lot of coordination between the field units and headquarters to get consistent data quality and formats," said Sperling. "The process was iterative, because new challenges came up all the time and we had to find solutions." Quarterly conferences brought together staff from all field offices and headquarters to review progress and establish common standards for developing TIGER.

Getting maps into and out of TIGER

As the team collected and digitized maps, the key questions for TIGER's architects were how to organize and how to manipulate the vast amounts of geospatial data efficiently. "We [began] digitization while working out the database structure," said Trainor. "We didn't know exactly what the file should or could look like. We started off with a blank piece of paper."

The Census Bureau's operational requirements and the limitations of available computing hardware shaped the design of TIGER. First, the

database had to support geocoding—the assignment of a mailing address to a census block—and map production. Second, the data had to be structured to identify inconsistencies between the three key geographic products: maps, address files, and geographic reference files. And third, the vast amount of data posed a special problem. No single hard drive could store the entire national dataset.⁸ Plus, speed mattered. The system had to produce maps for more than 300,000 separate enumerator assignment areas and more than 39,000 units of local government quickly, but at that time computer processors were slow.

Marx and Frederick Broome, the manager for mapping operations, turned to Hanan Samet, a computer science professor at the University of Maryland, for advice with regard to the underlying structure of the TIGER system. Samet proposed a networked database structure that would reduce the amount of disk space required and optimize the speed of data retrieval. The final TIGER system functioned as a single file even though it existed on multiple disks. Implemented as a linked set of subfiles, directories, and lists, the database integrated road information with address ranges and geographic boundaries.⁹ And the system's software ensured that changes to any item were immediately transferred to all other files simultaneously.

The cartographers and geographers worked closely with programmers in the bureau to write a host of software applications to build and maintain the database as well as routines to label files, partition the data into 3,200 counties, plot maps, and geocode addresses. “We designed a menu for inputting map features and coding them,” Trainor said. “All of the input software and output software, the geoprocessing routines, the quality and validity checks—all of it had to be developed.”

TIGER development also had to contend with rapid advances in computing during the 1980s. The age of paper punch cards and mainframe computers was coming to an end. The rapidly

growing availability of minicomputers, hard drives, graphical display systems, input devices, printers, and plotters facilitated high-volume digitization and map production. For the TIGER project, that revolution made certain tasks easier, but it also created new challenges. Although workstations and faster processors sped up operations, software required constant revision as technology changed.

OVERCOMING OBSTACLES

As the 1990 census approached, there were signs of trouble on the horizon. When the project started, the goal was to produce maps for 300,000 enumeration areas by the end of 1988 to enable canvassers to check millions of addresses before mailing people their questionnaires. But it took longer than anticipated to make the maps. Software failures and delays getting the current geographic information from regional offices ate into map production time.

The TIGER development team also had to find an alternative to the traditional method of producing maps, which required cartographers to sit in front of computer stations and spend hours generating each type of map. The bureau did not have enough such workstations or cartographers to complete the work in the allotted time using the existing techniques.

In the spring of 1987, Trainor's mapping-operations team decided that the maps should come from a production system that automatically determined scale, type of map sheet, labels, text, and placements of all features.¹⁰ To reduce computer-processing time, the mapping-operations team built a secondary database drawn from the TIGER system—referred to as the cartographic extract. Staff used that extract—a simplified version of TIGER data—to print out maps.

The task of geocoding address lists started as maps became available. The process had two parts. First, one bureau team compiled and checked addresses because the accuracy of the mail-out/mail-back survey depended on the quality and coverage of such information. Then the geography division worked with census staff to geocode

addresses by tagging each address with the census block in which it was located, along with the enumeration area, city, county, and state. (Because Title 13 of the US Code protected the confidentiality of information collected by the Census Bureau, TIGER contained only address ranges for every census block, not specific addresses.)

In the two years leading up to the census, the bureau worked to compile the mail-out list from multiple sources. For cities and suburbs, the bureau purchased lists from marketing companies. In rural areas, field-workers canvassed each road to pinpoint addresses. To ensure deliverability, the Postal Service cross-checked each address against its internal database, as it had done in previous decades. For the 1990 census, the aim was to have completed a Postal Service check of the full mailing list of roughly 106 million households by March 1989.¹¹ That deadline gave the bureau a year to correct any bad addresses before questionnaires started going out in the mail.

The change in the approach to map production helped speed up the process, but supply continued to lag behind demand, and that spelled trouble. Some of the field maps required for the canvassing were not available in time to complete the address list by early 1989. In March, the bureau submitted a partial list of about 90 million addresses to the Postal Service for validation and added another 16 million addresses later.

When the 1990 census launched, many addresses remained unchecked. The operation went forward, as required under the US Constitution, but errors in the final address file contributed to an overall response rate that was 10 percentage points lower than in 1980. Disputes over the count sparked 22 lawsuits against the bureau, alleging miscount and error—fewer than in 1980 but still cause for concern.¹²

Marx's team and other senior managers in the bureau threw themselves back into the effort to improve procedures for updating maps and to find a more reliable method for maintaining a national

address list linked with TIGER. The urgency of their work deepened in the early 1990s with the bureau's decision to introduce the American Community Survey, an annual sample survey designed to supplement the decennial count. Every month, the bureau asked questions of a representative group of 250,000 households with a view to collect detailed information on demographics, habits, languages spoken, occupations, places of work, housing, and other characteristics. Such a monthly sample survey promised to provide more-frequent and more-timely information for communities and businesses. However, to draw a representative sample, the bureau needed an accurate, current set of geocoded addresses for all of the housing units in the country.

The geography division proposed to meet that challenge by partnering with the Postal Service—the best source of mailing addresses in the country—to create the US Census Bureau's Master Address File. Both agencies had to seek special authorization to make this project work because of US laws about the confidentiality of individual addresses. In 1994, Congress passed the Census Address List Improvement Act, permitting the sharing of address information. The legislation also authorized the Census Bureau to share the list with state, local, and tribal governments so that officials could identify additions or deletions.

With that authorization in place, the Postal Service began to supply a copy of its address list to the Census Bureau twice a year. To create the Master Address File, staff in the bureau's geography division merged the Postal Service's address list with the list used for the 1990 census and updates from local partners. Using TIGER, staff then geocoded all the addresses, assigning each to a unique census block. Addresses that did not geocode pinpointed—within a specific ZIP code and along a mail route—where TIGER's information was out of date or incorrect. Regional office staff could then field-check the area or contact local governments to confirm changes such as new roads or houses.

Those steps helped, but the address-updating and geographic-updating processes still did not work perfectly. The geography division maintained the Master Address File and TIGER as two separate but linked databases, managed through a suite of software applications developed in-house. But changes made in TIGER sometimes broke the link that geocoded postal addresses to a census block. In 2000, TIGER failed to geocode 4.5 million of 175 million addresses included in that year's census. To resolve those issues, the bureau had to merge information from separate Master Address File and TIGER databases into a single relational database and upgrade its software.

After the 2000 census, the bureau decided to outsource the task of transforming TIGER and overhauling the suite of software tools used by the geography division. By 2000, commercial GIS software had outstripped the capabilities of the homegrown tools the geography division used. Private firms as well as many local governments and others were beginning to modernize their GISs faster than the bureau. Further, TIGER could not read geospatial data files generated by commercial platforms, and local and state governments felt frustrated by having to submit updates on paper maps.¹³

The bureau contracted with two private firms—Harris Corporation and Oracle Corporation—to merge the underlying databases of TIGER and the Master Address File and to correct the locations of roads, streams, boundaries, and other features.¹⁴ The contractors then transferred the database into a new software system it created for spatial information.

Merging the two separate files for addresses and locations into a bureau-wide Oracle database increased the efficiency of processing and the development of software and facilitated Internet access to the geographic information. The geography division also changed the distribution format of TIGER data from line files to a shapefile, a format for storing and exchanging geospatial information that enabled users to easily

download the TIGER dataset into desktop GIS applications.

The improved functionality and accuracy enabled the use of handheld GPS devices for field operations and enabled workers to collect the GPS coordinates of all housing units in the country before the 2010 census.

ASSESSING RESULTS

From 1990 to 2010, TIGER supported three decennial censuses, six economic censuses, numerous special censuses and census tests, several monthly household surveys, and the annual American Community Survey. At its debut, the database contained 23 million road features and 9 million streams, railroads, and dams. The Census Bureau used TIGER to print almost 40 million maps and maintain 140 million georeferenced addresses.

The switch to an automated mapping system reduced time and costs, improved precision, and expanded the ability to handle large datasets consistently. The bureau produced 1.3 million maps in three years for the 1990 census. For the 1980 census, it had taken four years for a combined workforce of 1,600 from the Census Bureau and a private contractor to draw 32,000 map sheets. TIGER had improved the overall quality and accuracy of census enumeration maps and tabulation maps. Digitization, too, facilitated address and street updates, thereby enabling the bureau to maintain a current list. With the distribution of computing equipment and TIGER data, each regional office acquired the ability to print the maps needed for fieldwork.

The geography division spent close to \$300 million to develop TIGER for the 1990 census, and \$500 million to modernize the system in the 2000s. However, said Trainor, the bureau's chief geospatial scientist, "the up-front investments required were offset by the benefits of the system in terms of greater flexibility and accuracy, and those benefits continue to accrue." Nevertheless, despite automation of the geographic support functions, the total cost of the decennial census

almost doubled every decade from 1970 to 2010, caused in large part by declining mail-back response rates. As more and more people chose not to participate in the census, the bureau needed more census takers to knock on doors.

One of the first examples of data sharing between agencies, TIGER laid the groundwork for similar collaboration across parts of the US government. The partnership between the Census Bureau and the Geological Survey spurred establishment of the Federal Geographic Data Committee to coordinate geospatial data management across departments. In 1994, the partnership also sparked creation of the National Spatial Data Infrastructure to collect and integrate data from all parts of government and make it available to the public.

TIGER was also open data from the start. To preview how external audiences might use the data, in April 1988 the geography division released a portion as a line file: a single table that contained roads, coordinates, address ranges, ZIP codes, census block numbers, and geographic area codes for Boone County, Missouri, where the bureau had conducted its final operational test of the system.¹⁵ The release did not include any software to support translation, visualization, or mapping of the data. “We just put the data on a reel of magnetic tape and said, ‘We have an example available.’ We sent out copies to everyone who wanted a copy,” Trainor recalled.

The response to the initial release was overwhelming. Software vendors took advantage of the preview version to develop applications for translating and importing the data into the rudimentary GIS platforms that were commercially available at the time, and local governments and private sector firms immediately wanted the files for additional counties. “People knew what we were up to. People were kind of sitting on the edge of their seats,” Trainor said.

Under its mandate to make its data freely available to the public—unless restricted by privacy concerns—the bureau provided copies on magnetic tape, charging only for the cost of

duplication. Charging for the data, the way one might charge for a product, was not an option. “That proposal went over like a lead balloon,” Trainor said. Moreover, anyone who purchased a copy could immediately duplicate the information and sell it.

A copy of the nationwide dataset on tape, including all counties plus Puerto Rico and outlying territories cost about \$90,000. “It was, effectively, free,” said Don Cooke, one of the first buyers of the TIGER dataset and founder of Geographic Data Technology, a company that developed map databases. “Let me say ‘free,’ considering the Census Bureau spent \$200 million creating this database.” Cooke’s company was one of several that contracted with the bureau to process data during TIGER development. With their deep knowledge of the TIGER system, these vendors were prepared to commercialize the line files. In the early 1990s, Cooke’s company developed and sold Dynamap—a translated and updated variant of the TIGER road network with addresses included—to a variety of clients, including shipping firm FedEx.

With the transition to CD-ROM, by 1991 the price of a set of CDs covering the entire nation had decreased to approximately \$10,000. During the next two decades, improvements in data storage technology and the growth of the World Wide Web progressively lowered barriers to access. And by 2010, users could download the data free from the bureau’s website.

Designed to meet the needs of the Census Bureau, TIGER had enormous uptake beyond the agency’s own activities. Because of the geography division’s local update and outreach efforts, state and local governments became accustomed to working with TIGER data and applied it to local planning. Applications ranged from river basin and coastal zone management in the Carolinas to analysis of ward-level voting patterns in Wisconsin. Local governments benefited from the access to digitized geographic information. In 1990, the city of Chicago could order the entire digital database of Cook County on CD-ROM for about \$2,000

rather than pay as much as \$1 million to build the same product from scratch.¹⁶ By 2010, barriers to information were even lower, and any organization could download the data from the bureau’s website. Nonprofit organizations such as PolicyMap, based in Philadelphia, used TIGER data to develop easy-to-use online mapping tools that assisted local planners and decision makers (See Textbox 4).

For the burgeoning GIS industry, TIGER was a blockbuster. Marx’s team had developed TIGER to solve a Census Bureau problem, and no one had given much thought to how others outside government might use the database. But even as TIGER development neared completion and a working prototype came online, team members began to envision the possibilities, as did the private contractors that had helped carry out some of the data processing work.

Entrepreneurs used the data to create new applications and tools—from car navigation systems to online maps, to sophisticated decision tools. TIGER’s road network data formed the basis of online mapping platforms MapQuest, OpenStreetMap, and the early Google Maps. In the early 2000s, Google licensed Cooke’s Dynamap for an early iteration of Google Maps, essentially importing all of the geospatial information the bureau had amassed.

As of 2017, many mapping and GIS companies were continuing to rely on TIGER for various aspects of their product lines. Google took the legal and statistical boundaries defined in TIGER as authoritative sources for the boundaries displayed online. Mapzen, an open and accessible mapping platform that published the code for all of its software, used TIGER’s address range information to provide search and routing services.

Box 4. PolicyMap

“TIGER data is so integral to everything we do that we end up not thinking about it a lot of the time,” said Bernie Langer, senior data analyst at PolicyMap. Spun out of the Philadelphia Community Reinvestment Fund in 2007, PolicyMap provided a one-stop shop of nationwide geographic information for policy makers and community groups. Founder Maggie McCullough envisioned a single, seamless mapping platform with jargon-free tools that let users focus on the policy and planning decisions at hand rather than on the technicalities of how to manipulate datasets.

Prior to the advent of PolicyMap there had been no single provider of the vast amount of geospatial data available for the country. Elizabeth Nash, vice president of product development and one of PolicyMap’s original employees, said the key questions in building the platform were, “What data did our users need? and, How do we make it simple for them in terms of display and navigation? Folks were fed up with downloading census data and not knowing how to use it or how to map it.” Among the platform’s users by 2010 were financial institutions, federal and local government agencies, universities, and community groups. Policy makers turned to the platform for answers to diverse questions—from how to identify the availability of housing in a particular county to how to evaluate a neighborhood’s access to healthy food.

Underpinning the varied applications and analyses was TIGER. PolicyMap’s platform combined a library of 150 data sources, with 80% of the information drawn from open-access federal and state sources and 20% from private firms. TIGER defined areas—jurisdictions, statistical areas, and school districts—and formed the basis of the geographic analysis of social, demographic, and economic data. “TIGER has also been crucial to understanding changes in boundaries over time,” said Nash. The differences between areas in 2000 and 2010 help identify areas of growth or consolidation—an important consideration in local planning decisions.

And Zillow, the online marketplace for real estate, drew on TIGER to display neighborhood demographic data for each property.

REFLECTIONS

When the Census Bureau and the Geological Survey launched a joint initiative in 1981 to create a digital representation of the United States, each focused on its own mandate and mission. Neither anticipated the impacts the project would have. At the time, both were leaders in digital-mapping operations and had the federal funding to tackle the enormous job of compiling geographic data for the entire nation in a matter of years. By 2017, both agencies remained important sources of freely available geospatial data for the country, but the private sector innovations that TIGER had helped spur were dominating the GIS landscape.

The concept of open data was at the core of the US Census Bureau's mission and legal mandate, and even though TIGER had not been conceived as an open-data project, the development team faced challenges in collecting, digitizing, and publishing information in accessible formats common to open-data initiatives that later blossomed around the globe. By using available resources, skills, and technology, the team overcame those challenges.

Not being in the business of creating maps, the bureau relied on partnerships with the Geological Survey and extensive outreach to state and local governments to collect and verify data. Early tests of map digitization and the Florida pilot confirmed that the goal of a geospatial database was achievable, and the bureau helped optimize the processes used in converting paper maps to digital information. The bureau hired private contractors to assist with the digitization effort for TIGER, thereby fostering the growth of an ecosystem of potential users who had the technical know-how and the computing hardware to transform the data for their individual purposes.

Looking back on the progress made over the decades, Atri Kalluri, associate director of the

Census Bureau's decennial census technology, said: "The geospatial industry has made tremendous progress. I feel that most of it was achievable because of the data TIGER made available in 1990. Only when you have the data can you conceive of what to do with it. If you didn't have the data, you wouldn't even have the thought process to do something with it."

Without having to incur huge data collection costs, users could immediately leverage the data to develop new applications and tools. OpenStreetMap put the data online and crowdsourced updates and extensions. Other companies invested in adding value to the data or transforming it and built businesses off it. "There is no way I could have run Geographic Data Technology without being able to get the stuff effectively for free from the Census Bureau," said company founder Cooke.

One of the key factors that made TIGER so valuable to users was its high degree of data accuracy and reliability. Given the high stakes attached to census results, TIGER had to be built to exacting standards. Despite a small percentage of errors and missing information in the final TIGER dataset, much of the geographical information was correct.

Moreover, the bureau's continued use of TIGER for decades meant the data would be updated and maintained over the long term. And users could trust that the government would continue to publish the data openly and in accessible formats.

The US government's open-data approach set a precedent for other nations as they looked to digitize national geospatial data. "The European Union did a study on whether maps should be crown copyright or be in the public domain—digital maps," said Cooke. "They largely accepted . . . the US example of TIGER's going into public domain, saving a lot of money, kicking off a lot of industries, allowing FedEx and UPS to operate more efficiently."

References

- ¹ National Research Council, *Modernizing the U.S. Census*, Barry Edmonston and Charles Schultze, eds., 1995, pp. 47-48; <https://doi.org/10.17226/4805>.
- ² Vincent P. Barbara, Richard O. Mason, and Ian I. Mitroff, "Federal Statistics in a Complex Environment: The Case of the 1980 Census." *American Statistician* 37(3); August 1, 1983:203–12; <https://doi.org/10.1080/00031305.1983.10483103>.
- ³ "The Census Bureau Needs to Plan Now for a More Automated 1990 Decennial Census." U.S. General Accounting Office, February 9, 1983; <https://www.gao.gov/products/GGD-83-10>.
- ⁴ Marx, Robert W. "The TIGER System: Automating the Geographic Structure of the United States Census." *Government Publications Review* 13, no. 2 (March 1, 1986): 181–201. [https://doi.org/10.1016/0277-9390\(86\)90003-8](https://doi.org/10.1016/0277-9390(86)90003-8).
- ⁵ Tomasi, Silla G. "Why the Nation Needs a TIGER System." *Cartography and Geographic Information Systems* 17, no. 1 (January 1, 1990): 21–26. <https://doi.org/10.1559/152304090784005804>.
- ⁶ Timothy F. Trainor, "Fully Automated Cartography: A Major Transition at the Census Bureau." *Cartography and Geographic Information Systems* 17(1); January 1, 1990:27–38; <https://doi.org/10.1559/152304090784005831>.
- ⁷ Carbaugh, Larry W., and Robert W. Marx. "The TIGER System: A Census Bureau Innovation Serving Data Analysts." *Government Information Quarterly, Special Issue Symposium on Geographic Information Systems*, 7, no. 3 (January 1, 1990): 285–306. [https://doi.org/10.1016/0740-624X\(90\)90026-K](https://doi.org/10.1016/0740-624X(90)90026-K).
- ⁸ Broome, Frederick R., and David B. Meixler. "The TIGER Data Base Structure." *Cartography and Geographic Information Systems* 17, no. 1 (January 1, 1990): 39–47. <https://doi.org/10.1559/152304090784005859>.
- ⁹ *Ibid.*
- ¹⁰ Trainor, Timothy F. "Fully Automated Cartography: A Major Transition at the Census Bureau." *Cartography and Geographic Information Systems* 17, no. 1 (January 1, 1990): 27–38. <https://doi.org/10.1559/152304090784005831>.
- ¹¹ 1990 Census of Population and Housing: History. U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1993
- ¹² 1990 Census of Population and Housing: History. U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1993.
- ¹³ Trainor, Timothy. "U.S. Census Bureau Geographic Support: A Response to Changing Technology and Improved Data." *Cartography and Geographic Information Science* 30, no. 2 (January 1, 2003): 217–23. <https://doi.org/10.1559/152304003100011054>.
- ¹⁴ Broome, F.R., and L.S. Godwin. "Partnering for the People: Improving the US Census Bureau's MAF/TIGER Database." *Photogrammetric Engineering and Remote Sensing* 69 (October 1, 2003): 1119–23.
- ¹⁵ Carbaugh, Larry W., and Robert W. Marx. "The TIGER System: A Census Bureau Innovation Serving Data Analysts." *Government Information Quarterly, Special Issue Symposium on Geographic Information Systems*, 7, no. 3 (January 1, 1990): 285–306. [https://doi.org/10.1016/0740-624X\(90\)90026-K](https://doi.org/10.1016/0740-624X(90)90026-K).
- ¹⁶ Patrick Reardon, "Computer map takes a very close look at U.S." *Chicago Tribune*, March 4, 1990.

Innovations for Successful Societies makes its case studies and other publications available to all at no cost, under the guidelines of the Terms of Use listed below. The ISS Web repository is intended to serve as an idea bank, enabling practitioners and scholars to evaluate the pros and cons of different reform strategies and weigh the effects of context. ISS welcomes readers' feedback, including suggestions of additional topics and questions to be considered, corrections, and how case studies are being used: iss@princeton.edu.

Terms of Use

Before using any materials downloaded from the Innovations for Successful Societies website, users must read and accept the terms on which we make these items available. The terms constitute a legal agreement between any person who seeks to use information available at successfulsocieties.princeton.edu and Princeton University.

In downloading or otherwise employing this information, users indicate that:

- a. They understand that the materials downloaded from the website are protected under United States Copyright Law (Title 17, United States Code). This work is licensed under the [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.
- b. They will use the material only for educational, scholarly, and other noncommercial purposes.
- c. They will not sell, transfer, assign, license, lease, or otherwise convey any portion of this information to any third party. Republication or display on a third party's website requires the express written permission of the Princeton University Innovations for Successful Societies program or the Princeton University Library.
- d. They understand that the quotes used in the case study reflect the interviewees' personal points of view. Although all efforts have been made to ensure the accuracy of the information collected, Princeton University does not warrant the accuracy, completeness, timeliness, or other characteristics of any material available online.
- e. They acknowledge that the content and/or format of the archive and the site may be revised, updated or otherwise modified from time to time.
- f. They accept that access to and use of the archive are at their own risk. They shall not hold Princeton University liable for any loss or damages resulting from the use of information in the archive. Princeton University assumes no liability for any errors or omissions with respect to the functioning of the archive.
- g. In all publications, presentations or other communications that incorporate or otherwise rely on information from this archive, they will acknowledge that such information was obtained through the Innovations for Successful Societies website. Our status (and that of any identified contributors) as the authors of material must always be acknowledged and a full credit given as follows:

Author(s) or Editor(s) if listed, Full title, Year of publication, Innovations for Successful Societies, Princeton University, <http://successfulsocieties.princeton.edu/>

