THE TIGER SYSTEM: GEOGRAPHIC IMPROVEMENTS FOR THE 1990 CENSUS AND BEYOND

Linda M. Franz, U. S. Bureau of the Census[1]
Washington, DC 20233

Introduction

The Geography Division of the Bureau of the Census has undertaken the development of a new geographic support system called the TIGER System (TIGER being an acronym standing for Topologically Integrated Geographic Encoding and Referencing). This effort is one of several by the Census Bureau to automate selected activities in order to improve the census count where it is feasible, cost effective, and appropriate to do so. This paper describes the geographic functions the TIGER System will support and describes the TIGER System itself. This paper also describes the processes for developing and using the TIGER System to support the 1990 decennial census and subsequent censuses and surveys.

Geographic Support

Geographic support means providing the wherewithall for two critical activities in connection with censuses and surveys: assigning individual responses to the correct geographic location; and classifying those geographic areas into the correct data tabulation units.

Three major tools allow the accomplishment of these geographic support objectives: maps, address reference files, and geographic reference files. Census Bureau maps guide data collection, depict the geographic areas for which data are tabulated and, in some cases, present data. Address reference files associate addresses corresponding to individual census or survey responses with their geographic location, or position on the earth's surface. Geographic reference files catalog the various geographic areas (for example, state, county, census tract, and so on) and define their relationships with one another so as to allow the presentation of data by geographic area in an orderly, understandable fashion.

For the 1980 census (and, to the extent they were used, for earlier censuses) the Census Bureau prepared these three major geographic tools in independent clerical operations. This was a massive undertaking. One would expect mistakes with an operation whereby literally millions of geographic areas and identifiers were hand-entered onto maps and into two independent computer files. As expected, the Census Bureau did not achieve perfect consistency among these three geographic products The overwhelming majority of the identifiers in the three products matched, but where inconsistencies did exist, problems arose. These problems occurred in data collection, data tabulation, and data use. Further, the labor-intensive processes by which the Census Bureau generated these products contributed to their late delivery. This brings us to our discussion of what the TIGER System is.

The TIGER System

The "I" in TIGER, which stands for "integrated," is a key to understanding how the TIGER System will allow the Census Bureau to overcome some of the difficulties mentioned above with geographic support. The TIGER File, the central component that along with its associated applications programs constitutes the TIGER System, will be a large database system with a computer-readable or "digital" map as its foundation. The TIGER File will cover the entire United States, Puerto Rico, the Virgin Islands, and the other Outlying Areas for which the Census Bureau conducts the census (Guam, American Samoa, the Trust Territory of the Pacific Islands, and the Northern Mariana Islands).

Integrated into this single digital map data base will be all the information needed to allow (through the use of applications programs) census mapping by computer-driven plotting devices, the assignment of census responses to geographic locations, and the cataloging and relational definition of areas. This means that all mapping and geoprocessing will be in complete agreement and that the problem of inconsistency between geographic products will be solved. The TIGER System also will allow us to produce maps of consistently high quality, which was not always possible in the past. The internal structure of the TIGER File was created by applying the concepts of topology, a branch of mathematics that describes the interrelationships of points, lines, and areas on the surface of a sphere. Construction of the file using these principles allows the creation of an elegant, powerfully self-checking computer data base containing all mapped features and boundaries, their names, address ranges, and geographic codes.

We store the geographic information in the TIGER File using lists and directories automatically linked and cross referenced to each other. The line segments, points, and geographic area classification codes all are tied together so that when any one of them changes, the other components and relationships change too.

Developing the TIGER System

The Census Bureau must complete the development of the TIGER File and the associated components of the TIGER System in time to support the activites of the 1990 decennial census. The three main stages of the file-building operation are as follows: First, we must create the initial digital map data base--the underlying map image in computer-readable form--so that the computer can manipulate all of the streets, rivers, railroads, and so forth.

Second, we must update this digital map data base with other information that makes the map useful for Census Bureau activities---adding new streets where subdivisions have been built; entering the names for all streets, rivers, and railroads so that Census Bureau field staff can

orient themselves; inserting the address ranges that go with each section of a street in the major urban areas; and adding the boundaries and names of all the political and statistical areas for which the Census Bureau tabulates data. Finally, we must use the file and enhance this information based on the results of the early file-use operations. Here the file-building and file-use activities interconnect because the file-use operations contribute information gathered via direct field staff observation to make the digital map data base match what exists on the ground at that point in time.

To help us achieve this ambitious undertaking, the Census Bureau entered into a major cooperative agreement with the United States Geological Survey (USGS) for the development of the initial digital map data base. The USGS, an agency within the Department of the Interior, is the "Nation's map maker". The USGS and the Nation will benefit from the early availability of this computer-readable map base as a result of the rigid time schedule for completing 1990 census preparatory operations.

TIGER File Building

The creation of the initial digital cartographic data base for the "lower 48" states and the District of Columbia as specified in the cooperative agreement between the Bureau of the Census and the USGS is complete. For the remainder of the United States and its territories the Census Bureau "digitized," or captured in computer-readable form on a point by point basis, information from printed maps. Theoretically the Census Bureau could have created the entire automated digital map data base in this fashion, but the Census Bureau and the USGS, by working cooperatively, achieved their mutual goals without duplicating cost and effort. The USGS used automated scanning equipment to capture line information from their 1:100,000-scale national map series. Each 1:100,000-scale map sheet consists of four separate layers used to make the map (roads, water, railroads, and miscellaneous transportation features). The USGS placed each required layer on a rotating drum where a light beam read the features shown on the base maps to create a computer file.

As a next step, the Census Bureau used its digitizing equipment to enter road type codes (freeway, city street, jeep trail, and so on) into each USGS road file by "tagging" each road segment with a classification code. The Census Bureau completed this task in May 1987. While the Census Bureau tagged the road files, the USGS entered the corresponding water, railroad, and miscellaneous transportation classification codes into the appropriate computer files for each map sheet. The Census Bureau transmitted the tagged road files to the USGS, where they incorporated this information, along with their three files, into the National Digital Cartographic Data Base.

The next major file-building step is called "vertical integration". The Census Bureau performs this process after the USGS returns the four computer files for each 1:100,000-scale map sheet. Vertical integration "sandwiches" together the four layers and aligns all features along the map edges.

At this same time another major file-building step was carried out by the Census Bureau's 12 regional offices. Geographers in the regions made contact with state, county, and local government officials in order to obtain up-to-date information on where new streets are being built and on the names for all streets, railroads, water features, and so on. Once they acquired these source materials, the regional office geographic staff plotted the updates on paper copies of the USGS 1:24,000-scale map series and wrote the feature name information on keying worksheets.

Four of the 12 Census Bureau regional offices are operating as field digitizing sites. The four are Atlanta, GA, Boston, MA, Dallas, TX, and Denver, CO. Using stand-alone digitizing stations, staff in these offices insert the updated feature information plus all the feature names into the vertically integrated 1:100,000-scale digital map data base for those areas beyond the coverage of our address reference files, called GBF/DIME-Files for the 1980 census. For the 345 areas covered by GBF/DIME-Files--the major urban areas of the Nation--contractors are inserting the updated feature information noted by the regional office geographic staff into the GBF/DIME-Files. These contractors also are inserting the 1980 geographic area codes into the GBF/DIME-Files. The following paragraphs describe these areas in more detail.

It is to the updated 1:100,000-scale computer-readable map base, which at this point will show an up-to-date feature network with names, that we are applying identifiers for the various geographic areas by which census data are collected and tabulated. These geographic areas fall into three major categories--political, statistical, and census administrative areas. Political areas are those with their basis in Federal, state, or local law. Some familiar examples are states, counties, incorporated cities, and American Indian reservations. Statistical areas are delineated for data presentation purposes, in some cases in cooperation with state and local officials and data users. Included among the statistical areas are census tracts, block groups, metropolitan statistical areas, and urbanized areas. Census administrative areas are those delineated for such purposes as assigning work to enumerators; in some cases administrative area delineation criteria are set so as to allow these areas to serve data presentation needs as well. An example is the census blocks where we are asking state redistricting officials to review the maps and ensure that the data collection blocks meet their redistricting needs. We describe this activity--the Block Boundary Suggestion Project --in more detail below.

The first geographic areas that we are incorporating into the digital map data base will be those used for the 1980 census. There are three reasons for this starting point: First, this information is available now. Second, because much of this information will not change between now and the 1990 census, having

this 1980 base means that we only need to enter changes later to form the 1990 areas. Further, establishing the 1980 political and statistical areas in the TIGER File allows for the automated comparison of 1980 and 1990 area information. The Census Bureau plans to automate the preparation of equivalency lists between the 1980 and 1990 census tracts for the 1990 census.

As part of another work step that is complete, staff in all 12 Census Bureau regional offices worked with state and local officials on the delineation of some of the statistical areas for 1990--specifically census tracts, block numbering areas, and block groups. These are the first statistical areas we delineate because we use them in the earliest phase of the census, the data collection phase.

Another program that the regional office geographers completed in cooperation with state officials is the Block Boundary Suggestion Project. This is the first phase in our operations to provide data to the states for redistricting of congressional and other legislative districts pursuant to P.L. 94-171. Under this program, which is voluntary for the states, state officials reviewed maps in the regional offices to identify features that they believed the Census Bureau should hold as block boundaries. The state officials did this to increase the possibility for voting district boundaries, when they are determined, to coincide with census block boundaries. Several years from now, states choosing to do so can delineate groups of 1990 census blocks representing their voting districts on copies of census maps. We will enter the resulting boundaries into the TIGER File. This will permit the tabulation of 1990 census counts by voting district, as required by P.L. 94-171.

The Geography Division Headquarters and several Census Bureau regional offices will enter the 1990 statistical areas into the TIGER File. Again, priority will be given to those statistical areas required for the data collection phase of the census: census tracts, block numbering areas, and block groups. The Census Bureau will assign the 1990 block numbers within the framework of the census tracts, block numbering areas, and block groups using the computer to identify and number each block. The work to delineate and enter into the TIGER File other statistical areas that are used for data tabulation purposes only--census designated places, census county divisions, and urbanized areas--will be done later. We will automate the delineation of the 1990 urbanized areas to a much greater extent than for previous censuses because of the existence of the TIGER System.

The process to obtain the political boundaries within which we will tabulate the 1990 census is ongoing. Every year the Census Bureau surveys all counties and those incorporated places with at least 5,000 population in the Boundary and Annexation Survey. Localities report their current legal boundaries in this survey. The 1988 through 1990 surveys will include all incorporated places regardless of their population size. With the 1988 survey, we will begin phasing in maps produced from the TIGER File as the base maps for the Boundary and Annexation Survey. The operation to enter the "current" (as distinct from 1980) political boundaries into the digital map data base is planned for the Data Preparation Division in Jeffersonville, IN. In the section below, where we describe the use of the TIGER File, we will discuss how the 1990 political boundaries (those effective on January 1, 1990, reported in January and February of 1990, and used for tabulating the results of the 1990 census) will be built into the TIGER File and used to identify blocks that require additional field checking.

A final major file building activity is the incorporation of address range information into the TIGER File. This information allows for the assignment of individual census responses to the correct geographic location in a mail census where, in most cases, enumerators at no point in the process visit housing units and so cannot determine their physical location. (The Census Bureau employed this methodology to count a large proportion of the population in both the 1970 and 1980 censuses and can use this method where mail is delivered using street name and house number addresses and a suitable automated address list either exists or can be compiled.)

As mentioned above, the address reference files that served this function for the 1980 census were called the GBF/DIME-Files; these files exist for 345 metropolitan areas. For selected GBF/DIME-Files, regional office staff will obtain reference materials to improve and update the address range information in the files. We will convert the GBF/DIME-Files to the TIGER File structure, merge these GBF/DIME-Files, now in TIGER File format, into the digital map data base to complete its coverage, and then update the TIGER File with this address information.

Use of the TIGER System to Support the 1990 Decennial Census

The charge to, and goal of, the Geography Division is to complete the development of the TIGER System for the entire United States and its territories in time to support the 1990 decennial census. The geographic products required earliest in the census cycle are those to support field operations to compile an address list for areas where the census will be taken by mail but no computerized address list already exists. (The Census Bureau hopes to expand the area within which the census is conducted by mail to lower costs and improve data quality--the TIGER System will help make this possible.)

We will prepare two products to support this operation: computer-generated maps for field and office use plus a computerized file depicting the streets bounding each 1990 census block for better control of the data collection and address list keying operations.

Where 1990 census field operations find new and better information, we will update the TIGER File to reflect these changes. We thereby can produce more up-to-date maps and other geographic materials for succeeding operations. Our map production and TIGER File update operations will be decentralized to our regional census centers.

For areas where the Census Bureau obtains a

computerized address list to mail the questionnaire, we will use the TIGER System to assign housing unit addresses automatically to the geographic areas used for data collection. We are improving our already good computer software for performing this function so as to minimize the number of addresses that need clerical research and resolution. Where there are unmatched addresses, we will use the results of the field checks to update the TIGER File.

Once we have assigned all housing units on the computerized address list to a geographic location, we will use the TIGER File to structure the list geographically so that field enumerators can, in advance of the census, completely canvass the areas covered by the list and update it as necessary. We also will use the TIGER File and computer-driven map plotters to prepare maps to guide the enumerators. As with other field operations, when this operation identifies additional changes in the geographic pattern, we will update the TIGER File.

Plans call for the Census Bureau to prepare precensus counts of addresses by census block and ask functioning governmental units to review the counts to identify address list deficiencies in their areas. We will use the computer to prepare maps for use by the localities in this precensus local review. The TIGER File will, by this time, reflect the most current political boundary information available so that we can automate the complicated logistics of sending maps covering the appropriate area to each governmental unit.

We will produce maps and a control list of geographic areas for use by Census Bureau field staff who will conduct a door-to-door enumeration in areas where no precensus mailing list exists or will be compiled. In addition, we will supply maps for nonresponse and failed-edit follow-up operations that are carried out after census day in mail census areas.

We expect to assign the addresses on various administrative lists that the Census Bureau obtains from other sources to geographic locations for use in the various coverage improvement and coverage measurement activities; some of these programs will require us to provide maps for field use as well.

The results of the 1990 Boundary and Annexation Survey, in which we collect the legal boundaries according to which the Census Bureau will tabulate the 1990 census data, will come in between January and March 1990. Because we could not use these boundaries for data collection but must use them for tabulation, we will enter them into the TIGER System immediately upon receipt. The automated system will determine where these new boundaries divide blocks, the lowest level geographic area for which the Census Bureau collects data. For these blocks, field staff will determine which addresses fall on each side of the boundary in question. This will be necessary only for boundary changes reported for the first time in the 1990 Boundary and Annexation Survey; we will enter the results of address location checks for earlier surveys into the TIGER File before the census.

Because the TIGER File defines the interrelationships of all the geographic areas it includes, these field operations are necessary only where a new boundary divides a census block. Once the results of these split block field checks are incorporated into the TIGER File, we can use the information to control the tabulation of the data according to the 1990 political boundaries. In 1980, this conversion from collection geography to tabulation geography involved a much more cumbersome operation, the clerical sorting out of the actual census questionnaires and extensive hand-revision to the census address registers by field staff using hand-prepared "split" maps.

Completion of this conversion allows the tabulation of data by governmental units as they are defined legally on the census reference date, January 1, 1990. Plans call for the resulting preliminary counts of both population and housing to be sent, along with updated maps, to all functioning governments for a postcensus local review of the counts. Like all other maps for the census, we will plot these postcensus local review maps from the TIGER File on high speed computer-driven map plotting devices. We will incorporate into the TIGER File any residual boundary corrections brought to light by the postcensus local review.

We will use the TIGER System during the post-data collection period to assign the place-of-work and migration responses on the individual questionnaires to their correct geographic location so that these data can be tabulated for transportation planning and analysis.

Another use to which we will put the TIGER System during this period will be the delineation of the 1990 urbanized areas based upon the 1990 census counts. While this job always will be one demanding the judgement of professional geographers, we will automate the job to a large extent to save time and money. A computer algorithm first will use the TIGER File plus the census counts to delineate a preliminary urbanized area boundary. Geographers then will review and interactively adjust the urbanized area boundary.

We will use the computer to generate maps to accompany the census counts provided to the states for redistricting pursuant to P.L. 94-171.

Finally, we will use the TIGER System during this same time period to computer-generate the maps the Census Bureau will publish along with the various 1990 data products.

Testing the TIGER System--Address List Compilation Test

Just as we have a charge to complete the development of the TIGER System for the entire United States in time to support the 1990 decennial census, so too do we have a charge to test this new system in conjunction with the various test censuses. Beginning in 1984, the Census Bureau conducted the Address List Compilation Test (ALCT) to compare alternative methods for preparing and/or updating address lists in both urban and rural areas. A major goal, in which we believe we succeeded, was to prepare geographic materials that would not bias the various test results in any way. We prepared maps that in several ways simulated those that we planned to produce using the TIGER System. For the urban test sites, Hartford and Bridgeport, CT, we made

maps for field use that depicted the cities on single rather than multiple map sheets. Although these maps were not computer plotted, they simulated the TIGER System's capability of efficiently mapping a given area on the fewest possible map sheets, thus eliminating the need to "piece together" parts of map sheets as was often necessary with our traditionally drafted 1980 census map products.

In the three rural ALCT sites (Gordon and Murray Counties, GA and Hardin County, TX), we completely redrafted the 1980 map products, incorporating updates by regional office staff (an operation that was not done for the 1980 census) and provided a set of maps that were the most legible ever supplied for rural area field operations. We scribed the base maps at a large scale and eliminated the many inserts generally found in the margin of the 1980 census maps for rural areas. Where multiple map sheets or insets were required, we used data collection area boundaries as match lines to facilitate map usage. The results confirm that the geographic products, prepared in a way that tailored the product to the need of the particular activity taking place in the field, were a significant improvement over the products used for 1980.

In addition to preparing new large-scale maps for field operations, we also digitized the maps in order to generate small-scale plots of each county without insets. Field staff used these maps in office operations such as recruiting and assignment planning. This capability to quickly generate maps at varying scales and formats once more demonstrates the flexibility that we will achieve when the TIGER System is fully operational.

We worked very closely with the key staff in all participating divisions to prepare the automated system for the assignment of housing unit addresses to geographic locations for the two urban sites. As with the map preparation activities for the ALCT, we worked to ensure that the system would handle the special kinds of requirements that existed in the two urban sites. In the process we further developed the specifications needed to improve the system in the future; an example is the handling of addresses that appear in the form of building names or special area identifications; these situations present difficult location assignment problems in urban areas.

In addition, we went to great lengths to eliminate the major types of problems in the 1980 geographic products, namely, inconsistencies among the geographic products.

Testing the TIGER System--1985 Test Census

For the 1985 Censuses in Tampa, FL and Jersey City, NJ, we again worked closely with the planners and designers to ensure that the geographic support materials enhanced the Census Bureau's ability to meet the specific objectives of the test census and that the geographic products would not contribute in any way to biasing the results of those tests.

Continuing the practice started in the 1984 ALCT, we custom designed the map products to ensure that they were centered on the test areas. This ensured minimum disruption of field and office activities by eliminating the need to deal with multiple map sheets. Although the Tampa test site was larger geographically than the urban sites in the 1984 ALCT, we were able to keep the number of map sheets for Tampa at three, with only one map sheet being required for Jersey City. Again, boundaries of interest were used as match lines where multiple map sheets were required.

We provided maps for use by enumerators for the Jersey City site at different scales according to the particular field operation for which the maps were intended. We worked with the responsible staff to gain a greater understanding of their requirements in this area and were able to provide a more useful product as a result.

In terms of the automated assignment of housing unit addresses to geographic locations, we made further progress in the development of the approach for the 1990 census. For the 1980 census, approximately 87 percent of the housing unit addresses for mail areas were assigned to geographic locations without last-minute field resolution; our goal for the 1985 test census was to improve that percentage to 98 percent or better. We improved the system to the point that we exceeded the goal and assigned 99.9 percent of the addresses to geographic locations. We seek to apply what we learn about the development and use of geographic products for the test censuses to develop the best possible TIGER System. An example of how we are maximizing this learning opportunity is our Reference File Enhancement Study, conducted in the Tampa test site. We acted upon the results of this study and added a reference file enhancement phase to our geographic location assignment work for the 1986 Test Census. Further, we conducted a survey of all users of our test census maps to ensure that we fully evaluated these products.

Testing the TIGER System--1986 Test Census

For the 1986 Censuses in East Central Mississippi and Central Los Angeles County, we took additional steps on the road to implementing the TIGER System. We acted on our commitment to use our new computer-driven map plotters to prepare most of the maps required for field operations. With automated map products for the 1986 Censuses, we demonstrated different methodologies for symbolizing the maps that would not have been possible in the more clerically based map production processes of the past. We also demonstrated the flexibility for producing map products at varying scales and in varying formats to meet the specific needs of different field and office operations. The types of format differences we tested included depicting roads with single lines where the accompanying names sit atop the line; showing the roads with double lines where the street names would appear inside the road symbol on the map; and using different sizes and styles of type for the feature names, the block numbers, and the other geographic identifiers shown on the map. From these various products we have gained experience and reactions from the Census Bureau's field staff in terms of which approaches work best. The results were used in refining the appearance of the maps in the 1987 Test Census

and the 1988 Dress Rehearsal, and ultimately will be reflected in the maps for the 1990 census.

We continued to refine and improve the automated processes for assigning addresses to geographic locations, to handle additional nonstandard address situations, and to experiment with approaches for handling address lists that cover larger areas. One of our major goals is to resolve the nagging problems associated with trying to associate mailing addresses of structures with their actual physical locations and to better understand the relationships between postal geography and census geography. Specifically, we experimented with the use of various types of postal geography directories that relate addresses on the mailing list to the specific portions of the geographic files for a particular area. In a national address list, for example, before we start looking at where an address fits along a street, we need to determine what post office that address is in (for example, Washington, DC or Alexandria, VA) to reach the proper section of the address reference file and begin our detailed search.

In the 1986 Censuses, we tried out approaches for linking the address control file--the geographically structured list of living quarters that controls the census field operations--to the geographic support system.

Further, we tested ways to produce maps for field researching addresses in blocks newly split by tabulation area boundaries, such as city limits, and methods for incorporating this information into the geographic support system.

Finally we tested ways to update the geographic support system with information brought to light by the test census field operations.

Testing the TIGER System--1987 Test Census

For the 1987 Census of North Central North Dakota, as well as in the 1987 Special Test Program, we have continued our testing activities. The scope of the 1987 Census is relatively narrow, however. It does not, for example, afford us an opportunity to test address location assignment applications. We produced all maps for field operations and for use by local officials in North Dakota using automated approaches and built in new alternatives gained from the experiences in the 1986 Censuses. As part of the 1987 Special Test Program, we performed a load test of our electrostatic map plotter.

Similarly, we expanded our activities to test further the Census Bureau's ability to identify the location, vis-a-vis a boundary, of the specific housing units located in blocks newly split by tabulation area boundaries.

We have, in the 1987 Census, continued testing to make refinements to our system for updating the geographic support system with changes discovered during test census field operations. In addition we made boundary updates to the TIGER File by automated means.

Demonstrating the TIGER System--1988 Dress Rehearsal

To a large extent we are demonstrating the fully developed TIGER System in the 1988 Dress Rehearsal; we still are working, however, to build in additional refinements and improvements for the 1990 decennial census.

We again computer plotted maps for field operations and use by local officials using the TIGER File, as we will do for all future Dress Rehearsal operations. In addition we computer plotted maps for use by Metropolitan Planning Organizations (MPOs) in supplying location information for major employers to assist us in our place of work coding.

We used the TIGER File to carry out the assignment of addresses on the vendor and special place lists to their correct geographic locations.

Further, we have updated the TIGER File for the Dress Rehearsal not only with information added to the census maps by enumerators, but also with boundary updates from the Boundary and Annexation Survey and corrections identified in our clerical geocoding resolution operations.

Summary

The TIGER System will improve the 1990 census by eliminating the inconsistencies among geographic support products that caused delays and confusion in the collection, processing, and use of the 1980 census data. The TIGER System will also allow us to provide our enumeration and processing staffs, not to mention the data user, with higher quality census maps. These benefits will also accrue to the Census Bureau's many other censuses and surveys during the 1990s and beyond.

[1]This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.