UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233

1989

PUTTING PEOPLE IN YOUR GIS

by

ROBERT W. MARX
Chief, Geography Division

Presented at the
American Agricultural Economics Association Meeting

July 28-August 2, 1989
Baton Rouge, Louisiana, U.S.A.

Geography, Economic 2183

#7622

Putting People in Your GIS

It's very appropriate that I should have the opportunity to deliver this talk at this location -- the campus of LSU and its famous Tigers.

A figure famous to geographers, Gilbert M. Grosvenor, President and Chairman of the illustrious National Geographic Society, frequently says, "If you don't know where you are, you're nowhere." In GIS, we can start to find the answer.

While it's nice to know where you are, it's even nicer to have some idea of where you are going. Many years ago Cervantes, the creator of that crazy but occasionally profound dreamer Don Quixote, recognized the value and intrigue of maps by exclaiming that "...in a map, one could journey over all the universe, without the expense and fatigue of traveling, without suffering the inconveniences of heat, cold, hunger, and thirst."

Even in his wildest dreams, old Don Quixote could not have conceived of the maps people are creating these days. Using computerized mapping and geographic information system techniques, we can travel from the vastness of outer space, toward the earth, seeing more and more detail, travelling through the familiar range of small-scale maps and larger-scale maps to the other end of the map scale continuum, where humans constantly strive for greater and greater detail, which demands greater and greater accuracy in observation and measurement.

And that brings me to the subject of geographic information systems (GIS). What does this have to do with GIS, you ask? And what does GIS have to do with the American Agricultural Economics Association? Well, that depends on your definition of a GIS.

The debate goes on endlessly. Some argue that the only "true" GIS is a system that processes overlays of information; others contend that a GIS must include natural resources data. Still others argue endlessly about the level of accuracy one needs in the coordinates of the features and data polygon boundaries, and I'm sure you are aware of the enormous amounts of energy people are expending these days to study the physical conditions that surround us under the rubric of "global change." Using a GIS certainly will help in these studies.

I believe that GIS technology can help you. It is in this context that I want to suggest another definition -- perhaps a bit more generic: "A GIS is a computer system that helps us discover relationships between and among sets of data that we could not see or understand before." In this broader context, I believe all the preceding views of a GIS are missing one important ingredient -- people!

People inhabit almost every part of the earth. They either affect what is going on or are affected by what is going on. This is even more stunning when you stop to realize that in the next 50 years, the earth's population will double. This growth likely will continue the dramatic shift from rural settlement to urban settlement -- the urban population of the world likely will triple in the same 50 years. Certainly these shifts are having...and will continue to have...a dramatic affect on the agricultural production and consumer demand situations. As a result, I suggest that people must be a "layer" in your thinking about GIS.

"People" information traditionally has not been in a form that computers could process conveniently in a GIS context. On the natural resources side of the aisle, we have satellite imagery and, as Carl Sagan would say, its "billions and billions" of pixels. I'd like to extend the notion of pixels to the context of people in a GIS and coin a new term -- censels.

The way in which the Census Bureau collects and tabulates data is by "polygon" -- just like those people collecting and studying soils or land use/land cover information. More importantly, in creating "people polygons," the Census Bureau routinely uses the same earth surface features most people want -- and include -- in their GIS: streets, roads, streams, railroads, governmental unit boundaries, and so forth. And this leads me to a discussion about tigers!

People normally visualize a tiger as a large, yellow, cat-like animal with black stripes. The TIGER to which I refer is an altogether different creature. It is an acronym we at the Census Bureau created to refer to our new computerized geographic support system; the Topologically Integrated Geographic Encoding and Referencing System. The TIGER System is not a GIS, per se. It is a very valuable source of information that can form one important component of a GIS. Let me explain.

The Geography Division, one component of the Census Bureau, developed the TIGER System in response to a major goal the agency set in 1981: to automate the full range of cartographic and geographic support processes in time to serve the data collection, tabulation, and dissemination needs of the 1990 decennial census -- the bicentennial census of the United States.

The charge included having the first geographic products from the TIGER System ready when the preparatory field activities for the 1990 census began in early 1988. This gave the Geography Division:

- Six short years to design, develop, test, and implement a computer data base like none in existence before: a data base that would handle both the cartographic and geographic tasks for which we are responsible.

- Six short years to identify, procure, install, and learn to use the graphic work stations, computer-driven plotters, and new host minicomputers required to build and use the TIGER data base.

- Six short years to build a computer file containing every known street and road in the United States, the name of each, and the range of address numbers located along each section of every street in the 345 largest urban areas of the United States.

- Six short years to include all the railroads in the United States and the names of their operating companies, along with all significant hydrographic features and their associated names.

To create an automated map base of this magnitude, the Census Bureau entered into a major cooperative agreement with the U.S. Geological Survey -- our Nation's map maker. Under the terms of the agreement, the USGS used its automated scanning equipment and sophisticated computer processing techniques to convert the roads, rivers, lakes, railroads, and so forth on its conventional maps for the lower-48 states into computer files.

In this era of limited resources, especially for those of us in public service positions, such sharing is critical. This process provided a fully sufficient cartographic base for taking a census -- even at the city block level. I suggest that TIGER -- the derived cousin of the USGS files -- will provide a fully adequate base for most agricultural planning applications. But I digress...

The Census Bureau followed the production steps of the cooperative project with the vertical integration of the separate USGS data layers and the update of the street network in these files with new streets, street names, and address ranges, using source materials collected from city, county, and state governments. We followed this with the horizontal integration and edge matching of the adjacent enhanced USGS files with the updated GBF/DIME-Files prepared for earlier censuses -- to form county-based files rather than quadrangles, the format in which the information came from the USGS.

- In addition to building an up-to-date feature base for the whole U.S., we had six short years to enter and verify the boundaries, names, and numeric codes for all the geographic areas used by the Census Bureau to tabulate the results of both the 1980 and 1990 censuses -- which also provides the more limited set used for the agriculture censuses.

I am delighted to report that this development task is complete; the overall goal, nearly met. I'm telling you all this background of the TIGER System because the TIGER data base structure provides a bold new approach in automated cartographic and geographic files. It adapts the theories of topology, graph theory, and associated fields of mathematics to provide a disciplined, mathematical description of the geographic structure of the United States. (See Figure 1)

At the base of this structure, and linking all other components together, is the spatial or geometric framework. The spatial information in the TIGER data base includes the coordinate locations of all known street intersections -- and the points at which streets cross rivers, railroads, and other mapped features -- roughly 28 million points. It also includes the information that shows which of those points are connected by the 40 million lines that define the transportation and hydrography network of the United States and which sets of those lines enclose the 12 million polygons that comprise the TIGER data base.

The TIGER data base is much more than points, lines, and areas, however. The majority of the TIGER data base structure is composed of the attributes describing the various components of the spatial structure. It is this specific set of attributes that make the TIGER data base useful for the Census Bureau; and it is this set of attributes that will put people into GISs. Let me describe these attributes briefly.

First, the feature attributes. This information includes street names and river names along with address ranges and ZIP codes. It also includes landmarks -- named areas such as major parks and military bases -- plus named apartment and office buildings that are important as alternate ways to address mail; for example, the Empire State Building in New York City or the Premier Tower here in Baton Rouge.

Most importantly, the attributes in the TIGER data base include all the geographic entities for which the Census Bureau tabulates the data it collects. We have included both the 1980 census set of areas and the 1990 census set. (See Figure 2) These geographic entities are the "censels" I discussed earlier. Each censel has a string of data associated with it in Census Bureau data files -- just like a pixel in a satellite transmission. There are nearly 3,300 censels that we call counties -- but that's pretty coarse resolution for people characteristics.

There are over 39,000 units of local government subdividing those 3,300 counties -- townships and cities and villages. They provide more than a 10-fold increase in the number of censels, but still offer fairly coarse resolution. On the plus side, demographic data are available for all of these governmental units and economic data are available for many of them.

To zoom in closer, there are over 65,000 census tracts for the 1990 census -- a big increase over the 1980 census because these areas (and their cousins -- the block numbering areas) cover the entire

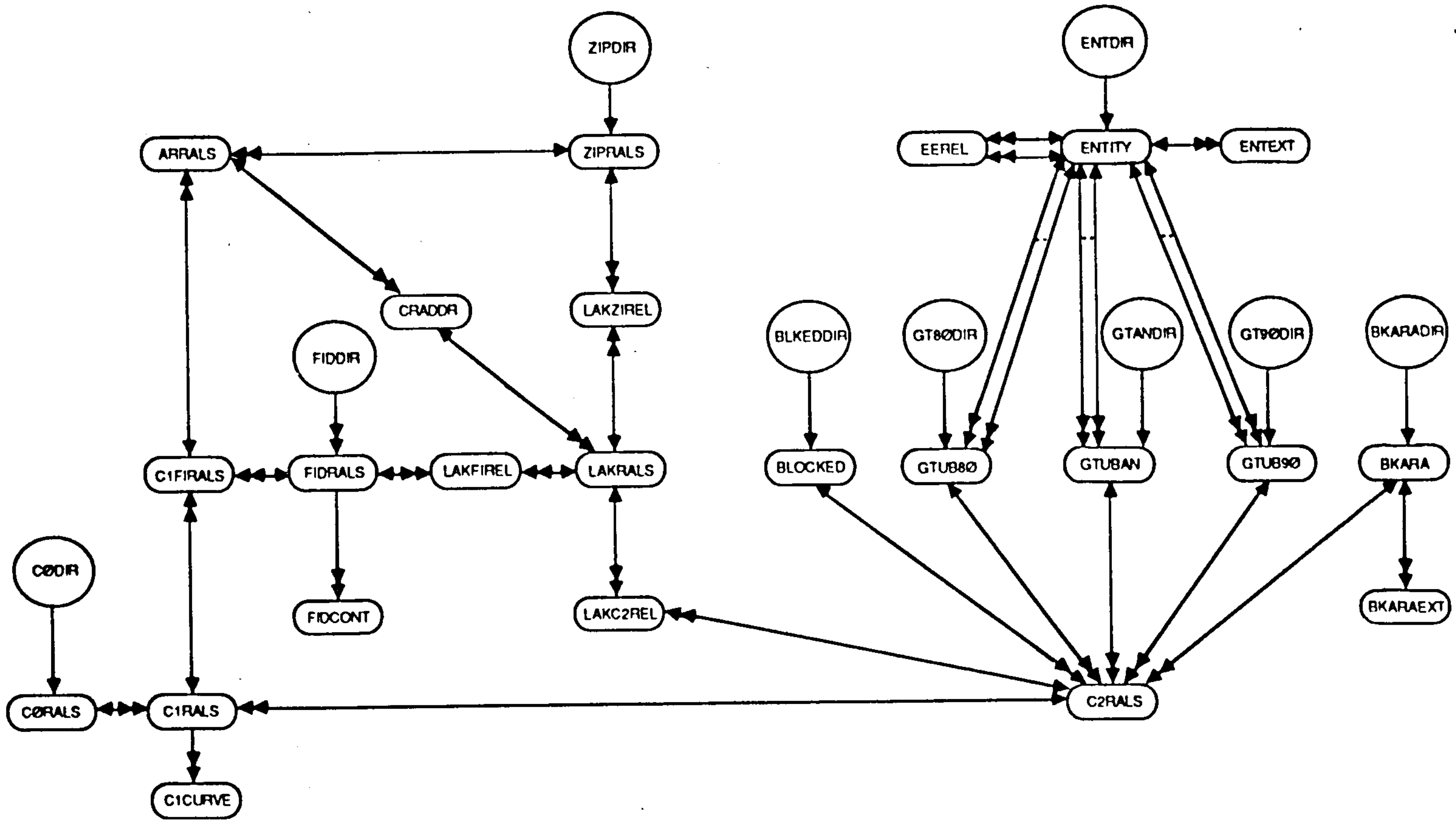

Figure 1: TIGER Data Base Structure

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
LAKC2REL - LANDMARK, AREA, KEY GEOGRAPHIC LOCATION/ 2-CELL RELATIONSHIP
LAKFIREL - LANDMARK, AREA, KEY GEOGRAPHIC LOCATION/ FEATURE IDENTIFIER RELATIONSHIP
LAKRALS - LANDMARKS, AREAS, KEY GEOGRAPHIC LOCATIONS
LAKZIREL - LANDMARK, AREA, KEY GEOGRAPHIC LOCATION/ 5-DIGIT ZIP CODE RELATIONSHIP
ZIPDIR - ZIP CODE DIRECTORY
ZIPRALS - ZIP CODE

United States. These census tracts and BNAs are subdivided into over 200,000 block groups that further segment the governmental units for purposes of data presentation. At this level, a GIS user has access to the full range of decennial census responses -- 100 percent and sample.

Finally, there are between 7 and 9 million census blocks -- people polygons -- for which we will tabulate the 100 percent data from the 1990 census. As with the census tracts, this represents a huge increase over the 1980 census when block-level data were available primarily for the urban cores of metropolitan statistical areas. These "millions and millions" of blocks provide a very fine-grained censal resolution to the demographic data sets available from the Census Bureau.

The TIGER System links the Census Bureau's geographic areas directly with this network of features, and thus, to many other data sets available in a local GIS. The codes used to identify the Census Bureau's geographic areas provide the capability for a computer to link with, and perform analyses on, the data sets that flow from the decennial, economic, and agriculture censuses, and from the Census Bureau's sample surveys.

In earlier days, we thought about such uses in terms of thematic maps and struggled to get the computer to help us with limited data sets. For a GIS user now days the problem is not a lack of data in machine-readable format nor a lack of computer hardware on which to process those data. Quite the opposite -- asking for an item of Census Bureau data can be likened to trying to take a sip of water from a fire hydrant!

In a GIS environment, one can examine other geographically distributed data sets -- soil categories, hazardous wastes, water quality, land use/land cover, and so forth -- in the context of the governments responsible for managing an area, and in the context of the characteristics of the people who occupy the land, their farms, their business, and their industrial activity. The feature and landmark attributes discussed earlier provide an additional linkage mechanism to local data sets.

Knowledge of these landmarks may be a critical piece of information in understanding other component information in a GIS, such as explaining why housing units in the center of a city have incomplete plumbing -- because they are student housing near the university where the students share these facilities.

The TIGER data base, the demographic, economic and agricultural statistics available from the Census Bureau, the geographically referenced information maintained by state and local governments, and the analytical power available to study the interrelationships of

POLITICAL AREAS	STATISTICAL AREAS
United States	Regions (4)
States & State Equivalents (57)	Divisions (9)
States (50)	Standard Consolidated Statistical Areas — SCSA (17)
District of Columbia (1)	Standard Metropolitan Statistical Areas — SMSA (323)
Outlying Areas (6)	Urbanized Areas — UA (373)
Counties, Parishes, & Other County Equivalents (3,231)	Census County Divisions — CCD (5,512)
Minor Civil Divisions — MCD (30,491)	Unorganized Territories (274)
Incorporated Places (19,176)	Census Designated Places — CDP (3,733)
American Indian Reservations (275)	Census Tracts (43,383)
Indian Subreservation Areas (228)	Block Numbering Areas — BNA (3,404)
Alaska Native Villages (209)	Enumeration Districts — ED (102,235)
Congressional Districts — CD (435)	Block Groups — BG (156,163) (Tabulated parts —(197,957)
Election Precincts (36,361) [In 23 participating States]	Blocks (2,473,679) (Tabulated parts — 2,545,416)
School Districts (16,075)	Traffic Analysis Zones (≅ 160,000)
Neighborhoods (28,381)	
ZIP Codes (≅ 37,000)	

Figure 2: 1980 Census Geographic Areas

these data sets using the GIS technology now available in the private sector to merge and manipulate these massive and diverse data sets, make better understanding a very real possibility.

Meetings such as this promote the exchange of ideas that will lead, ultimately, to much more powerful and productive systems. The TIGER data base is available now. Many of the GIS vendors already have interfaces prepared to read the TIGER/Line files -- the first public version of TIGER -- into their GIS. Many of the other vendors have development of such interfaces underway. For more information on the TIGER/Line files contact the staff at the address and telephone number shown below.

State and Regional Programs Staff
Data User Services Division
Bureau of the Census
Washington, D.C. 20233

(301) 763-1580

Reference Materials

Boudriault, G., *Topology in the TIGER File*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp. 258-263, 1987

Broome, F. R., Beard, C., and Martinez, A., *Automated Map Inset Determination*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp. 466-470, 1987

Callahan, G. M. and Marx, R. W., *The 1:100,000-Scale Data Base: Implications for GIS in Illinois*, **Illinois Mapnotes**, Issue Nos. 10/11, pp. 6-8,1988

Fairchild, D., *The Display of Boundary Information: A Challenge in Map Design in an Automated Production System*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp. 456-465, 1987

Kinnear, C., *The TIGER Structure*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp. 249-257, 1987

LaMacchia, R. A., Tomasi, S. G., and Piepenburg, S. K., *The TIGER File: Proposed Products*, paper distributed at the National Conference of State Legislatures, November 1987

Martinez, A., *Application of Expert Rules in Automated Cartography*, paper presented at the 10th Annual Applied Geography Conference, October 1987*

Marx, R. W., *Computer Mapping and the Census Bureau: The Future and You*, paper presented at the AM/FM International Eastern Regional Conference, Williamsburg, Virginia, May 1986 *

Marx, R. W., *LIS/GIS Activities of the United States Census Bureau*, paper presented at an information exchange sponsored by the United States Embassy, Bonn, Federal Republic of Germany, November 1988

Marx, R. W., *The TIGER File Supports Census Bureau Tests*, **Illinois Mapnotes**, Issue No. 9, pp. 2-5, Winter 1987

Marx, R. W., *The TIGER System: Automating the Geographic Structure of the United States Census*, **Government Publications Review**, Vol. 13, pp. 181-201, 1986

Marx, R. W., *The TIGER System: Six Years to Success*, presented at the 13th International Conference of the International Cartographic Association, Morelia, Mexico, October 1987

Marx, R. W., *TIGER: An Automated Cartographic/Geographic Data Base*, **Illinois Mapnotes**, Issue No. 6, pp. 1-2, Spring 1985

Marx, R. W. and Saalfeld, A. J., *Programs for Assuring Map Quality at the Bureau of the Census*, **Proceedings of the 4th Annual Research Conference**, pp. 239-259, 1988

McDowell, T., Meixler, D., Rosenson, P., and Davis, B., *Maintenance of Geographic Structure Files at the Bureau of the Census*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp. 264-269, 1987

McKenzie, B. Y. and LaMacchia, R. A., *The U.S. Geological Survey-U.S. Bureau of the Census Cooperative Digital Mapping Project: A Unique Success Story*, paper presented at the Fall Meeting of the American Congress on Surveying and Mapping, Reno, Nevada, October 1987

Meixler, D. and Saalfeld, A. J., *Polygonization and Topological Editing at the Bureau of the Census*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp. 731-737, 1987

Sobel, J., *Principal Components of the Census Bureau's TIGER File*, **Research in Contemporary and Applied Geography: A Discussion Series**, Vol. X, No. 3, pp. 1-16, 1986

Trainor, T. F., *Attribute Coding Scheme: Identification of Features in the U.S. Census Bureau's TIGER System*, **Proceedings: AUTO CARTO London 86**, Vol. 1, pp. 117-126, September 1986

U.S. Bureau of the Census, *TIGER Tales*, Washington, DC 20233 *

U.S. Geological Survey, *1:100,000-Scale Digital Cartographic Data Base For Federal Requirements*, **Year Book, Fiscal Year 1984**, Reston, Virginia 22092

Wanie, T. L., *Estimating Production Parameters For Large-Volume Automated Mapping on a Raster Plotter*, **Proceedings: Eighth International Symposium on Computer Assisted Cartography**, pp 270-277, 1987

White, M. A., *Technical Requirements and Standards for a Multipurpose Geographic Data System*, **The American Cartographer**, 11, 1, pp. 15-26, 1984

* A package of all unpublished papers is available for $5 from Customer Services, Data User Services Division, Bureau of the Census, Washington, DC 20233, (301) 763-4100.