

GEOGRAPHIC AREAS AND COMPUTER FILES FROM THE
1980 DECENNIAL CENSUS

Robert A. LaMacchia
Geography Division
United States Bureau of the Census
Washington, D.C. 20233

ABSTRACT

For the 1980 Decennial Census, the United States Bureau of the Census produced two major types of computer files to document the geographic information related to the map sheets used for the census. For areas where GBF/DIME-Files were prepared, these files document the name of every street and other physical feature on the map, along with the address ranges and block numbers pertaining to both sides of each street segment. These files also have latitude and longitude values assigned to each street intersection. The Master Reference File documents the interrelationship of census political and statistical geography, such as census tracts, places, townships, counties, urbanized areas, and so forth. Work is underway to add latitude and longitude values to the centroid location of each major tabulation area. In addition, the Bureau has produced a digital file documenting the boundaries of every county and county equivalent in the United States as of January 1, 1980.

INTRODUCTION

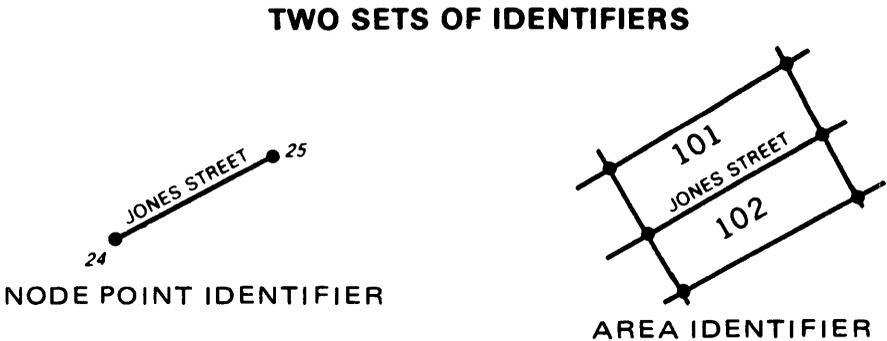
For the 1980 Census of Population and Housing, two computer files were developed, which enable a user to assign geographic location codes to an address. These files are the GBF/DIME-File and the Master Reference File (MRF).

These files were not developed for the purpose of computer-assisted mapping; rather, they were developed to accomplish the task of assigning the correct census geography to a questionnaire enabling the Bureau to tabulate data for the required political and statistical units, such as county, city, or census tract. As a post-census operation, earth coordinates are being inserted into the GBF/DIME-Files, thereby enabling the user to produce maps in a computer-assisted fashion. A version of the MRF available to the public, called the 1980 Master Area Reference File (MARF), will be available as MARF II and have coordinate values for centroids of major tabulation areas. In addition to these two files and as a by-product of the county area measurement project for the census, a new county boundary file also is available and replaces the 1970 county boundary (DIMECO) file.

THE GBF/DIME-FILE

The GBF/DIME System consists of the GBF/DIME-File, a working set of the Census Bureau's Metropolitan Map Series known as the CUE maps, and a computer program to create and maintain the file. The GBF/DIME-File itself is a computerized representation of the CUE map; within the area covered by the file, it contains all features shown on the CUE map plus potential address ranges and ZIP codes for all streets. The file structure is based on graph theory. Each feature on the map is considered a series of lines, and each feature intersection is considered a point (node). An entire map sheet is viewed as a series of interrelated nodes and segments; when connected they create enclosed areas (blocks). This approach is referred to as DIME (Dual Independent Map Encoding). DIME refers to the fact that the computer file which represents the map is created from two independent sets of identifiers for each segment: (1) the node points for each intersection, and (2) the enclosed areas separated by each segment. In Figure 1, a segment of Jones Street may be identified either as that segment between nodes 24 and 25, or as the segment which separates blocks 101 and 102.

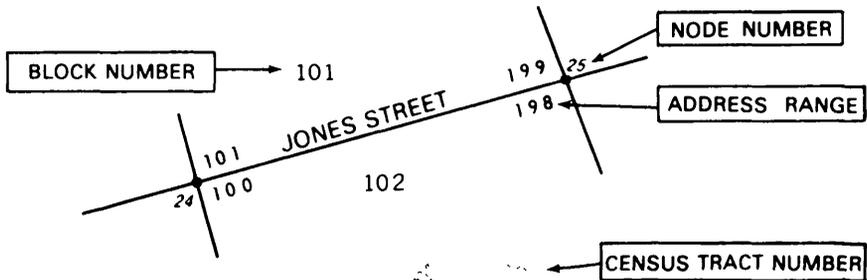
Figure 1. A line Segment with DIME Identifiers



Each computer record in the GBF/DIME-File identifies a single segment of a feature between two node points and includes all of the geographic information related to that segment. A feature on the map is divided into a series of segments (computer records) where it intersects with other features, or for the purpose of indicating curvature (curves are represented as a series of straight lines). Each address range is tied to the proper area (block). If the potential address ranges along a blockside are added to a traditional map, the same segment of Jones Street would appear as shown in Figure 2.

Figure 2. A DIME Segment with Address Information

TYPES OF INFORMATION



For each segment side, the 1980 GBF/DIME-Files contain the numeric identifiers for state, county, minor civil division, place, census tract and block, the potential address range, and ZIP code. The GBF/DIME-Files were developed to permit the automated assignment of geographic codes to an address without reference to a printed map. The geographic area covered by each file varies by SMSA, but generally the coverage includes the central city and immediate suburbs. These files exist for the 277 Standard Metropolitan Statistical Areas (SMSAs) established by the Office of Management and Budget by 1977 (except in Puerto Rico). Work is now proceeding to create similar files for SMSAs established as a result of the 1980 census. As the GBF/DIME-File is a computer representation of a map, the addition of coordinate values to each node point permits the computer to use the file to display the map. As a post-census operation, each node point is being digitized and the latitude-longitude value inserted into the file. State plane coordinate values are calculated from the latitude-longitude and also inserted into the file. During the digitizing operation, computer plots of each map sheet are reviewed. A completed plot of the Elmira, New York GBF/DIME-File appears in Figure 3. This plot is a mosaic of parts of five map sheets made from page-size copies of the graphics terminal display.

The structure of the GBF/DIME-File enables it to be used in a variety of computer-assisted mapping applications. As the next paper will indicate, the files can become the nucleus for an automated geographic system. The maps produced by the files are accurate only for use in showing feature relationships and not for engineering-type application as the computer can only reproduce a map of the quality that was put into computer-readable form. The Bureau has experimented with using GBF/DIME-Files to produce maps which may be used by a field enumerator in the 1990 census. Automating map production would give the Bureau the capability to produce a map at the proper scale for each enumerator assignment as well as for publication. An experimental map made from a GBF/DIME-File

Figure 3. Elmira, New York GBF/DIME-File

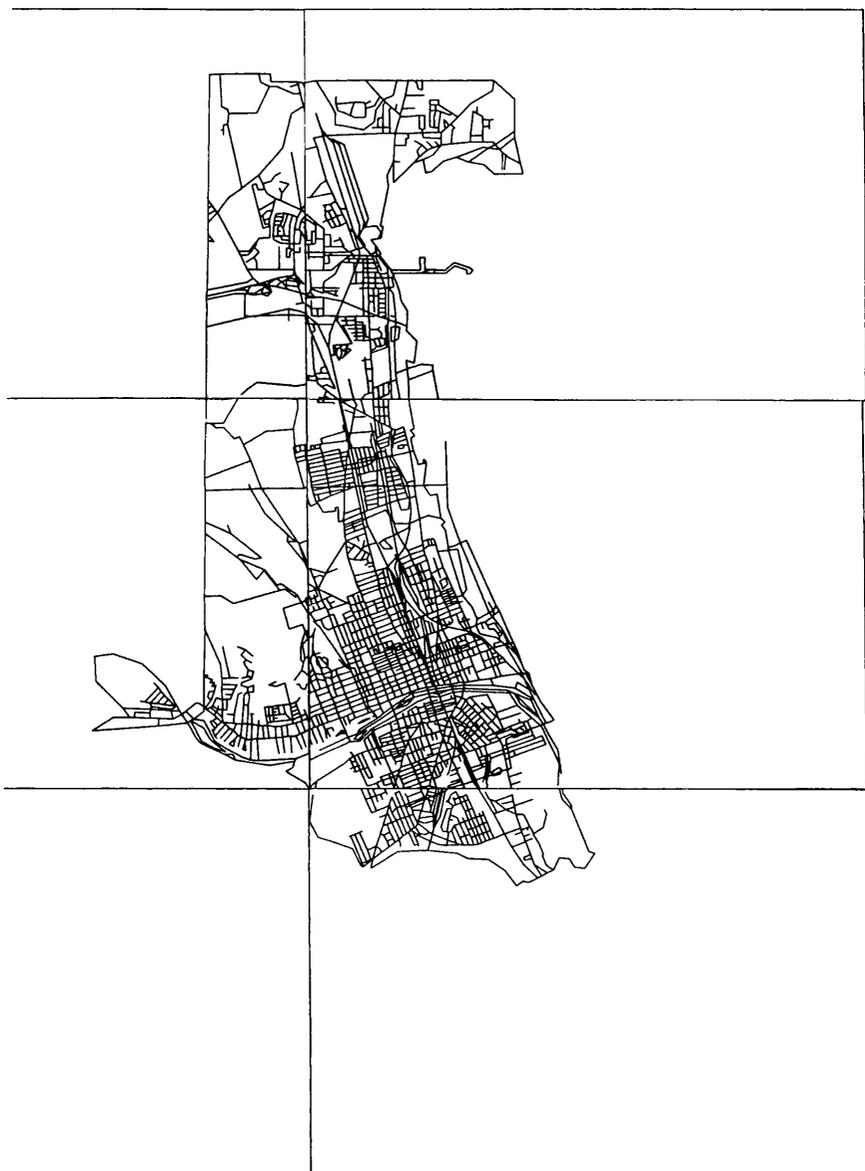
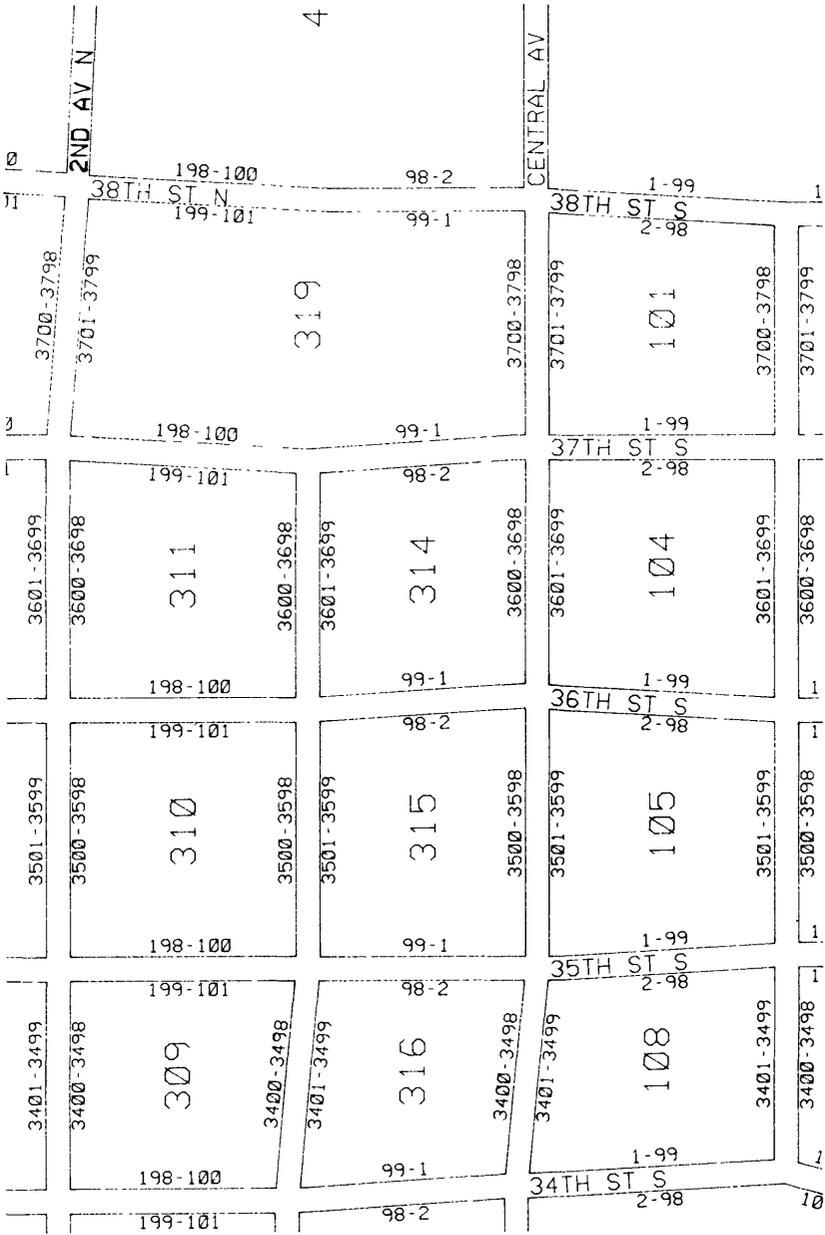


Figure 4. Experimental Field Map



is presented in Figure 4. In lieu of the simple single-line plot shown in Figure 3, the computer has turned each segment into a double-lined street, added feature names, and the potential address ranges. This work is still in the experimental phase, but it clearly shows that computer-drawn maps can simplify the cartographic requirements for 1990.

The fact that the GBF/DIME-File enables a user to assign geographic location codes to local data based on an address, such as police and fire calls, health visits, and so forth, also means that the file can be used to plot the approximate location of an address along a street or determine an approximate coordinate value for the address, assuming linear interpolation of the address range along the street. Combining an address list with the double-line map opens the door to additional use by the Bureau, when one considers that over 325,000 individual ED maps were prepared for the 1980 census. These ED maps were made from over 32,000 1980 Census Map sheets of the type discussed in the first paper, and often required photographic enlargement to meet the various enumerators' requirements. The amount of time and clerical involvement can be substantially reduced for areas covered by GBF/DIME-Files using computer-drawn maps of the type shown in Figure 4.

THE MASTER REFERENCE FILE

The Master Reference File (MRF) was created prior to the 1980 census and was corrected and revised throughout the taking and tabulation of the census. Its purpose is to link all state, county, minor civil division (MCD) or census county division (CCD), and place names with their numeric codes, to indicate the hierarchical relationships between these units and all other political and statistical units, such as census tract, census block, and urbanized area, and to identify the lowest level of geographical code combination which must be tabulated. This smallest geographic area that must be used in tabulation is called a Geographic Tabulation Unit Base (GTUB). Within each GTUB, one or more enumeration districts (EDs) are defined. Associated with each ED are the census blocks or block portions (in block numbered areas) covered by the ED. In the public files, EDs are replaced by block groups (where blocks exist) for tabulation purposes. Census tract (or block numbering area (BNA)) is an inseparable corollary to each census block. The combination of tract (BNA)/block relates to other levels of geography. In unblocked areas, EDs (without relation to census tract) served this function. Even in block number numbered areas, the ED was the basic unit for enumeration.

A section of an MRF might appear as shown in Table 1, where the "R" column indicates the geographic level. A code of all "9's" indicates that the geographic level contains more than one code for the specified type of activity, or has area territory inside and outside the geographic area.

TABLE 1

R	ST	CO	MCD	PL	TRACT	CD	UA	NAME
S	51					99	9999	VIRGINIA
C	51	001				01		ACCOMACK
M	51	001	005			01		ATLANTIC
P	51	001	005	0580		01		HALLWOOD
T	51	001	005	0580	9903	01		
G	51	001	005	0580	9903	01		
E	51	001	005	0580	9903	01		ED 56
								BL 106 107 108 109
								110 111 112 113
P	51	001	005	1100		01		SAXIS
T	51	001	005	1100	9903	01		
G	51	001	005	1100	9903	01		
E	51	001	005	1100	9903	01		ED 57
								BL 201 202 203 204
E	51	001	005	1100	9903	01		ED 58
								BL 301 302 303 304

From this example, geographic interrelationships are illustrated. For example, BNA 9903 appears in at least two places, Hallwood and Saxis. The 100 block group (BG) is in Hallwood and the 200 and 300 block groups are in Saxis. (A block group is that set of blocks within a census tract sharing the same first digit.) Both Hallwood and Saxis are in Atlantic MCD which is in Accomack County, Virginia. Other codes in the MRF would indicate that Atlantic is an MCD and not a CCD, and that Hallwood is an incorporated place while Saxis is a Census Designated Place (CDP). All political boundaries represented in the MRF are as of January 1, 1980 as reported to the Bureau.

The MRF was clerically constructed from the 1980 Census Maps. All of the required boundaries were plotted and the smallest unit required for tabulation (GTUB) was determined. EDs were defined within GTUBs, first on the maps and then translated to the MRF. A public version of the MRF is called the 1980 Master Area Reference File (MARF) and contains 1980 population and housing counts. It does not contain GTUBs, EDs within block numbered areas, or block numbers. In block numbered areas, it contains a record for each block group (BG). The 1970 equivalent of the MARF is known as the MEDList, a version of which (MEDList-X) contains coordinate values for the approximate population centroids of each ED. For 1980, a second version of the MARF (MARF II) will contain the approximate population centroid of each ED or BG. (Budget constraints had eliminated MARF II as a 1980 census product. Current indications are that MARF II will be reinstated.)

When MARF II becomes available, the user will have a nationwide tool which can be used to graphically display and analyze census data. Although MARF II will only give the user a single set of coordinate values for an ED or BG, that will be sufficient in most cases to make close approx-

imations of the population within a specified distance of a desired location. Combined with the digital county boundary file, which can be used to plot the county boundary, a user will be able to display census data available by ED/BG assuming the ED/BG extends half-way to the next centroid or to the county boundary. This is only one example of what users may do with MARF II provided they develop the necessary software.

CAVEATS

The GBF/DIME-Files were created or updated for the most part during 1976 through 1978, and at best represent 1977 vintage on the ground features, political and statistical geography, address information, and ZIP codes. Over 300 local agencies worked on this project and were responsible for content and geographic extent of the files. At the tract/block level, agreement between the CUE maps (and GBF/DIME-Files) and the 1980 census publication maps (as published in the block statistics reports) is about 98 percent for all GBF/DIME-File areas. For any single file, the rate of disagreement between the tract and block numbers could be as high as 7 percent. Reference has been made to CUE maps, which are not the same as the 1980 Census Maps or 1980 Metropolitan Map Series (MMS). CUE maps are copies of the working maps used by local agencies to create or update the GBF/DIME-Files and are the only set of maps that contain node numbers. As they were the basis for the update of the artwork used to produce the 1980 Census Maps, differences occurred because of the clerical nature of the work and census requirements. Because the GBF/DIME-Files predate the census, no geographic codes other than the tract and block numbers should be used when matching to census data. With the exception of the New York, NY SMSA, tract numbers are unique within an SMSA and can be used as a match key to the MARF to obtain any other geographic codes required. In the New York, NY SMSA, tract numbers are unique within county.

The MRF, which is also the basis for geographic codes in all of the 1980 census Summary Tape Files (STFs) available to the public, also was created through a clerical process and prone to clerical error. Users will sometimes find a block number (available in STF 1B and report PHC80-1) which does not appear on the 1980 Census Map, or vice versa. The number of these occurrences is very small nationwide, but the user attempting to match the MARF/STF, 1980 Census Map, and GBF/DIME-File must be aware that there may be some cases in which these three sources will be inconsistent.

The digital county boundary file is distributed in a new file structure, called the Geographically Encoded String Structure (GESS). The GESS is easy to convert to both polygons and segments and can be reformatted into a DIMECO type file.