THE TIGER STRUCTURE

Christine Kinnear
Geography Division
U.S. Bureau of the Census
Washington, DC 20233

ABSTRACT

In an effort to automate the geographic support process for censuses and surveys, the United States Bureau of the Census has undertaken to design, develop, and implement an integrated geographic support system that will produce computer generated maps, assign addresses to geographic units and delineate geography for the collection, tabulation and publication of census data. The core of this effort is the Topologically Integrated Geographic Encoding and Referencing (TIGER) System, a collection of computer files and software. The central file of this system is the TIGER data base that houses the data required to perform the geographic support functions. This paper will describe the TIGER File structure, the elements that make up the file and the linkages between the elements.

OVERVIEW

The overall structure is shown in Figure 1. While it is one interlocking structure, it can be viewed as many smaller substructures, each performing a specific function. Four broad views are presented here: geometry, geography, linear attributes, and landmarks and related entities.

The TIGER File contains roads, hydrography, boundaries, railroads and some miscellaneous features that are represented as single lines. The location in space of the intersections of these lines is known to the TIGER file; and the areas enclosed by the intersecting lines are the atomic areal unit maintained in the file. These three units are known as 1-cells, 0-cells and 2-cells, respectively. They represent the geometry in the file and are related such that 1-cells are bounded by 0-cells and 2-cells are bounded by 1-cells. This relationship is the topological base upon which the remaining data are dependent.

The geographic elements are the areal cover and are linked to the topology via the 2-cells. There are three general sets of geographic combinations, made up to five specific components, that cover the 2-cells directly: the 1980 blocks and Geographic Tabulation Unit Base (GTUB), the 1990 blocks and GTUB, and the 1990 Ancillary GTUB. The 1980 data are kept for postcensus comparability studies. The Ancillary GTUB will be used to store the geographic entities not required for initial census tabulations, but needed for special postcensus tabulations. Individual geographic entities are linked to the three GTUBs. Each entity is made up of many GTUBs and each GTUB in turn is made up of many 2-cells.

Each 1-cell may be labeled with one or more feature identifier (a name), though it is expected it will be primarily roads and hydrography that will be named. Each feature identifier in turn may have one or more 1-cells attached to it. Each 1-cell/feature identifier combination may have one or more address range pair linked to it if the 1-cell is an addressable feature. The five-digit ZIP code is a linear attribute in the TIGER File and is linked directly to the address range pair. Therefore, the address range/ZIP code combination is linked indirectly to 1-cells.

The last view is the remaining miscellaneous items of Landmarks, Areas, and Key Geographic Locations (KGLs), collectively known as LAKs. Landmarks and KGLs are free floating points and are associated with 2-cells and 1-cell sides, respectively. Landmarks are cartographic enhancements, are linked directly to a 2-cell, and are generally unnamed. KGLs, mainly commercial buildings, are in the file as alternative addressing schemes and therefore are linked directly to an address range and subsequently to a 1-cell side; KGLs are always named. Areas are specially defined areas that are not part of the normal geographic hierarchy but are significant enough to warrant recognition, such as national parks and large water bodies. Areas consist of one or more 2-cells and may or may not be named.

The TIGER physical file is made up of many logical subfiles, each of which is generally one of the components shown on the structure diagram. Each subfile is made up of fixed length records that contain pointer data and descriptive data for the item represented. A subfile has data stored either ordered by a key (a balanced tree logical subfile), or randomly (Random Access Logical Subfile or RALS). Each subfile is linked to at least one other subfile. The balanced trees are the directories and provide the entry points into the data base.

The relationships in the TIGER file among the subfiles are explicitly represented by pointers. These relationships are specifically defined and are described as lists.

LISTS

An owner or head of a list is a single entity that is related potentially to many others. A member or tail is an entity that is in a subordinate relationship to another entity. Pointers are linkages between records and may be either a directory key or a subfile position number. (Unless otherwise noted, reference to pointers means position number.) There are three classes of pointers: 1) first or head to tail; 2) head/owner or tail to head; 3) next or tail to tail.

A list is a chain of like elements all of which have a common attribute. A list is effected by a combination of a first, head, and next pointers. It has an owner record on which is stored the common attribute and a first pointer to a tail record. The chained elements are the tail records that may have an owner pointer and may have a next pointer

to another tail record.

There are five characteristics of lists in the TIGER File.

1. Simple: The owner has a first pointer and the tail may contain a next pointer. Simple lists can be one-way, where there is either a first or last pointer; or two-way where there are both.

2. Intertwined: A form of a simple list where the owner has a first pointer and the tail has a pair of next pointers, either of which may be used for the next record. Each tail is in two list of the same type simultaneously.

3. Many-to-many: This is a combination of two related simple lists each list owner having a first pointer to a shared relationship record. These tail records have next and owner pointers for each head.

4. Multilist: These are parallel simple lists where there are many owners with first pointers and the tail records contain parallel next and owner pointers.

5. Indexed: A simple list whose head is an ordered directory record with a first pointer and the tail may or may not contain a next pointer. Directory keys, rather than record numbers are used for head pointers.

Lists are the traversal mechanism in the TIGER File. Each is specifically defined in terms of its owners and its members. Referring to the structure diagram, lists are represented by the lines and arrows between subfiles. The double arrows indicate the next pointers of a list; or in other words, the subfile pointed to is a tail record of a list with next pointers. The single arrow on the other end of the double-arrow line represents the presence of an owner pointer on the tail record. A line with a single or double arrow at only one end represents a tail record without a pointer back to its owner. The many-to-many lists are represented on the structure diagram by the presence of a relationship subfile. Extending each owners double-arrow line through the relationship subfile to the other owner shows this relationship.

SUBFILES

The following discussion is a brief summary of the primary subfiles, including comments on the their relationships to other elements.

Ø-Cells

Ø-cells as the TIGER File views them are, in addition to points of intersection, terminating points of a line and intermediate points on a line that represent major attribute changes. Each Ø-cell is represented twice in the file, once on the Ø-cell directory in the form of a Peano key and once on the Ø-cell RALS in the form of longitude

and latitude. Each Ø-cell is the owner of an intertwined list that has the 1-cells as tails.

A Peano key is a merge of the alternating bits of the coordinates of a point. It allows storage of the coordinates in a one-dimensional array.

2-Cells

2-cells are the atomic areal unit in the file. The TIGER File allows internal structures, both contiguous to and isolated from the bounding 1-cells, to be part of a 2-cell. The 2-cell RALS contains one record for each unique 2-cell. Though some data is stored on the record, the critical function of the 2-cell is the link between all geography and the topology. Each 2-cell is the owner of the intertwined list between 2-cells and 1-cells; and it is the tail on all GTUB and block related lists.

1-Cells

1-cells are the joining elements in the TIGER File. They link to feature identifiers, addresses, ZIP codes, KGLs; are bounded by Ø-cells; they bound 2-cells that link to all geography and landmarks and areas. The 1-cell RALS stores each unique 1-cell once, uniquely defined by the bounding Ø-cells plus curvature. Stored on the 1-cell record is an envelope, a minimum area that encloses the 1-cell plus its curvature. This is used to reduce spatial searches involving 1-cells. The remaining data consist mainly of boundary flags that identify the major geographic and statistical areas that the 1-cell bounds. The 1-cell is the tail of the two intertwined lists from the Ø-cells and 2-cells. It is these two list that represent the topological relationship in the TIGER File. A 1-cell may be the owner of one of the related simple lists in the many-to many relationship between 1-cells and feature identifiers. A 1-cell may also be the owner of the only two-way list in the TIGER File, the 1-cell to its curvature points.

1-Cell Curvature

1-cell curvature records can be perceived as an extension of the geometry of the 1-cell. Each 1-cell to curvature list contains all the coordinates for the intermediate points between the two Ø-cells. There is no limit to the number of curvature points a 1-cell may have. For plotting and computational convenience, the 1-cell to curvature list can be traversed both forward and backward.

Feature Identifiers and Feature Name Continuations

The feature identifiers are the labels of nongeographic entities, such as streets, rivers, buildings, and parks. Each unique identifier is represented only once in the feature identifier RALS even if it labels both a linear and an areal feature. Each identifier consists of at least a name and may include directions, such as east or northwest, and feature types, such as avenue, river, or building. There is a directory for feature identifiers and that directory is ordered by a Soundex key plus a truncated, packed version of the name. Therefore a directory entry may link to one or more feature identifier RALS records.

Since records are fixed length, a name continuation subfile is used to store name text that exceeded the fixed length. A feature identifier may be the owner of up to two related simple lists of two many-to-many lists: the feature identifier to 1-cell and the feature identifier to LAKs.

Address Ranges

Address ranges are stored as a left/right pair, with orientation being the same as the 1-cell they are associated with. Since address ranges are not linked directly to a 1-cell but to the combination of a feature identifier and 1-cell, each alternate identifier may have its own unique set of address ranges, allowing parallel address systems to be represented. The majority of the information on the address range RALS consists of flags that indicate the type of range, potential error conditions, and verified anomalies. The nine-digit ZIP code is considered part of the address range; the five-digit ZIP code being implicitly on the record as the ZIP code owner pointer of ZIP code to address range list. The four-digit add-on is stored explicitly as 1-cell side data. An address range may be the owner of a cross-reference address.

Cross-Reference Addresses

The Key Geographic Locations in the TIGER File are part of the overall automated address geocoding process and therefore need to be linked to a 1-cell side. Links to 1-cell side are via an address range; therefore, a KGL needs to have some link to an address range. The cross-reference address record serves this purpose with an actual or imputed address. A cross-reference address is the owner of an address to KGL list. This subfile is not intended for storage of the 1990 census individual addresses.

ZIP Codes

ZIP codes in the TIGER File are looked upon as a linear or a point attribute. Each five-digit ZIP code that occurs in the file is stored once on the RALS record and once in the directory. The ZIP code RALS record is the owner of two lists, the ZIP code to address range and the one of the related simple list in the ZIP code to LAK many-to-many list.

Landmarks, Areas and Key Geographic Locations

This subfile is the collection of the three LAK entities described in the overview. There is one record for each unique LAK. Some information stored on the record is unique to specific entities, but flags indicate what usages are intended for each LAK. The LAK subfile may be the owner of up to three related simple lists of three many-to-many lists: LAK to feature identifier, LAK to ZIP code, and LAK to 2-cell.

Geographic Entities, Geographic Entity Extensions

Each of the 26 geographic entities recognized for decennial census processing is represented on the entity subfile; each entity is identified with its own record type. The data varies from entity to entity but each record contains

as a minimum the combination of FIPS and Census codes required to uniquely identify the entity. An open ended entity extension subfile stores additional data for various entities. Currently record types exist on this subfile for areal data and entity names. Other record types may come into existence as required. The entity directory and RALS each have one record for each unique entity.

The entity subfile records are both the owners and the members in the many-to-many relationships among entity types. This allows an entity to be either dominant or subordinate to any other entity. The entity record is the owner in all the entity to GTUB multilists.

1980 Geographic Tabulation Unit Base

Each unique combination of 1980 tabulation geography is stored on both the 1980 GTUB RALS and directory. Except for detected errors, the 1980 GTUB will not be updated once it is associated with its 2-cells. Because of residual errors, some 2-cells may have no 1980 GTUB owner and some GTUBs may have no 2-cell members.

1990 Geographic Tabulation Unit Base

Each unique combination of 1990 primary publication geography is stored on both the 1990 GTUB RALS and directory. The geography will be current until the final update, at which time it will represent the final 1990 geography. Each GTUB will be the owner of a 1990 GTUB to 2-cell list.

1990 Ancillary Geographic Tabulation Unit Base

The Ancillary GTUB RALS and directory contain one record for each combination of additional geographic codes. The geography on this subfile represents additional postcensus tabulation requirements. It also documents historical areas. Each GTUB will be the owner of an ancillary GTUB to 2-cell list.

1980 Block/Enumeration District

This subfile contains the 1980 block numbers and enumeration districts (ED). There are two record types, a block record and an ED record. There is no overlap of coverage; either a 2-cell was blocked in 1980 and it therefore has only a block record; or it was not blocked and has an ED record. There is one record for each unique block code or enumeration district on both the RALS and the directory.

1990 Block/Address Register Area, Block/ARA Extension

This subfile contains information for the smallest geographic area defined for the 1990 census. The block RALS has two record types, a block group (first digit of a block number), and a block. The 2-cells initially are linked to the block group record, but once collection blocks are assigned, the link is to blocks only and block group records become summary data only. Collection blocks become tabulation blocks once splits by final 1990 geography are effected. The Block/ARA extension subfile provides for additional data relating to blocks. Currently

areal data is stored, but additional record types may be added if required.

The directory contains three record types, one for address register area (ARA), which is equivalent to a block group, but assigned a code unique within district office; one for block group; and one for block. Both the ARA and block group records link to the same block group record on the block RALS

Relationship Subfiles

This group of subfiles exists primarily to allow many-to-many relationships. This relationship shows two complementary views: each owner may have many members and each member may have many owners. Each one of these subfiles in the TIGER File has data that relates only to the combination of the two heads of the related simple lists. The 1-cell/feature identifier relationship record is the only one of this category that becomes an owner of another list. This occurs because address ranges are not associated directly with a 1-cell or a feature identifier, but rather with that 1-cell when it is known as that feature identifier. Therefore, this relationship record may become the owner of a list with address ranges as tail records.

The three relationship subfiles that are associated with the landmarks, areas, and key geographic locations are shown as separate entities, but they are in fact stored in one physical subfile with the data distinguishing one from the other.

SUMMARY

The data base described in this paper is designed to be resident on one of the Census Bureau's mainframe Unisys (Sperry) 1100 series computers. The software to manipulate this data base is written in FORTRAN-77 with Unisys supported extensions. The TIGER file will made up of many physical files partitioned by geographic units, mainly county or groups of counties.

The TIGER File structure described here is not the final structure, but is the TIGER-I(nterim) File to be used for the 1990 Decennial Census and the 1992 Economic and Agriculture Censuses. The TIGER System will continue to evolve throughout the next decade until it reaches its full expectations.

REFERENCE

U.S. Bureau of the Census, Geography Division, 1986, TIGER-I System Documentation.

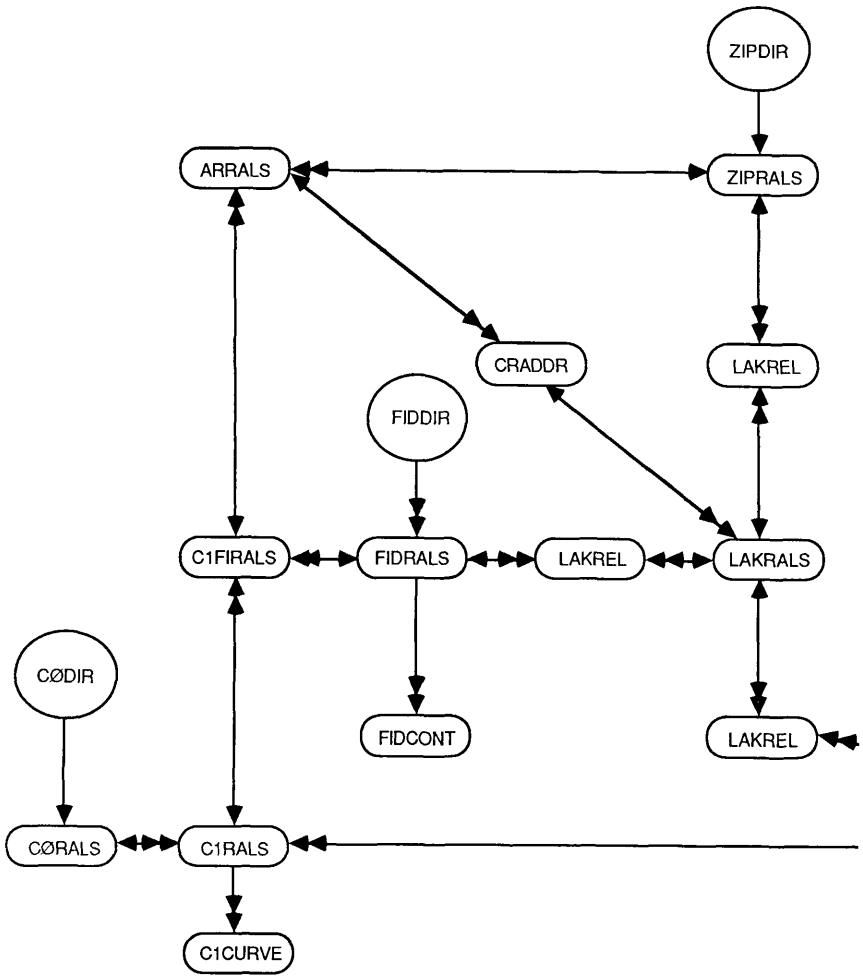


FIGURE 1. TIGER-I SYSTEM DATABASE STRUCTURE

KEY TO SYMBOLS:

RANDOM ACCESS LOGICAL SUBFILE

BALANCED-TREE DIRECTORY

ONE-TO-ONE RELATIONSHIP

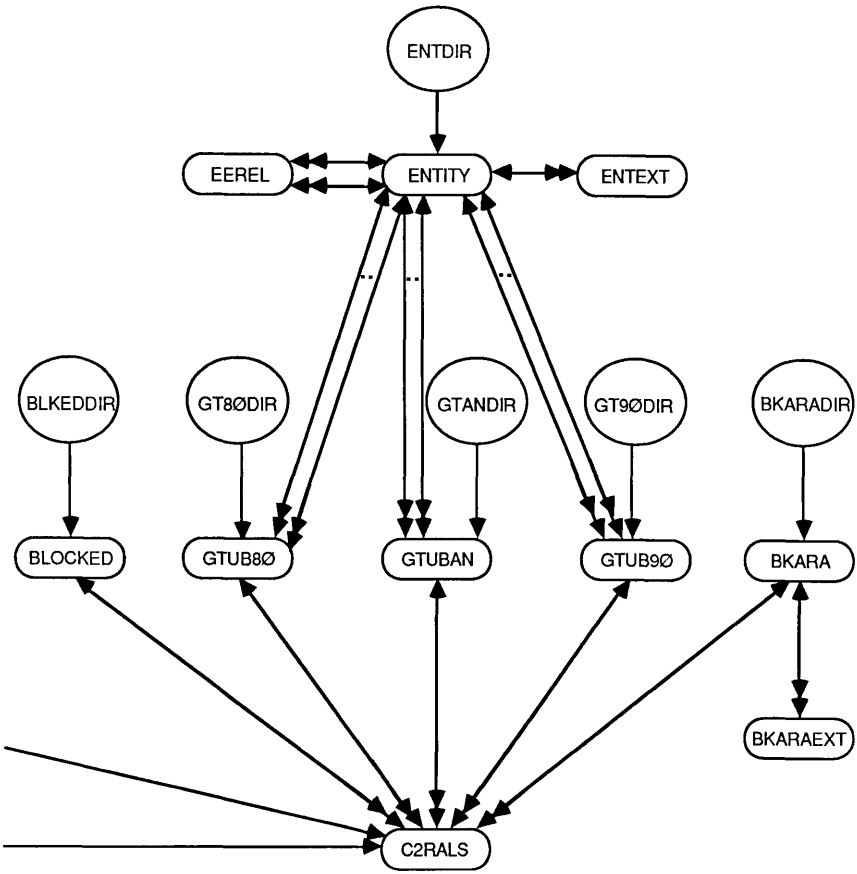
ONE-TO-MANY RELATIONSHIP

ONE-TO-MANY RELATIONSHIP WITH MEMBER POINTING BACK TO ITS OWNER

•• PARALLEL MULTIPLE LINKAGES

KEY TO SUBFILE ABBREVIATIONS:

ARRALS - *ADDRESS RANGES*
BKARA - 199Ø BLOCK/ADDRESS REGISTER AREA (ARA)
BKARADIR - 199Ø BLOCK/ARA DIRECTORY
BKARAEXT - 199Ø BLOCK/ARA EXTENSION
BLKEDDIR - 198Ø BLOCK/ENUMERATION DISTRICT DIRECTORY
BLOCKED - 198Ø BLOCK/ENUMERATION DISTRICT
CØDIR - Ø-CELL DIRECTORY
CØRALS - Ø-CELL
C1CURVE - 1-CELL CURVATURE
C1FIRALS - 1-CELL/FEATURE IDENTIFIER RELATIONSHIP



KEY TO SUBFILE ABBREVIATIONS (CONTINUED):

C1RALS - 1-CELL
C2RALS - 2-CELL
CRADDR - CROSS-REFERENCE ADDRESS
EEREL - ENTITY TO ENTITY RELATIONSHIP
ENTDIR - GEOGRAPHIC ENTITIES DIRECTORY
ENTEXT - GEOGRAPHIC ENTITIES EXTENSIONS
ENTITY - GEOGRAPHIC ENTITIES
FIDCONT - FEATURE NAME CONTINUATION
FIDDIR - FEATURE IDENTIFIER DIRECTORY
FIDRALS - FEATURE IDENTIFIER
GT8ØDIR - 198Ø GEOGRAPHIC TABULATION UNIT BASE DIRECTORY
GT9ØDIR - 199Ø GEOGRAPHIC TABULATION UNIT BASE DIRECTORY
GTANDIR - 199Ø ANCILLARY GEOGRAPHIC TABULATION UNIT BASE DIRECTORY
GTUB8Ø - 198Ø GEOGRAPHIC TABULATION UNIT BASE
GTUB9Ø - 199Ø GEOGRAPHIC TABULATION UNIT BASE
GTUBAN - 199Ø ANCILLARY GEOGRAPHIC TABULATION UNIT BASE
LAKRALS - LANDMARKS, AREAS, KEY GEOGRAPHIC LOCATIONS
LAKREL - LANDMARK, AREA, KEY GEOGRAPHIC LOCATION/ FEATURE IDENTIFIER, 2-CELL, ZIP CODE RELATIONSHIP
ZIPDIR - ZIP CODE DIRECTORY
ZIPRALS - ZIP CODE