WAYS OF STRUCTURING DATA WITHIN A DIGITAL CARTOGRAPHIC DATA BASE

Dean T. Edson
George Y. G. Lee
U.S. Geological Survey
345 Middlefield Road
Menlo Park, California 94025

BIOGRAPHICAL SKETCH

Dean Edson, a Californian, began his career in mapping with the Corps of Engineers during WWII. He joined USGS in 1947 and worked in all phases of mapping in the Western States, including both field surveys and photogrammetry. In the austral summer of 1962-63, he was a member of a survey team establishing map control in Antarctica. On assignment to the research staff, he helped build the computer service division of USGS, including purchase of a large computer. Since then he has been particularly associated with automated techniques in cartography, organizing a symposium in Reston, Va., in December 1974, and serving for several years as the U.S. representative on Commission III, Automation in Cartography, International Cartographic Association. He is currently Chief, Branch of Cartography, in the Western Mapping Center of USGS.

George Lee, a civil engineer currently assigned to the Branch of Cartography, Western Mapping Center, is working on digital applications in practical cartography. He received a B.S. in engineering mathematics and an M.S. in photogrammetry from the University of California, Berkeley in 1973 and 1974. He is a member of the American Society of Photogrammetry.

ABSTRACT

The National Mapping Program includes the activities for providing basic map data and a family of general-purpose maps. The U.S. Geological Survey is developing a Digital Cartographic Data Base (DCDB) within the National Mapping Program to include such features and reference systems as the public land survey network, State and county boundaries, transportation systems, and hydrography. Although the public land survey network is administered by the Bureau of Land Management, it is included as a DCDB component because of its relationship to almost all cultural development beyond the thirteen original States. Such data included in DCDB have no legal status and are intended for display and map reference only. The interrelationship of these categories, particularly of the first three, requires care in structuring the digitized data so that valuable time is not wasted in producing cartographic spaghetti. This paper presents some ideas about structuring the components of DCDB.

INTRODUCTION

The National Mapping Program, a Geological Survey responsibility within the Department of the Interior has been established to serve the basic cartographic data needs of the United States. The program includes the activities for making basic cartographic data and a family of general-purpose base maps available to users. Most of the features shown on the maps, such as roads, buildings, relief, streams, lakes, and shorelines, are identified as base categories of map data. Other map data of public value are to be incorporated but identified as nonbase categories. The Geological Survey is responsible for the base categories of data for the National Mapping Program and will collect and make the data available in forms that contribute to timely and efficient use. In the past, the USGS cartographic data base has been exclusively graphical, in the form of topographic map sheets. However, with the ever-increasing demand for cartographic data in machine-readable form and the rapid growth in automated cartography, USGS is developing a Digital Cartographic Data Base (DCDB) as a part of the National Mapping Program.

DCDB DEVELOPMENT

The base data categories of the National Mapping Program include reference systems. Although the Public Land Survey (PLS) network is excluded from base category status, PLS data have not been ignored, particularly since the topographic base maps include a cartographic representation of PLS. The areas and boundaries defined by PLS are the responsibility of the Bureau of Land Management.

DCDB will include corner and closing-point data and monument data in addition to boundary lines and enclosed areas where the data are available. Land grant areas, such as those in Louisiana and California, and new protracted survey data will also be included as required. All data will carry codes indicating accuracy of the coordinates, agency source, and year field checked. Section-corner data will include type of monument and method of map location if available. It must be emphasized that even though USGS includes PLS data in the DCDB, the data have no legal validity, either explicit or implicit. Responsibility for legal description and position

Siggraph '77, July 20-22 San Jose, California

of section corners remains with the Bureau of Land Management. PLS data in the DCDB will be solely a digital representation of the graphically displayed network on standard topographic maps.

Another base category concerns State and county boundaries, specifically the boundaries of the 50 States and the county boundaries within each State. One approach will be to describe States and counties as areas enclosed by boundary lines and define the boundary lines as to type (State, county) and feature identification where the lines correspond to physical features, such as rivers or highways. The areas will be described by FIPS codes, and access to the data will be through the FIPS codes or geographic coordinates.

The next closely related category is the transportation network. This category could contain primary elements, such as roads, railroads, and power lines and pipelines. The type of road will need to be considered--limited access, heavy duty, medium duty, light duty, and unimproved. Several classifications for roads are in use, and it will be a goal of DCDB to identify and integrate them. The DCDB will also need to provide for the recording of single and multiple track standard-gage railroads and selected pipelines and powerlines.

The last category to be studied is surface hydrography. The category normally includes all perennial drains, intermittent drains longer than 600 meters, and perennial open water with a minimum dimension of 15 meters.

DATA STRUCTURE

The concept of the DCDB is to be an integrated data base directed not only at map production but also at direct application to problem solving. The development is beginning to focus on organization, entry, maintenance, access, and display of digital cartographic data as a user product. The potential benefits will only be realizable if the data are structured (organized) in a way that meets the system objectives. One objective of the DCDB will be to provide easy data entry, access, and retrieval. Since it is to be an active, working data base, the data must be organized so as to minimize massive searches for data access, updating, and retrieval.

In 1970, a Geographic Base File was developed by the Census Use Study, sponsored by the Bureau of the Census, using a technique called Dual Independent Map Encoding or DIME. The geographic base file, commonly referred to as GBF/DIME, is characterized by an editing capability that improves the accuracy of the files.

The principle underlying the GBF/DIME files is derived from graph theory. Each street, river, railroad, tract, municipal boundary, or other feature that bounds a census block can be considered as one or more straight line segments. Curved streets and similar features can be divided into series of straight line segments. Where streets or other features intersect or change direction, node points are identified. Figure 1 illustrates street segments and nodes.

A GBF/DIME file is built up in a street-segment routine in which each segment record contains the appropriate codes for both sides of a street between two nodes. Uniquely identifying each segment, each node, and their geographic relationships provides a description that can be checked by computer for accuracy.

USGS has tested some aspects of the DIME system in the first efforts to collect digitized PLS data. In considering topological structure, we find that the relatedness of features can be expressed in two-dimensional space by the intersections or junctions of like or unlike features. These intersections or junctions, in the DIME notation, are defined as nodes. In the context of PLS data, if each node is assigned some sort of identification and has a spatial reference, such as geographic coordinates, then a group of points forms a topological framework to which other data can be related. The PLS could then be a building block for other categories, such as State and county boundaries. Figure 2 illustrates nodes identified in a land-net. In DIME notation, the digitizing record does not merely contain information about a single feature, such as a section corner, but instead the individual features are coded with their topological relationships to other features. In the digitizing record (fig. 2), we see that the land line will be digitized as line segments. Each segment carries identifiers for a pair of nodes at the ends and a pair of polygon features for the sides. The advantage in digitizing in DIME format is that the DIME structure allows for topologically based file verification by computer.

The digitizing records are bulky. Moreover, because the file will be fragmentary, directories linking segments together into a complete PLS unit are almost as large as the file itself. Hence, in a data reduction phase, the digitizing record will be processed to form a section directory and a coordinate directory (fig. 2). These two directories will be the data base record. The data base structure of PLS will become, in fact, a point directory file, with the result of improving retrieval and area-computation operations. As a result of using DIME notation, feature data need not be digitized more than once, thus eliminating the messy process of editing coordinate pairs common to several sections.

The data structure for State and county boundaries is expected to be much like that of the PLS system. However, a chain directory must be included for handling a string of related points that define a feature from node to node; these are called chains. Figure 3 illustrates the chain identifier for a string of points between nodes. A chain therefore represents a line segment between two nodes. In general, any random line can be defined by a chain.

With points, nodes, and chains defined, the structure for transportation and hydrographic features is simple. Networks of roads and drains, which can be identified as networks, will appear in a network directory as two specific types, a branch network for hydrographic features and block networks for transportation features. The directory will permit important name identifiers, such as Interstate and State route numbers, to become descriptors for data retrieval. (See fig. 4.) Similarly, name identifiers, such as those of drainage basins and major river networks, will be descriptors for data retrieval in the surface hydrography category. Figure 5 illustrates points, nodes, and chains used to define point and line features in a drainage network.

The data base structure is summarized in fig. 6. Note that coordinate-level information of the several categories is retrieval through directories that also link the categories to each other.

It is important to understand that points, chains, and chain groups can be shared by almost any number of features. This structuring technique will enable potential users to obtain copies of the point, chain, and node data and establish their own higher order directories for special use without redigitizing the basic data.

RETRIEVAL METHODS

Digital and graphic data in the DCDB will be made available to users through the National Cartographic Information Center (NCIC). Requests will be handled by a combination of index graphics of approximate coverage over geographic areas and detailed computer data base search and retrieval routines. Identification of subsets of the data base, by both locational and feature criteria, may be derived from the DCDB.

Digital data will be collected in files with each file containing information for one cartographic unit (CU). For the present, a CU will consist of a 7.5-minute quad sheet. Although USGS has chosen that format as the CU, the DCDB will be designed with enough flexibility to accommodate other formats on retrieval. For example, if the user required PLS data in township format, then as many as four CUs may be needed to complete one township. (See fig. 7.) Through the use of directories, data can be retrieved for reformatting into townships. The processing to merge the separate data sets is aided by the false corners that appear in adjacent CUs. Similarly, the need to plot particular sections or irregular sections is simple with this data structure. Area groups, such as counties, States, and sections, established in a way similar to networks; that is, a complete closed area is identified at the chain-group level, county or land grant. These smaller units can then be aggregated into States, national parks, management areas, or whatever the user requires.

SUMMARY

The Digital Cartographic Data Base (DCDB), to be developed and maintained by the U.S. Geological Survey, will be a standardized source for base categories of digital cartographic data, principally for the United States. The DCDB will provide selected cartographic data in digital format to meet known requirements and organized to expand and evolve as users gain experience with digital data and refine their requirements. Included among base map categories will be Public Land Survey (PLS) data. Although the digital PLS have no legal status but represent in numerical forms the surveys as shown on topographic maps, they can be used for computations, statistical analyses, graphical outputs, and other user requirements. A pilot project will also include data for State and county boundaries, transportation networks, and surface hydrography.

REFERENCES

Departmental Manual, Department of the Interior, March 30, 1976.

Data Access Descriptions, 1970 Census Geography: Concepts, Products, and Programs, U.S. Department of Commerce, Bureau of the Census, DAD No. 33 (Series CG-3), August 1973.

Digital Cartographic Data Base--Preliminary Description, Office of Research and Technical Standards, Digital Applications Team, U.S. Geological Survey, Reston. July 1975.

Edson, Dean T., Lee, George Y. G., Land-line Survey Data Within a Digital Cartographic Data Base, in Proceedings of the 37th Annual Meeting, American Congress on Surveying and Mapping, March 1977.

GBF/DIME File Record

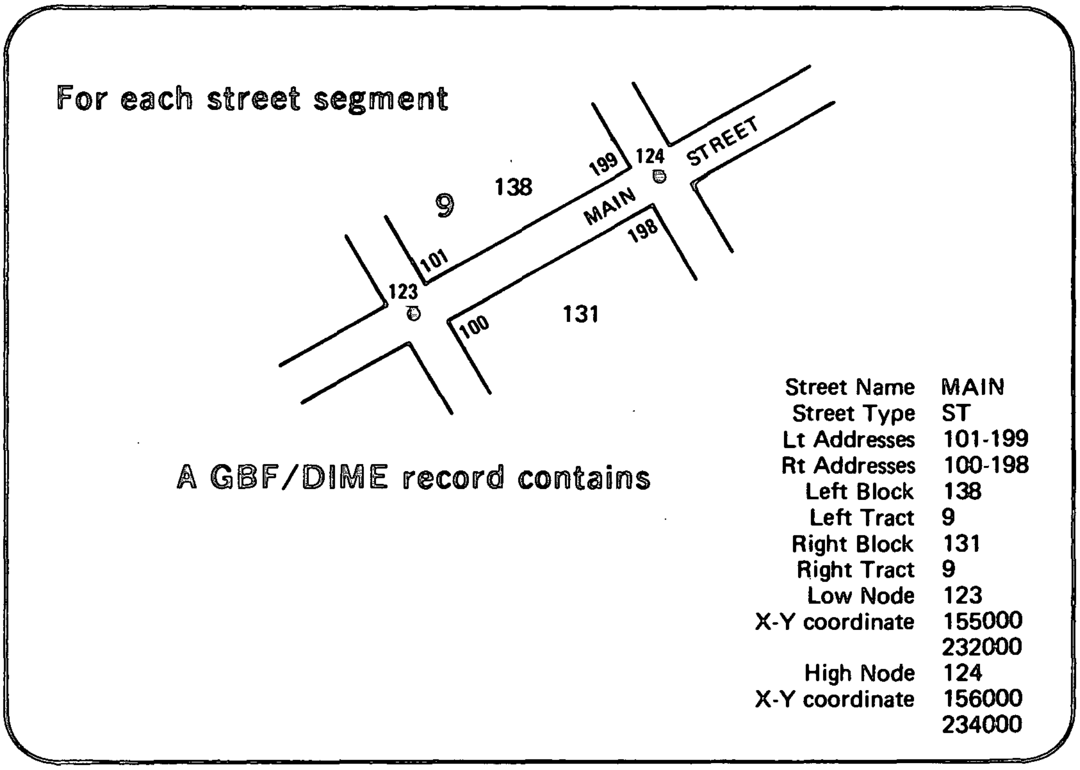
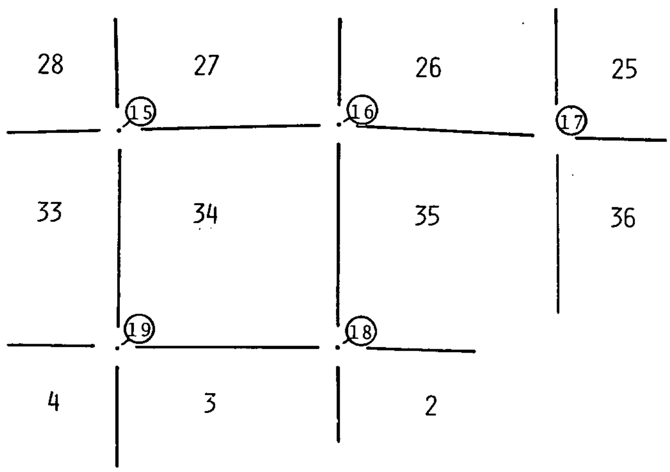


Figure 1.

RECTANGULAR SURVEY (LAND NET)



DIGITIZING RECORD

FROM NODE	LEFT	RIGHT	TO NODE
15	27	34	16
16	35	34	18

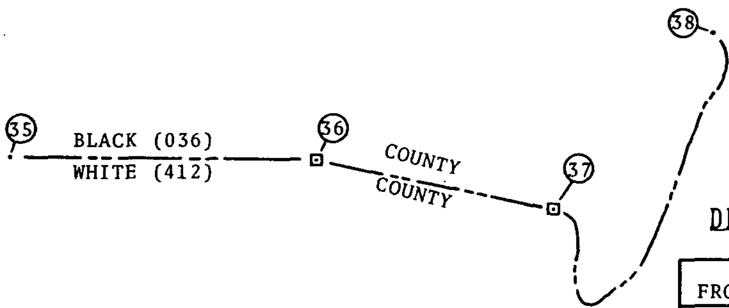
DATA BASE RECORD

SECTION DIRECTORY	
SECTION	NODES
34	15, 16, 18, 19, 15

COORINDATE DIRECTORY	
POINT	COORDINATES
15	X_{15}, Y_{15}
16	X_{16}, Y_{16}

Figure 2.

CIVIL BOUNDARIES



NAME DIRECTORY	
NAME	FIPS CODE
BLACK	036
WHITE	412
:	:

DIGITIZING RECORD

FROM NODE	LEFT	RIGHT	TO NODE
35	036	412	36
36	036	412	37
⋮	⋮	⋮	⋮

DATA BASE RECORD

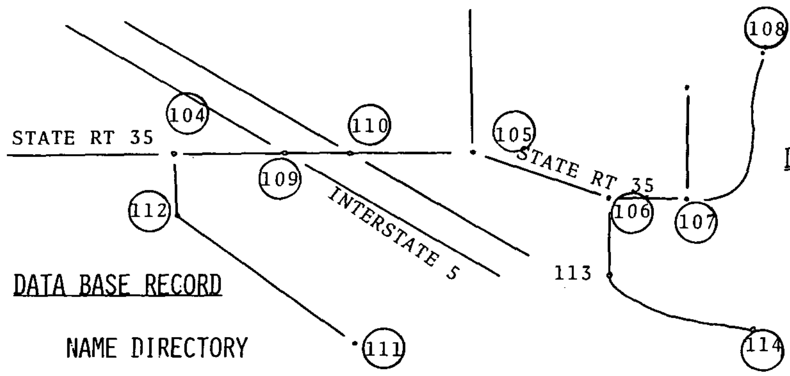
COUNTY DIRECTORY	
COUNTY CODE	NODES
412	...35,36, 37*,38*,...
⋮	⋮

COORDINATE DIRECTORY	
POINT	COORDINATE
35	x_{35}, y_{35}
36	x_{36}, y_{36}
⋮	⋮ ⋮

CHAIN DIRECTORY	
END POINTS	POINTS
37-38	1005-1185

Figure 3.

TRANSPORTATION NETWORK



NAME DIRECTORY

ROUTE NO.	QUAD. NO.
S 35	419
I 5	419, 420
⋮	

DIGITIZING RECORD

FROM NODE	TO NODE	CODE
104	105	135
105	106	135
⋮		

DATA BASE RECORD

NAME DIRECTORY

INTER-STATE ROUTES

ROUTE NUMBER	NODES
5	...110...
⋮	

STATE ROUTES

ROUTE NUMBER	NODES
35	...104, 109,110... 108
⋮	

GENERAL ROADS

CLASS	NODES
3	105,...
⋮	

COORDINATE DIRECTORY

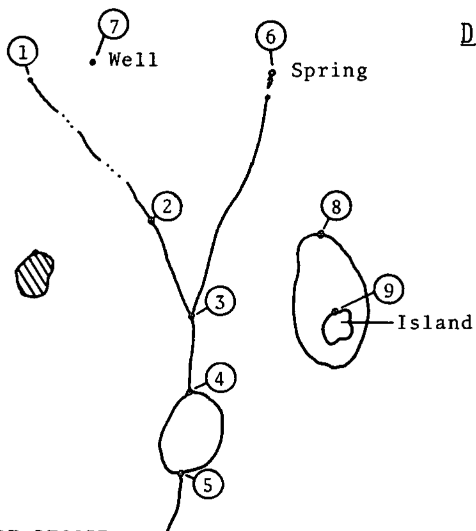
POINT	COORDINATE
104	x_{104}, y_{104}
105	x_{105}, y_{105}
⋮	

CHAIN DIRECTORY

END-POINTS	POINTS
107-108	2000-2189

Figure 4.

DRAINAGE NETWORK



DIGITIZING RECORD

POINT FEATURES & NODES	
Point/Node Number	Type
1	N
2	N
6	P
⋮	

STREAM OF COORDINATES		
FROM NODE	END POINTS	TO NODE
1	1-1001	2
2	1002-1940	3
⋮		

DATA BASE RECORD

COORDINATE DIRECTORY	
POINT	COORDINATE
1	X_1, Y_1
2	X_2, Y_2
3	X_3, Y_3
⋮	

CHAIN DIRECTORY	
END-POINTS	COORDINATES
1-2	1-1050
8-8	1051-5000

Figure 5.

DATA BASE STRUCTURE

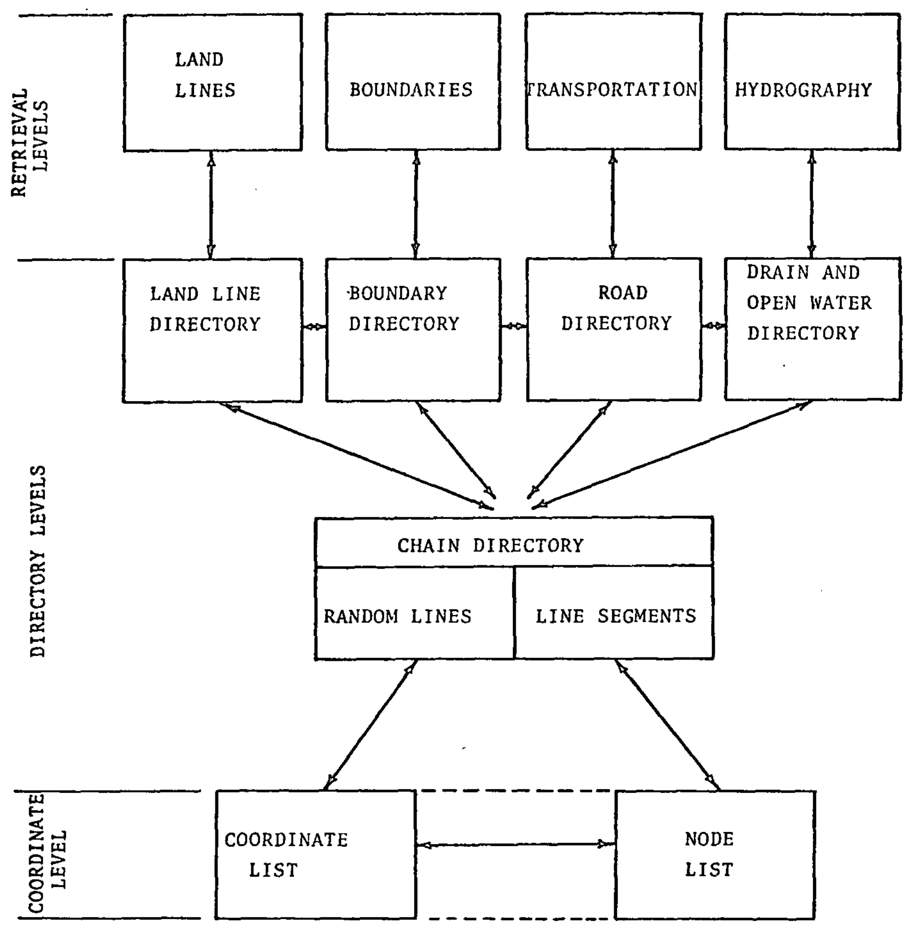


Figure 6.

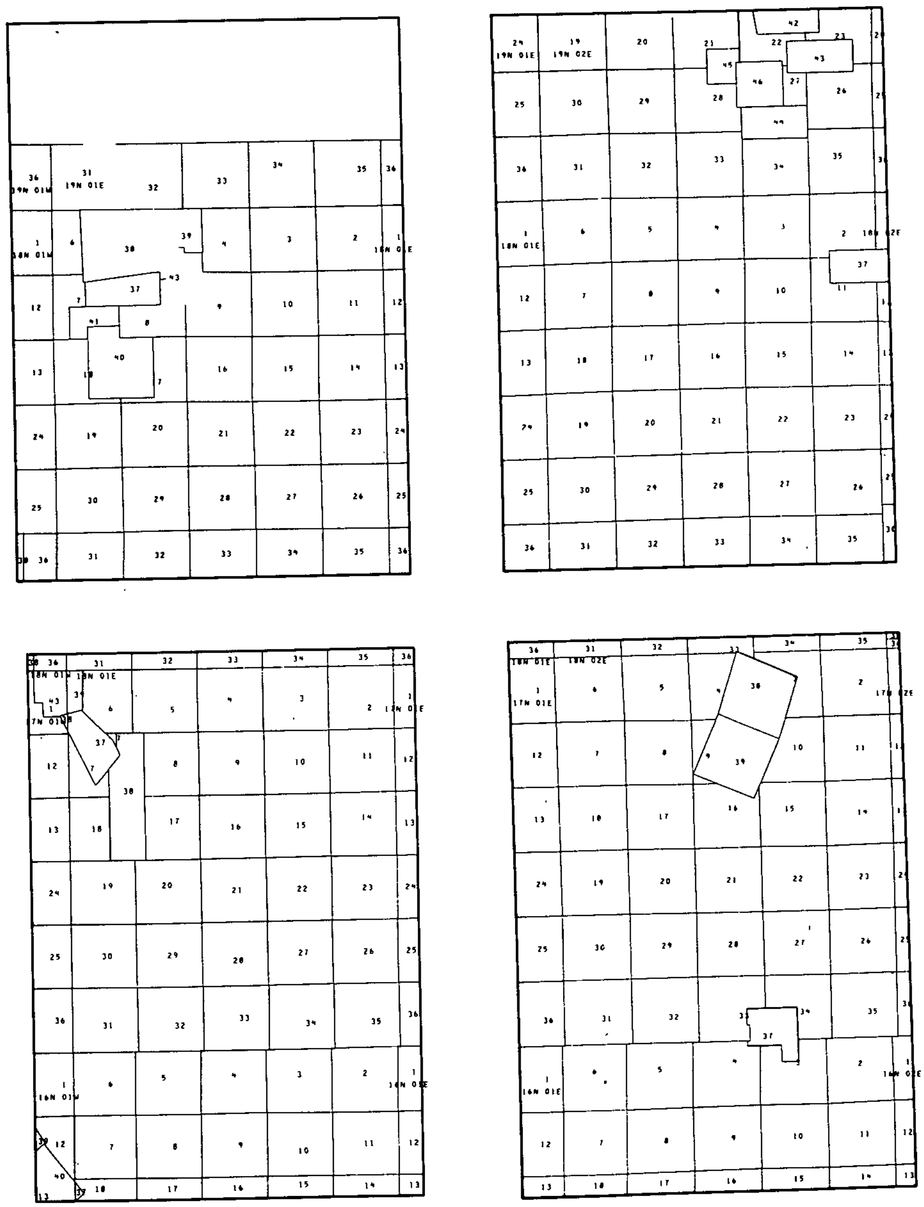


Figure 7.