A CARTOGRAPHIC EXTRACT OF THE TIGER FILE: IMPLICATIONS FOR MAPPING APPLICATIONS

Amy Bishton
Geography Division
U. S. Bureau of the Census
Washington, DC 20233

ABSTRACT

The Topologically Integrated Geographic Encoding and Referencing (TIGER) File is the Census Bureau's primary data base for map production. Collecting and chaining the TIGER File's topological elements of 0-cells (points), 1-cells (lines) and 2-cells (areas) into meaningful cartographic elements is a major step in map creation. Rather than have each mapping application directly access the TIGER File and independently perform the chaining process, the Census Bureau creates a secondary data base -- a cartographic extract from which mapping applications retrieve the coordinate strings of pre-chained cartographic elements. While developed initially to reduce overall computer processing requirements, the cartographic extract has additional implications for mapping applications, such as providing a structure for approaching the mapping task and enhancing preproduction decision-making processes.

INTRODUCTION

The Topologically Integrated Geographic Encoding and Referencing (TIGER) System represents the culmination of efforts by the United States Bureau of the Census to automate the geographic support processes for the 1990 decennial census and beyond. A major component of these geographic support processes is the production of maps, initially to aid data collection and later to complement data dissemination. The Census Bureau has produced a variety of map types from preliminary versions of the TIGER File to support pre-1990 test census applications. Additionally, the Census Bureau currently is producing maps from the "live" TIGER File in preparation for the 1990 decennial census.

The TIGER File breaks down the network of roads, railroads, hydrography, political boundaries and other features found on the earth's surface into a system of 0-cells (points), 1-cells (lines), and 2-cells (areas). Latitude and longitude coordinate values for 0-cells and 1-cells are stored explicitly, while coordinate values for 2-cells are derivable from the 0-cells and 1-cells. Attribute information about these topological elements, such as the feature name of a 1-cell or the geographic cover of a 2-cell, is stored in numerous related subfiles (Kinnear, 1987).

Collecting and chaining the TIGER File's topological elements into meaningful cartographic elements based on attribute information is a major step in map creation. Furthermore, many mapping applications require the retrieval and chaining of topological elements in the same or a similar way. Rather than have each mapping application directly access the TIGER File and independently perform the chaining process, the Census Bureau creates a secondary data base, the cartographic extract from which mapping applications retrieve the coordinate strings of pre-chained cartographic elements (Bishton, 1988). While initially developed to reduce overall computer processing requirements, the cartographic extract has additional implications for mapping applications.

APPROACHING THE MAPPING TASK

Perhaps most significantly, the cartographic extract provides a structure for approaching the mapping task not readily offered by the topologically-based TIGER File. First, the extract's data base structure suggests a conceptual framework that is consistent with traditional mapping methods. Second, the extract's content supports specific mapping strategies that are compatible with an automated environment.

A Conceptual Framework

The cartographic extract consists of logical subfiles linked by a system of pointers (see Figure 1). Each subfile contains fixed-length records, with each record holding pointer data followed by nonpointer or descriptive data. Balanced tree (B-tree) logical subfiles, also referred to as directories, store records ordered on a key; random access logical subfiles (RALS) store records randomly.

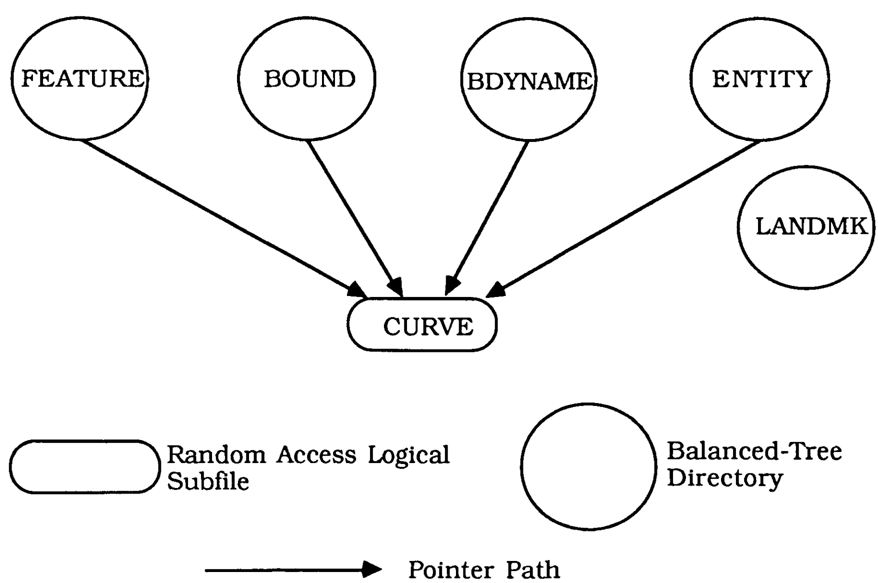

Figure 1. Cartographic Extract Data Base Structure

The FEATURE directory holds records for linear and areal features, including physical features such as streams, glaciers, and ridge lines, and cultural features such as roads, railroads, and airports. The LANDMK directory holds records for physical and cultural point features such as mountain peaks, lookout towers, and churches. Together, those two directories embody the concept of a feature base and a feature names overlay.

The BOUND directory holds records for unique combinations of political and statistical boundary coincidence; for example linear chains that function simultaneously as a county boundary and an American Indian reservation boundary represent a unique boundary combination that is different from linear chains that function simultaneously as a county boundary and a city boundary. Furthermore, because the individual identity of each geographic entity has no bearing on the boundary combination, a linear chain that forms the boundary of county A and county B is not uniquely different from a linear chain that forms the boundary of county A and county C. As the source for boundary symbology, the BOUND directory equates with the concept of a political base.

A political names overlay may be thought of as residing in the BDYNAME and ENTITY directories. The BDYNAME directory allows the linear placement of names parallel with boundary symbology; the ENTITY directory allows the areal placement of names centered within a specific geographic entity. The BDYNAME directory holds records for unique combinations of opposing entity names along state and county boundaries. Unlike the BOUND directory, the individual identity of each geographic entity has a bearing on the boundary name combination; consequently a linear chain that forms the boundary of county A and county B is uniquely different from a linear chain that forms the boundary of county A and county C. The ENTITY directory holds records for most of the political and statistical geographic entities recognized by the Census Bureau for data collection and data tabulation purposes. A record exists for each discontiguous piece of an entity so that every geographic area can be labelled individually.

The map base and map names overlay concepts cannot be fully realized without the CURVE RALS. Each record in the FEATURE, BOUND, BDYNAME, and ENTITY directories has a pointer to a series of records in the CURVE RALS to link a cartographic data element with its chained coordinates. A LANDMK directory record does not require a link to the CURVE RALS because its cartographic data element possesses only one coordinate point, which is stored directly on the LANDMK directory record.

Mapping Strategies

The record content of individual subfiles allows a progression from a conceptual framework toward specific strategies for approaching the mapping task. Figure 2 shows the record layout for each subfile, excluding pointer data. Shading identifies the fields that make up a directory's key. Some noncritical fields have been omitted and others rearranged for the purposes of discussion.

FEATURE DIRECTORY

COORDINATE ENVELOPE	FEATURE CODE	FEATURE NAME	CHAIN NUMBER	ENDPOINT COORDINATE	ADJACENT PARTITION	BOUNDARY FLAGS

BOUND DIRECTORY

COORDINATE ENVELOPE	BOUNDARY CODE	CHAIN NUMBER	ENDPOINT COORDINATE	ADJACENT PARTITION

BDYNAME DIRECTORY

COORDINATE ENVELOPE	ENTITY CODE RIGHT	ENTITY CODE LEFT	ENDPOINT COORDINATE	ADJACENT PARTITION	ENTITY NAME RIGHT	ENTITY NAME LEFT

ENTITY DIRECTORY

COORDINATE ENVELOPE	ENTITY CODE	CHAIN NUMBER	ENDPOINT COORDINATE	ADJACENT PARTITION	ENTITY NAME

CURVE RALS

COORD 1	COORD 2	COORD 3	COORD 4	COORD 5	COORD 6	COORD 7	COORD 8	COORD 9	COORD 10

LANDMK DIRECTORY

POINT COORD	FEATURE CODE	FEATURE NAME

Figure 2. Cartographic Extract Subfile Record Layouts

The coordinate envelope field describes the spatial extent of a chain. The field contains four subfields: the chain's minimim longitude, maximum longitude, minimum latitude and maximum latitude. As the first field of a directory's key, the coordinate envelope is the primary data element by which the directory orders individual records. Directory records are ordered spatially to allow mapping applications to determine quickly which chains fall within a rectangular area to be mapped. While the TIGER File is conceptually a single computer file, it actually consists of many physical files or partitions, with each partition usually covering the area of one county. Because an extract file exists for each TIGER File partition, the cartographic extract also is county-based. For large-scale mapping applications, the area to be mapped often represents a small subset of the county covered by an extract file. By ordering records spatially, many chains falling outside the mapping area may be skipped without a direct comparison between the mapping area's coordinate envelope and each chain's coordinate values.

Just as the coordinate envelope provides a spatial filter within an extract file, the adjacent partition field provides a spatial filter across extract files. If the mapping area extends beyond a county, the mapping application must access more than one extract file to produce a single map. A cartographic problem arises, since each 1-cell along a partition edge is found in two TIGER File partitions and the 1-cell coordinates are incorporated into chains in two extract files. To avoid overprinting symbology, the mapping application must access a chain from only one extract. The adjacent partition field indicates if a chain occurs on a partition edge, and contains the code of the adjacent partition. Following preset rules, such as accessing only from the lower-code partition, the mapping application selects or rejects chains falling on the partition edge.

The feature code and boundary code fields provide potential for cartographic classification and generalization. Turning to the feature code first, every 1-cell and some 0- and 2-cells in the TIGER File carry a Census Feature Class Code (CFCC) (Trainor, 1986). A feature code represents a predefined grouping of CFCCs that can be considered equivalent for mapping purposes; for example, feature code 1 combines 10 CFCCs all under the general heading of "Interstate, U.S., and State highway, not in tunnel or underpassing." If the extract feature groupings are still too numerous for a specific mapping application, groupings can be combined by changing the feature code of several similar groupings to the feature code of one representative grouping. By referencing application-specific lookup tables to direct the feature code changes, one generic mapping program easily produces a variety of map designs.

The boundary code for a chain represents its unique combination of boundary coincidence. The code is the integer equivalent of a boundary number to the 21st power where the nth bit is "on" if the nth boundary type is present; for example, the boundary code of a chain forming a state boundary (bit 20) and a county boundary (bit 19) but no other boundary type equals 786,432 ($2^{20} + 2^{19}$). No mapping application ever requires all 21 boundary types to be symbolized, and some mapping applications require two or more boundary types such as incorporated place boundaries and census designated place (CDP) boundaries to be symbolized as the same boundary type. The bit structure of the boundary code easily accommodates this cartographic generalization and classification. The user redefines the boundary code by turning off all bits of boundary types not required, and by turning on the bit of a representative boundary type (e.g., incorporated place) while simultaneously turning off the bit of any other boundary type it represents (e.g., CDP). As with the feature code changes, a lookup table directs boundary code changes.

Simply altering a feature code or boundary code does not automatically trigger the merging of similar chains into a single, larger coordinate chain. If a mapping application symbolizes all railroad-related feature codes identically, chaining together all like-named chains within the appropriate range of feature codes will maximize the regularity of the railroad symbol pattern and minimize the repetition of the railroad name. If the same mapping application symbolizes only 5 of the possible 21 boundary types, chaining together all extract chains with the same newly-defined boundary code promises a more regular boundary symbology pattern. The mapping application accomplishes this secondary chaining by first identifying which chains belong to a common group using a work directory, and then matching chains within a group on their endpoint coordinates.

Frequently, Census Bureau mapping applications call for certain feature types, such as pipelines and powerlines, to be shown only where coincident with a mapped boundary. Including all pipelines and powerlines leads to unnecessary clutter, while showing no pipelines or powerlines may create confusion at the mapped boundary. The boundary flag field found on the FEATURE directory record provides a connection between feature and boundary chains. Within the boundary flag field, the nth bit is turned on if the nth boundary type is coincident anywhere along the feature chain. By checking whether or not certain boundary bits are turned on, the mapping application selectively suppresses feature chains within a feature group.

The entity code field provides a means of merging statistical values from external files with the extract's cartographic base to produce choropleth maps. The entity code is a character string where the first alphabetic character represents the geographic entity type (e.g., C=county) and the remaining numeric characters identify a specific geographic entity (e.g., 20005=Atchison County, KS). The entity code adheres to the Census Bureau's highly developed and extensively recognized geographic code scheme whereby each geographic entity receives a unique code incorporating hierarchical relationships between entity types (U.S. Census Bureau, 1983). Given the Census Bureau's current focus on data collection activities rather than data dissemination activities, no choropleth maps have been produced from a cartographic extract, although the potential remains.

ENHANCING PREPRODUCTION PROCESSES

Prior to map production, the mapping application selects an appropriate map scale and determines the grid sheet layout including any insets. Increasingly, mapping applications use the cartographic extract to enhance these preproduction design-making processes. In particular, the accessibility of the coordinate envelope or the full coordinate chain of a geographic entity from the extract's ENTITY directory represents an invaluable time-savings in preproduction processing.

The County Block Map, a developing map series designed to clearly display each census block within a county usually requires multiple map sheets at a single scale to cover one county. The map scale is determined by a process that measures overall feature density via the TIGER File (Martinez, 1987). The grid sheet layout for the county is then a function of map sheet size and map scale. When the grid sheet layout approximates the county subdivision boundary network depicted on the map, slivers of county subdivisions appear along map edges. Slight shifting of the grid sheet layout to maximize the number of county subdivisions that fall on a single sheet reduces the slivering effect. After comparing the map sheet coordinate envelopes against the county subdivision coordinate envelopes located in the ENTITY directory, the number of single-sheet county subdivisions is summed for nine different shift directions: a shift to the north, to the south, to the east, to the west, to the northeast, to the northwest, to the southeast, to the southwest and no shift at all. The shift direction producing the largest number of single-sheet county subdivisions is applied to the grid sheet layout.

Insets for a County Block Map are identified by a process that locates local pockets of high feature density via the TIGER File (Martinez, 1989). Since high feature density typically occurs within urban areas, many insets encompass incorporated places and CDPs. When a rectangular inset only approximates the irregular limits of a place, slivers of a place remain outside the inset area. By comparing the inset coordinate envelopes to the place coordinate envelopes located in the ENTITY directory, insets are expanded selectively to reduce the clutter of place slivers.

Another map series, the County Locator Map, uses the extract exclusively for all preproduction processes. The County Locator Map shows the boundary of and centers a label within each address register area (ARA) (a 1990 census collection unit) within a county while showing few cultural or physical features. The preproduction process attempts to strike a balance between maximizing the map scale for label legibility and minimizing the number of map sheets and inset areas for ease of use. The ARA coordinate chains are scaled and ARA label placement is tested through several iterations of grid sheet layouts, progressing from the simplest grid sheet layout (that is, 1 x 1 sheet) to more complex grid sheet layouts (for example, 2 x 3 sheets). Groups of ARAs that cannot be legibly labelled within an iteration form inset areas. The iterative process stops when a more complex grid sheet layout (in effect, increasing the map scale) does not reduce the number of inset areas significantly.

CONCLUSION

As a growing number of mapping applications accesses the cartographic extract, the initial investment in computer processing time to create the extract is returned many times over. Moreover, new and unplanned uses of the extract can initiate modifications to the original extract concept, resulting in a more robust cartographic data base. Modifications currently in development include storing the left-side and right-side geocodes for each linear chain in the FEATURE, BOUND and BDYNAME directories. The left/right geographic information can be used to selectively suppress linear chains falling within specified geographic entities. As we move toward the 1990 decennial census, the cartographic extract concept will continue to evolve to meet the Census Bureau's mapping needs.

REFERENCES

Bishton, A. 1988, Designing and Using a Cartographic Extract: Mapping from the TIGER System: Proceedings URISA, Vol. II, pp. 130-141.

Kinnear, C. 1987, The TIGER Structure: Proceedings, Auto Carto 8, pp. 249-257.

Martinez, A. 1987, Applications of Expert Rules in Automated Cartography, presented at Applied Geography Conference.

Martinez, A. 1989, Automated Insetting: An Expert Component Embedded in the Census Bureau's Map Production System: Proceedings, Auto Carto 9.

Trainor, T.F. 1986, Attribute Coding Scheme: Identification of Features in the U.S. Census Bureau's TIGER System: Proceedings, Auto Carto London, Vol. I, pp. 117-126.

U.S. Bureau of the Census, 1983, 1980 Census of Population and Housing, Geographic Identification Code Scheme, U.S. Government Printing Office, Washington, DC.