

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Working Paper No. 14
English only

Work Session on Geographical Information Systems
(Brighton, United Kingdom, 22-25 September 1997)

Item (iii) of the provisional agenda

**GIS DEVELOPMENTS AT THE CENSUS BUREAU:
MAINTAINING THE TIGER SYSTEM**

by

Joel Sobel, U.S. Bureau of the Census

I. INTRODUCTION

1. At the 1996 ECE Work Session on GIS we reported on the TIGER System's new directions this decade. Most notably, the Census Bureau (CB) is now:
 - a) creating and maintaining a nationwide file of dwelling units, using the 1990 census address file and the U.S. Postal Service (USPS) file of individual mail delivery points;
 - b) matching this file to the TIGER data base to determine address locations down to the census block level; and
 - c) updating the streets, street names, and address ranges in the TIGER data base by identifying address ranges recognized by the USPS that cannot be matched to TIGER and resolving them accordingly. Furthermore, the Geography Division (GEO) in the past 18 months has planned or implemented significant changes to the structure of the data base, the hardware and software used to maintain it, and the methods for correcting or adding data.

II. RESTRUCTURING THE TIGER DATA BASE

2. The GEO is making several changes to the TIGER data base to reflect customers' changing needs as Census 2000 approaches. One of these changes affects the way national parks and military reservations appear in TIGER. Previously, these items were treated as "area landmarks." This meant that they could appear on maps produced from the TIGER data base, but could not be identified and processed as if they were geographic "entities" such as counties, cities, census tracts, and census designated places. In 1997, the data base structure will change, and national parks and military reservations will be recognized in the TIGER data base as true geographic entities. Once the change is made, the CB will be able to both perform batch processes on these entities and prepare maps and files of them without specifying each of the individual blocks that comprise these entities.

3. Many other changes will be more noticeable to customers outside the GEO. The CB will likely recognize new geographic entities for Census 2000, including settlement clusters and ZIP (postal) code tabulation areas. Just as important, we will add Census 2000 tabulation areas to the data base structure, while retaining the 1990 census tabulation areas. This will enable data users to compare geographic area and demographic changes from 1990 to 2000 in a GIS environment with a single geographic data base.

III. TIGER DATA BASE MAINTENANCE

4. For the 1990 census, the GEO developed up-to-date software known as GUS, written mostly in FORTRAN, and for use exclusively with Tektronix 4335 graphics workstations, which were purchased in 1987. The TIGER data base and GUS software resided on a single computer cluster in the CB's Charlotte (North Carolina) Computer facility. GUS operators at CB headquarters in Washington, in the Data Preparation Division in Jeffersonville (Indiana), and in 12 census regional offices nationwide, updated the TIGER data base by "remotely" accessing the desired TIGER partition (usually a single county) or partitions.

5. Since then, the GEO has adopted an "open systems" policy in which functions are less hardware dependent. Changes were implemented beginning in 1995, and continue today. The Tektronix workstations have been replaced with Silicon Graphics, Inc. Indigo work stations and Intel-based personal computers, all of which function as "stand-alone" UNIX-based machines. Instead of accessing remote TIGER partitions, operators now transfer the desired TIGER data

base partition to their work stations, and use newly-developed GUS-X software, written for the most part in ANSI C, to make the necessary updates to TIGER.

6. The CB has benefited in several ways from these changes. First, by decentralizing files during update operations, work proceeds more efficiently; updates no longer are delayed nationwide by problems at a centralized site. Second, the switch to the open UNIX environment and away from a hardware-dependent system allows more flexibility and will enable the CB to take greater advantage of improved hardware in the coming the years without needing to completely revise the GUS-X software.

IV. EXCHANGE OF MACHINE-READABLE GEOGRAPHIC BASE FILES

7. Since the early 1980s, the GEO and other divisions in the CB have worked to develop software that would enable the automated updating of the TIGER data base by matching it to a locally-maintained digital geographic base file. In 1996, the GEO successfully began this procedure for automated processing, with rather strict limitations to minimize the introduction of erroneous information and the need for intensive manual intervention.

8. As part of this program, the GEO will:

- a) add ONLY street features, and NOT hydrographic features, railroads, and other non-street features, to the TIGER data base;
- b) accept attributes such as names, address ranges, postal codes, and geographic coordinates ONLY for added street records OR for existing street records that do not contain these attributes, and NOT as replacements for data that the TIGER data base already contains (in other words, the GEO will not attempt to improve the accuracy of existing geographic coordinates or address ranges in TIGER).

9. The purpose of these updates is to allow the GEO to assign census block codes to individual dwelling unit mailing addresses through an automated match with the TIGER data base (a process called "geocoding"). As of the end of May, 1997, the GEO has used 27 locally-maintained digital files, and saved an estimated 55 person months of clerical geocoding.

V. "RURAL" DWELLING UNITS

10. Enumerators in previous censuses who were responsible for "listing" dwelling units in a census register book placed "map spots" on several hundred thousand census maps to identify the approximate location of each unit. The CB subsequently used these same maps, whose condition often deteriorated significantly over the course of the listing operation, for later census questionnaire delivery and/or follow-up operations. Although the CB eventually updated the TIGER data base to reflect new and corrected features indicated by these enumerators, the map spots never were entered into TIGER.

11. As part of the 1994 Census Test in several parishes in northwest Louisiana, and the 1996 American Community Survey test in Otero County, New Mexico, the CB is testing the feasibility of adding these map spot coordinates and numeric identifiers to the TIGER data base. Similar tests will be conducted in affected areas of South Carolina and Wisconsin as part of the dress rehearsal for Census 2000. By including these map spots in the TIGER data base, the CB will be able to produce new follow-up enumerator maps that include map spot locations. (Note: Although the TIGER data base will contain map spot data as a result of these operations, these data are confidential, and, as such, will not appear in the TIGER/Line_ files or census maps that are available to the public.)

V. DATA CAPTURE

12. As Census 2000 approaches, the CB is experimenting with new techniques to handle the large volume of TIGER data base updates.

13. Currently, the CB is testing a method for "scanning" Census maps that contain enumerator updates, especially the map spot locations. As part of the test, these maps are scanned by Kodak 990 Scanners, and an output file is produced. Digitizing staff then may digitize the map spot locations using the files of scanned data. The advantages of this procedure are: a) time-consuming map registration activities may be eliminated; b) map spot locations are collected in a flat ASCII file and then later loaded into the TIGER data base through batch processing; c) training time is reduced due to the "user friendly" nature of the software.

The CB will use similar scanning technologies to produce machine-readable files of the dwelling units that appear in the address registers that census field enumerators compile during their canvassing activities. These data include the map spot numbers that also are on the enumerator maps that the CB is scanning. The CB then will create a Master Address File (MAF) from the scanned dwelling unit files, link the MAF with the TIGER data base (which by that time includes the digitized map spot locations), and resolve any discrepancies.

VI. MAP IMAGE METAFILES (MIM)

14. The CB automated its mapping efforts for the 1990 census. As part of the TIGER System, the GEO supported census field operations by developing computer programs for producing hundreds of thousands of maps using a non-interactive computer batch process. All field enumeration maps thus were produced from digital plot files designed for specific large-format electrostatic plotters. One significant disadvantage of this method was that the map files were machine dependent (that is, the maps could be plotted only by using a specific output device). A second disadvantage was that maps could only be "viewed" through use of paper plots as opposed to on-screen viewing.

15. In the early 1990s, however, the GEO devised a new map file format to support publication mapping, known as the Map Image Metafile (MIM_), which is a simple sequential ASCII file that is "human-readable" and device independent. The open format allows for creation of simple conversion routines for any output device. The CB also created MIM_ viewing software for use on PCs, thereby extending automated mapping utilities. Functions such as map sheeting, scaling, and text placement are separate interchangeable modules that the GEO developed to support diverse map requirements.

16. For Census 2000, staff in the hundreds of decentralized census field offices will have the ability to produce their own maps by initiating standard mapping programs to meet their particular requirements. This reduces their dependence on centralized computer facilities, and allows them to determine map requirements and control production. Just as important, in these offices where usable space is vital, less room is needed to store paper maps, because on-demand plotting capabilities will be available.

References

Franz, Linda M., The Master Address File/TIGER System: Infrastructure for Innovative Data Collection, presented at the Conference of European Statisticians Work Session on Geographic Information Systems, April 1996.

Rosenson, Peter, Automated Updating of TIGER from Locally Developed Geospatial Data Bases (unpublished paper).

Torrieri, Nancy K., Revising the Census Bureau's Urban and Rural Definitions for the 2000 Census, presented at the Conference of European Statisticians Work Session on Geographic Information Systems, April 1996.

Acknowledgments

I would like to thank Linda Franz, Jack George, Catherine McCully and Brian Swanhart for reviewing portions of this report and providing information. I also am especially grateful to Timothy Trainor for preparing the initial text for section 6 of this report dealing with Map Image Metafiles.