AUTOMATING THE GEOGRAPHIC AND CARTOGRAPHIC ASPECTS OF THE 1990 DECENNIAL CENSUS

Robert W. Marx
Assistant Division Chief for Geographic Areas
Geography Division, U.S. Bureau of the Census
Washington, D.C. 20233

ABSTRACT

The mission of the Census Bureau is to provide basic statistics about the people and economy of our Nation to the Congress, the Executive Branch, and the general public. To accomplish this mission, geographic support must be provided to assign each housing unit and business establishment to the correct geographic location, and then classify that location according to the tabulation areas represented in the census or survey. Recording all relevant geographic information about an area in a single computer file will permit the assignment of geographic location codes to households and businesses, provide the ability to generate maps showing the geographic information for field operations and subsequent publication, and provide the geographic structure needed for tabulation of the data to every area whose boundaries have been recorded in the file. The Geography Division of the Census Bureau has developed a long-term plan for building and maintaining this computer file.

INTRODUCTION

The mission of the Census Bureau is to provide basic statistics about the people and economy of our Nation to the Congress, the Executive Branch, and the general public. The successful collection and tabulation of these data are dependent on the availability of a variety of geographic materials and services. These geographic materials and services are all geared toward doing two things: First, assigning housing units and business establishments to a geographic location--a block, and then classifying that block into the full range of geographic areas recognized in the data tabulations of that census or survey.

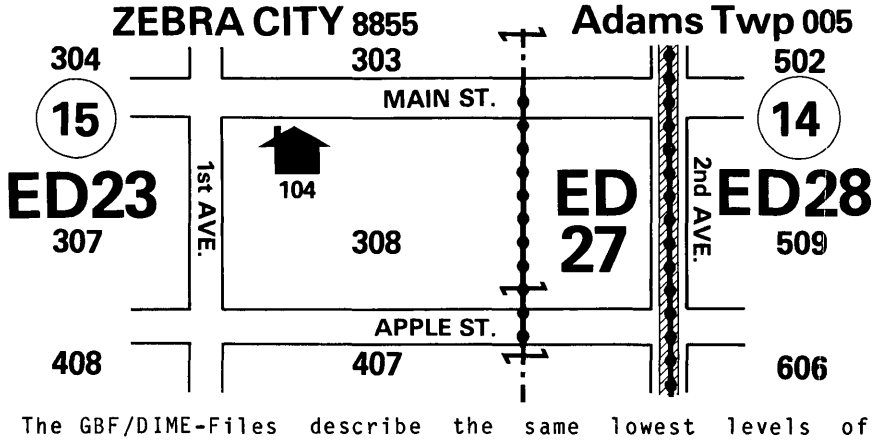
You have heard in the first two papers of this session about the three major tools prepared by the Geography Division to assist the field data collection staff in the completion of their task and the subsequent use of these tools to assist in the tabulation and understanding of these data. You also may have heard from some sources that there were problems with these geographic materials in 1980; problems that caused confusion on the part of the field staff and the data using public.

THE PROBLEM

Understanding what caused the problems with the geographic products in 1980 turned out to be the first major step in planning for the geographic support of the 1990 census. If you think about what you have heard described in the previous two papers, you may recognize that maps, GBF/DIME-Files, and the Master Reference File all have several items in common; that is, they are simply three different ways of describing the surface of the earth.

Maps describe the earth in graphic form showing both the lowest levels--streets, railroads, streams, and the polygons, which we call blocks, delimited by those features. Maps also show the higher level geographic units into which we classify blocks--census tracts, cities, townships, counties, urbanized areas, and so forth. (See Figure 1.) Using this map, a census enumerator can walk around a block, list the addresses seen along each side of the block in a book called an address register, and at the same time record the number of the block containing that address. For example, that 104 Main Street is in block 308. In that simple act of writing down the block number, the enumerator has "geocoded" the address; that is, assigned it to a geographic location. This is how censuses have been taken for the 200 years of America's history.

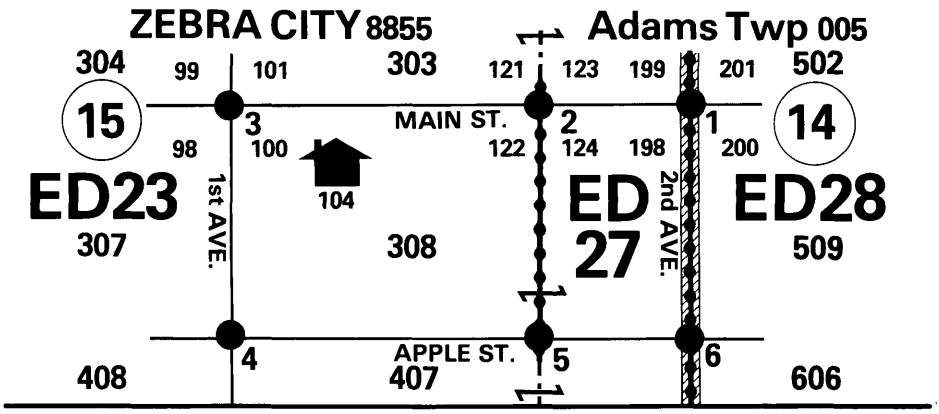
Figure I



The GBF/DIME-Files describe the same lowest levels of geography--streets, railroads, streams, blocks, and census tracts--in a manner that is understandable by a computer. (See Figure 2.) The GBF/DIME-Files add two critical pieces of information to the computer's knowledge of the map.

° First, they record the address ranges that apply to each street segment. This additional information allows the computer to "see" what addresses fit into each block using computer matching algorithms which perform the geocoding function previously done by an enumerator. For example, 104 Main Street "fits" in the range 100-122 on Main Street and is, therefore, in block 308.

Figure 2



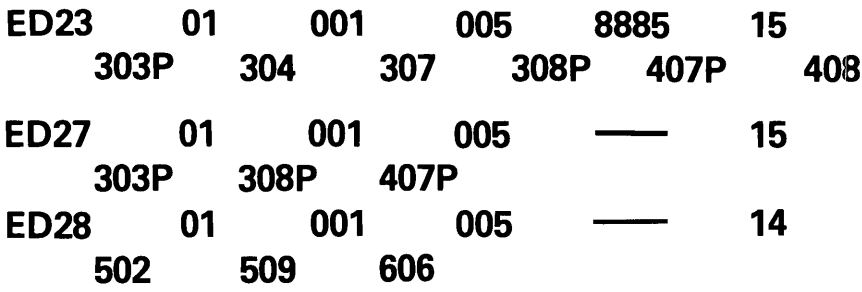
Street Name	From	To	Place	Tract	Left Block	Right Block
MAIN ST.	3	2	8855	15	303 101-121	308 100-122
MAIN ST.	2	1	——	15	303 123-199	308 124-198

° Second, they record the geometric description of the map in the form of latitude and longitude values for each intersection (and some intermediate points) shown on the map. These points are called "nodes" and have been assigned numbers for use in linking the coordinate values read from the maps to the computer files. The GBF/DIME-Files can geocode addresses without the coordinates, but they can't drive a plotter unless this information is present.

The Master Reference File describes the same map in a third way--by recording in computer-readable form the relationships of the lowest level geographic area, the block, to all higher level geographic areas for which data will be tabulated. (See Figure 3.) This is done enumeration district-by-enumeration district. For example, this file tells you that ED 23 contains block 308 which is in census tract 15, Zebra City (code 8855), Adams Township (code 005), county 001, state 01, and so forth.

As you can see from this example, we have a problem in the Master Reference File: The place code for Zebra City was written as 8885 instead of 8855. This error in the Master Reference File now causes this file to "mismatch" with the GBF/DIME-File and the maps. In 1980, this type of mismatch caused a cascade of problems in all subsequent geographic products related to Zebra City and resulted in much of the discontent expressed by our field staff and data users.

Figure 3



Now, that we have seen what the products were for 1980 and what kind of problem existed, let's look at the cause of our problems. To create the three products described we used people. One group of people (about 900 of them) drew the maps, named all of the streets they could, plotted all of the boundaries, and rubbed down by hand all of the block numbers-2 1/2 million of them--to make the 32,000 separate map sheets that covered the United States for the 1980 Decennial Census. This was all done initially during the 2 year period before the census with cleanup activities continuing using a greatly reduced staff for 2 years after the census.

At a slightly earlier period in time, over 300 local agencies scattered across the country were busy creating GBF/DIME-Files by transcribing the street names shown on earlier versions of many of these same maps, all of the block numbers within the areas covered by the files, and the address ranges that went with those street names and block numbers. All of this information subsequently was keyed and converted into a series of computer files.

At a slightly later period in time, another group of people (about 300 of them) took these same maps and transcribed the same block numbers, along with the ED numbers, census tract numbers, place names and codes, county names and codes, and so forth in which the blocks were located. This information also was keyed and converted into a series of computer files, state-by-state.

Said another way, three different times, with three different groups of people, we recorded the essential information about the earth in census terms. We then matched all three of these products together and were confronted with the fact that they didn't match perfectly. For example, in Figure 3, the place code in ED 23 was written as 8885 instead of 8855, which caused a mismatch between the Master Reference File and the maps and GBF/ DIME-Files.

Perhaps we should not have been surprised because people do make mistakes as they perform repetitive clerical tasks. Also, it is not reasonable to assume that mistakes will be made on the same area by different people--they each make mistakes in their own way. It is the process, not the people, which is the cause of the problems; this series of complex and functionally separate operations which were used to create the 1980 geographic products... and the 1970 geographic products... and the 1960 geographic products.

THE PLAN

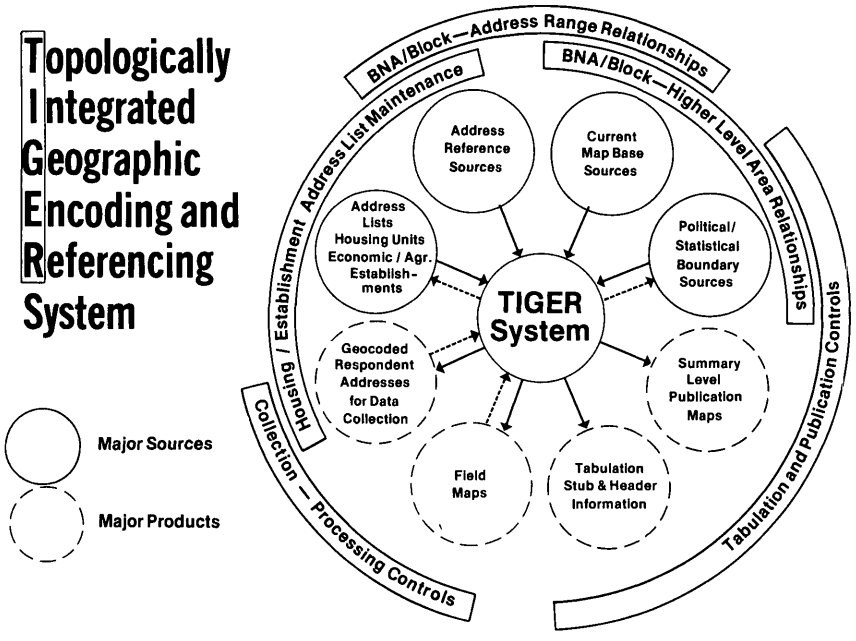
The expanding role of censuses and surveys in meeting the Nation's data requirements necessitates a substantial improvement in the delivery of these critical products and services. To this end, the Bureau has enunciated a series of long-term goals, one of which is:

> "To improve the efficiency of Census Bureau activities by inaugurating an automated system of geographic support services."

The improved geographic support program designed to meet this goal in time to benefit the 1990 Decennial Census involves the recording of all relevant available geographic information about an area in a single computer file. This file, once completed for an area, will permit the assignment of geographic classification codes to individual residential and business addresses, allow for the generation of maps for field operations and publications, and provide the geographic framework for tabulation of data to any unit whose boundaries have been recorded in the file. Further, because all products will be produced from one file, the possibility no longer exists for the maps to show one block number, the GBF/DIME-Files to show a second block number, and the geographic reference files to have a third block number. The same block number will appear in all three products.

Based on the assimilation of advice and suggestions from a broad spectrum of academic disciplines and acknowledged experts in several fields, a preliminary design for the structure of the desired computer file has been developed. The file structure adapts the mathematical theories of topology to cartographic presentation and geographic relationship storage and retrieval problems. For discussion purposes, this file is called the Topologically Integrated Geographic Encoding and Referencing File, or TIGER file. (See Figure 4.)

Figure 4



To convert the functionally separate, clerically intensive geographic support operations of the past into an automated system capable of delivering high quality, timely geographic products and services for 1990 is an enormous undertaking. It will require a "one-time" major allocation of resources to make a fundamental change in the way geographic support is provided to the Bureau; to provide for the transition from the methodologies of the past to the approaches made possible through automation. This fundamental change in approach will enable us to deliver not only the previously required types of geographic support services, it will provide the Bureau with the capability to automate many other aspects of the data collection/processing/publication continuum--especially, those which are geographically based.

The geographic support program described in this paper includes both activities which will be automated and those related activities which will continue to use traditional processes. For example, the collection and preparation of the geographic information which will be entered into the computer file (current base maps, address range listings, geographic area boundaries, and so forth) will be done in traditional ways; the entry of that information into the computer file, the editing of the computer file for completeness and consistency, and the production of maps and other geographic materials will make use of automated processes that will greatly increase the efficiency and quality of the services and products provided.

The computer file that will be at the heart of the automated geographic support program will integrate four major functional components:

1. The capability to document the nominal and hierarchical relationships of the geographic areas represented in each census or survey: the 50 states, the 3,137 counties or county equivalents, the 18,000 plus minor civil divisions or equivalent areas, the more than 20,000 incorporated places, and so forth. Each area's name and its identifying code(s) will be recorded. The area's geographic relationship to all other areas in the file (the Master Reference File of 1980) will be derivable automatically.

2. The capability to assign household or business establishment responses to the appropriate geographic location based on the structure address. This is the process referred to as "geocoding" that was done by the 1980 GBF/DIME-Files.

3. The capability to accurately portray the boundaries of all political and statistical areas recorded in the geographic component of the TIGER file plus the streets, roads, and other relevant map features needed for field operations as recorded in the geographic assignment component of the TIGER file. This component must provide a mechanism for update of the features and boundaries comprising the file to reflect changes. It will then be possible to produce the maps using automated methodologies.

4. The capability to provide a geographic control system that meets the operational requirements of the field data collection activities, the processing center activivities, and the data tabulation/dissemination functions. For example, the production of maps for use in followup assignments, the printing of lists showing blocks in relation to various higher-level geographic units, the preparation of management reports showing the status of various geographic operations, and so forth.

Each record in the file will contain information, such as:

- The name and type of the feature; for example, whether the feature is a road or street, a waterway, a railroad, a political boundary, and so forth.

- The coordinate values defining intersection points along the feature, along with other geometric characteristics of the feature; for example, the curve vectors defining the shape of the feature.

- The range of addresses located between intersection points for those records representing streets or roads, in addition to the post office name and ZIP code associated with that address range.

- The codes for the geographic area(s) applicable to each segment of the feature based on the geometric relationship of the boundaries in the file to the feature.

- Other special situations associated with the record; for example, major employment centers or residential structures located along the feature.

Once constructed, the TIGER file provides a geographic framework from which the maps required for field operations or publication products could be generated; a means through which the housing units or business establishments in a census or sample survey could be assigned to a specific geographic location; a basis upon which an automated questionnaire check-in and control system could be established, permitting generation of follow-up assignments based on structure address or serial number in geographic perspective; and, a source file from which geographic table stubs and summary cartographic products could be generated for tabulation purposes.

While this plan is very ambitious, it is "do-able" by 1990 if we organize our resources properly and take advantage of the work done by others. We look forward to discussing this plan with you in the days and months ahead.