GEOGRAPHIC DATA ENCODING ISSUES

by

Kenneth J. Dueker

Technical Report #43

April, 1975
(Revised June, 1975)

Institute of Urban and Regional Research
University of Iowa
Iowa City, Iowa 52240

Dr. Dueker is Director of the Institute of Urban and Regional Research and Professor of Urban and Regional Planning and Geography, University of Iowa.

Support of the Energy and Environmental Systems Division, Argonne National Laboratory (ANL) is gratefully acknowledged. ANL work performed under the auspices of the Office of Land Use and Water Planning, U.S. Department of the Interior under Interagency Agreement P-7434A.

GEOGRAPHIC DATA ENCODING ISSUES

Abstract

Means of capturing and encoding geographic data into machine-readable form is a major issue in the design and the development of statewide land use information systems. The available and potential technology constitutes a bewildering array of choices for system designers from which to select formats and processing capabilities to meet user applications. Unfortunately, the data use/application areas are not well specified; consequently designers are not provided a very precise statement of data requirements (or fidelity requirements for the data capture and processing system). Therefore, the tolerable distortion rates and tolerable information losses are not well expressed so as to assess coding efficiencies. Nor is there substantial agreement as to the appropriate breadth of comparative tests, as some systems can replicate coverages well for cartographic purposes, while other systems may provide stronger analytical capabilities for processing encoded geographic data.

Although the debate continues among technicians as to relative merits of grid versus polygon encoding, it appears that systems that have both capabilities are emerging. Grid encoding is often favored for planning-oriented systems to simplify overlaying coverages, although choice of cell size is crucial; once made accuracy and data volume are set. Polygon encoding can more accurately represent coverages, but the processing requirements are greater. Unfortunately, polygon encoding technology does not handle well, errors in capture and encoding of geographic data. Source document errors, digitizing errors, and logical errors pose considerable problems for fully-automated systems and are leading system designers to man-machine systems to purge files of errors.

One of the major problems in encoding geographic data is the lack of measures by which to assess the effectiveness of the coding. One set of effectiveness measures relates to the ability to replicate the source document in map form, while a second set of effectiveness measures relates to the use of the map or coverage data. There are a number of effectiveness measures that need developing in order to compare and test the effectiveness of alternative system approaches. A source document constructed to possess features that will test the effectiveness of alternative systems fairly and equitably needs development. This would result in the ability to establish benchmark tests by which systems could be compared. In this context, the question is whether one benchmark test could be constructed for both digitizer and scanner based data capture systems, and whether there should be separate benchmark tests for source documents, for encoding errors, and for logical errors.

Until some of these issues are resolved, system designers should caution planners of statewide land use information systems of the potential for errors, delays, and cost overruns when attempting to encode and replicate a large number of complex coverages. Presently, coarse polygon or large grid based systems are more appropriate for most states, while at the same time undertaking prototype developments in smaller study areas to test more sophisticated encoding techniques and to develop staff skills.

INTRODUCTION

Capture and encoding of geographic data to a machine-readable form is a major issue in the design and development of statewide land use information systems. Data capture technology -- manual coding, digitizing, scanning -- and formats -- pixels, cells, grid units, points, line segments, or polygons -- constitute a bewildering array of choices for system designers. The system designer is faced with selecting data capture technology, a format that provides a capability to meet user applications, and a data processing system that is commensurate with the choice of data format and volume of data the format and region size imply. Extending applications/uses require data formats that more closely capture the fidelity of source data. Yet, some degree of spatial aggregation of geographic data is necessary. There is no easy choice with respect to fidelity requirements for data capture or the appropriate level of aggregation of data. Too little is known as to the data requirements for applications for statewide land use planning and the data capture and manipulation technology for large-scale applications.

A region consists of spatially varying sets of characteristics. In this paper, each set is considered a coverage, say soils, and each coverage is categorized, say into soil classes. In other words, geographic data is considered to consist of various areas of like characteristics separated by networks of lines. A single such partitioning of a region into non-overlapping zones will be referred to as a coverage (Goodchild, 1975, page 2). For example, there may be coverages showing soil characteristics, land uses, vegetation cover type, political division, or combinations of these. Linear data, such as streams, roadways, railroads can also be considered a coverage, but with the emphasis on the network of lines rather than the bounded areas. This paper is concerned only with coverages as two dimensional objects, thus excluding pictorial representation of the three spatial dimensions as well as time-varying pictorial information, e.g., on-line character recognition. This restriction also rules out picture processing, computer generated movies, and computer typography. Finally this paper treats the subject of coverages from a primarily problem oriented rather than technique oriented standpoint, in that the emphasis is on the relationship between encoding coverages and applications, rather than hardware/software techniques for encoding coverages.

The first section of this paper attempts to identify encoding coverages and discusses geocoding options. Next the paper deals with the degree of data aggregation, system requirements, and analytical tasks as a function of geocoding options and data application/use. Then the data capture, formatting, storage and output of coverages is related by analogy to information theory for purposes of illuminating parallels between encoding spatial data and encoding messages for transmission, receiving and use. This analogy proves useful in that it identifies the dilemma of system designers, in that fidelity requirements for the spatial processing have not been developed and consequently effectiveness measures for the encoding of coverage data cannot be specified. Next, the paper attempts to identify the potential for error in data capture so that system designers can be alert to situations that can occur and make allowances for remedying these errors. Then, a discussion of test coverages is presented that provides a basis for testing encoding and processing systems. Typical data encoding errors and other built-in features are incorporated into the test coverages that enable application of effectiveness measures and means to compare different systems. Finally, the paper calls for the development of comparative benchmark tests that agencies could employ to evaluate vendor systems.

ENCODING COVERAGES

As indicated in the introduction, this paper focuses on encoding geographic data consisting of coverages, i.e., areas of like characteristics separated by networks of lines. A bounded area within a coverage is referred to as a polygon, a face, or a map segment.

A coverage can be encoded as a one-dimensional representation in which the basic record is a single contiguous homogeneous tract or "polygon" coded by locating the boundary as a series of connecting points and by indicating the character or descriptor of the enclosed territory. Ignoring any error in the drawing of boundaries or in ascribing characteristics, the accuracy of polygon encoding is limited only by the precision with which boundaries can be coded as series of connective points. One disadvantage of the polygon method occurs when two coverages have to be compared. With polygon data it is time consuming to identify the polygon in which a certain point lies, and thus its characteristics, and to compare the same point on another coverage. The advantages of organizing the data so that the character of any location can be retrieved quickly, often leads to gridding polygon data prior to overlaying or the initial adoption of a grid data structure. In a grid structure the coverage is encoded by recording the nature of each of a series of cells ordered in some conventional sequence. Compression of grid data structure is possible to eliminate a data set containing a separate entry for each cell by means of using some form of multiplier convention for sequences of repetitious cells.

With grid data it is a simple matter to compare the characteristics of a point on two coverages. The most troublesome characteristic of grid encoding is its approximation of coverages. Accuracy is directly linked to the size of the grid cell and precise replication of a coverage requires a large number of infinitesimally small cells.

Both polygon and grid cell encoding are widely used. The polygon form is adopted by systems whose major concern is data storage and accurate cartographic retrieval, and the measurement of area. The grid system is more widely used in various forms of planning, but where accuracy is less important and the ability to overlay coverages is essential, and where the range of likely demands on the system is perhaps much better defined. (Goodchild 1974b, pp. 3 and 4.)*

Table 1 displays the uses, encoding methods, advantages, and disadvantages of various geocoding options.

ISSUE SET #1 GIVEN THAT INFORMATION SYSTEMS OR STATEWIDE LAND RESOURCE PLANNING HAS BOTH CARTOGRAPHIC AND OVERLAY REQUIREMENTS WHICH FORM OF ENCODING IS MOST APPROPRIATE?

CAN BOTH POLYGON ENCODING AND GRID ENCODING BE INCORPORATED INTO A SINGLE SYSTEM?

WHAT CONVERSION CAPABILITIES BETWEEN POLYGON ENCODING IS NECESSARY FOR VARIOUS APPLICATIONS?

Exploration of these issues can begin by:

COMPARING THE COST-EFFECTIVENESS OF DIRECT POLYGON ENCODING (DOUBLE DIGITIZING LINE SEGMENTS BETWEEN JUNCTIONS) WITH LINE SEGMENT AND DESCRIPTOR ENCODINGS AND THEN GENERATING POLYGON RECORDS.

---

*The terminology in this section draws heavily from Goodchild. Goodchild goes on to examine two specific problems; the first being overlaying of polygon data, which because of the inaccuracy of boundary locations creates slivers or spurious polygons. The second problem is in analysis of grid cell size or its relationship to accuracy in grid manipulation.

Table 1

CHARACTERISTICS OF GEOCODING OPTIONS

Geocoding Options

	Grid			X-Y Coordinates			
	Predominant	Qualitative	Area Meas.	Points	Line Segment	Simple Polygons (few vertices)	Complex Polygons (many vertices)
Use	Small cells, predominant use	Presence or absence of features	Distribution of uses	Point location, area centroid, sample points	Linear data, e.g., streams, roads type of network of areas, e.g., dual encoding of lines and areas	Regular areas, e.g., jurisdictional boundaries, coarse coverages	Irregular areas, e.g., soil, cover, slope complex coverages
Encoding Method	Scanner	Record primary, secondary, and tertiary occurrences	Manual grid, grid of polygon data	Point digitizing	Point or stream digitizing with or without area coding	Point digitizing	Stream digitizing automatic line following
Advantages	One category per cell	Encode all data once for each cell		Minimal encoding	Replication of linear data for cartographic purposes	Explicit encoding of area data	Explicit encoding of complex area data
Disadvantages	Separate encoding for each coverage many cells	Exact coincidence of categories not encoded		Boundaries not encoded	Line segments need additional processing to generate polygons	Must process polygon data to ensure exact matching of adjacent polygons	

COMPARING THE COST-EFFECTIVENESS OF OVERLAYING POLYGONS WITH GRIDDING POLYGONS PRIOR TO OVERLAYING.

Selection of a geocoding option is dependent upon the type of application. The next section discusses interrelationships of geocoding options and applications with other design options.

DATA AGGREGATION, SYSTEM REQUIREMENTS, AND ANALYTICAL TASKS

The level of data aggregation, system requirements, and analytical tasks are a function of: 1) data use/application (within the general area of statewide land use planning), and 2) geocoding options. The system designer must select an appropriate level of abstraction in converting a map to a coverage, and then again in converting a coverage to digital form, both as a means of reducing the sheer data volume. Level of data aggregation and geocoding options should be determined by the intended applications. Developing these interrelationships is crucial to the design of systems having multiple uses, but at the same time it is difficult to relate the designers' options to abstract data use/application categories. Yet, systems must be designed for classes of problems, not specific applications.

Although most readers will find fault and be unsatisfied with this section, the attempt to define interrelationships should not be viewed as definitive but as a first approximation. The essential point of this section is to speculate on geocoding options, system requirements, aggregation levels, and analytical tasks, necessary for varying classes of problems or applications that are emerging for consideration in statewide land use planning. The chief difficulty is that each problem class or application type may be approached from a variety of levels of sophistication and detail. Spatially varying detail makes it difficult to define information system requirements for classes of applications, although it is essential that decomposition of classes of applications continue so as to gain specificity in assessing system, geocoding and analytical requirements.

Data aggregation has two components, aggregation of phenomena to categories and spatial aggregation. The level of detail for coverage categories and spatial units should be compatible. A large number of categories, for say cover or soils generates a complex coverage, which if encoded to large spatial units, such as LUNR's (Shelton, 1973) one kilometer grid imposes a high degree of spatial aggregation. Similarly a coarser coverage classification such as MLMIS (Land Management ... , 1972) utilized finer spatial units (40 acre) for predominant use assignment. Consistency of coverage categories with spatial units is crucial to selecting a geocoding option which meets intended data uses/applications.

It is essential to recognize that comparison of systems is made difficult if they have been designed for different applications. For example, a system designed principally for cartographic uses may more exactly replicate coverages, but may be incapable of extended applications, while a system less capable of replicating coverages may have more flexible geocoding structures that enable more powerful analytical processing, such as overlaying or generating slope maps.

It is particularly difficult to categorize data uses/applications for a still emerging statewide land use planning process. On one hand a narrow interpretation dealing with critical area delineation, focusing on data categories and spatial units could be explored. On the other hand a broader interpretation of statewide land use planning is deemed more fruitful in the long run. Data use/application, for purposes of this discussion, are categorized below. Each category is illustrated with an example of types of analysis.

1. Policy Planning, state or county level projections of population and employment with varying assumptions regarding birth rate, investment, consumption, exports, unemployment, labor force participation, etc., reflecting policy alternatives to assess alternative growth and development strategies for a state.

2. Program Planning, locational analysis of public services, that state or substate regional agencies are to coordinate, such as recreation sites, health services, detention centers, educational facilities, etc., and service area delineations, such as redistricting.

3. Land Inventory, inventory of cover type, land use, water sheds, type of land ownership, governmental jurisdiction, soil type, terrain to produce maps at 1:500,000 and 1:250,000, and area measurements tabulated to town/townships.

4. Impact Assessment, environmental impact assessment of key facilities and large-scale developments.

5. Land Capability, initial screening for critical area analysis, overlay inventory data for small areas to determine where critical resources coincide and/or are within areas subject to developmental pressures. Similarly site suitability analysis overlaying inventory data to determine locations possessing desired characteristics.

6. Regulation or control of land, delineating extent of flood plains, shorelines, with respect to property lines. Regulation of land requires access to a land records system that can be linked to land inventory data and socio-economic data.

These six categories are intended to include a full range of applications with respect to land use analysis. These applications can take place at a statewide, regional, or local level, although the implication is that policy planning would be more at regional or state levels while land regulation traditionally lies more at the local level. Each of these uses of data has differing demands for coverages, geocoding options, system requirements, level of aggregation, and analytical tasks. While it is recognized that manual procedures can be employed for many of these applications, this discussion assumes coverage data in machine-readable form for the purpose of identifying data and system requirements for the different applications.

Depending upon the data use/application coverages of different detail will be required and consequently, geocoding options must be considered in capturing these coverages. The geocoding options considered here are (in increasing levels of detail):

1. External index systems
    a. place name
    b. place code

2. Coordinate index systems
    a. grid
        1. predominant use
        2. qualitative
        3. area measurement
    b. points
        1. location of activities or events
        2. area centroid
        3. sample points for spatial distributions
    c. lines
        1. linear features
        2. boundary or flow networks
    d. coarse polygons
    e. fine polygons

External index systems allow handling a few large areas in conjunction with maps, while the coordinate index systems enables capturing more complex coverages and facilitates their processing and compositing.

As one would expect, more complex coverages and geocoding options require more sophisticated systems for processing. Consequently, system requirements (for increasing levels of detail) are categorized as:

1. General purpose computer,
   card and tape files,
   general data management software,
   no specialized peripheral equipment;

2. General purpose computer,
   digitizer, plotter,
   specialized software for data capture;

3. Dedicated computer with digitizer, CRT, plotter,
   specialized software for data capture.

Different data use/application needs require the aggregation of both social economic data and physical data to common levels of aggregation. Common levels of aggregation or land use planning and analysis are:

1. Arbitrary Areas
    a. State
    b. County
    c. Minor Civil Division/Tract
    d. Enumeration District/Block Group

2. Grid
    a. Sq. Mile/Sq. Kilometer
    b. 40 acre/9 hectare
    c. One acre/one hectare

3. Polygon
    a. Natural areas
    b. Ownership parcels.

Finally, different data use/application needs require various analytical tasks. These are defined as follows:

1. Trend projection models
2. Optimal location/allocation models
3. Spatial interaction models
4. Spatial association, measurement and display techniques
5. Diffusion models
6. Record keeping and monitoring .

In terms of designing geographic information systems, the essential analytical task is Spatial Association, Measurement, and Display Techniques. The techniques consist of:

1. Area measurement
2. Overlay
3. Tabulation
4. Cross-tabulation
5. Functional analysis of variate relationships
6. Display - maps, graphs, reports.

These techniques provide the basis for processing land use data; which when used in conjunction with other analytical tasks, provide a system to support a multifaceted planning process.

Interrelationships between level of data aggregation, system requirements, and analytical tasks as it relates to data use/application and geocoding options is illustrated in Table 2. Table 2 posits a mix of these interrelated elements, although a reader should not take too literally the match-ups, as the process of specifying design options for classes of applications is extremely hazardous and subject to extensive debate at this time. Although the matching is speculative at this time, the attempt is made to help identify that the range of options and needs probably precludes the design of a single system with data at a single level of aggregation that meets multiple uses.

Interpretations from Table 2 should be made with caution. At the present time the following limited interpretations are offered:

1. The choice of a geocoding option determines the requirements for a computer system. External index or grid geocoding options only requires a general purpose computer without peripheral equipment, while x-y coordinate geocoding requires peripheral equipment for digitizing and plotting, and fine polygon geocoding will likely require a dedicated system for on-line editing when encoding map data.

2. Each data use/application category of planning problems has requirements for varying levels of detail depending on the specific application. Land inventory, impact assessment, and land capability application can be approached from various options for spatial detail and geocoding options. Consequently, this discussion is not able to recommend design choices for these data use/applications. More research and development is needed to decompose applications to discern data requirements and then identify common applications which justify common data and systems.

3. The specific choice of design option combinations in Table 2 assume the type of problems for each data use/application category that was described when the categories were introduced, earlier in this section of the paper. It should be recognized that specific problems might deviate from the combination of design options specified in a given cell of the table.

4. Empty cells in a row indicate combinations of design options that are less appropriate for that class of application.

Table 2

LEVEL OF DATA AGGREGATION, SYSTEM REQUIREMENTS AND ANALYTICAL TASKS AS A FUNCTION OF DATA USE/APPLICATION, AND GEOCODING OPTIONS

	Geocoding Options									
	External Index	Grid			X,Y Coordinate					
	Place Name or Code	Predominant	Qualitative	Area Measurement	Points	Line Segment	Coarse Polygon (few vertices)	Grid	Fine Polygon (many vertices) Polygon	Network
Policy Planning	A0 S1,A1,M1				S1,A1,M3					
Program Planning					A1,M2 S1,A2,M3 A3,					
Land Inventory		S1,A5,M4	———	S1,A4,M4	S2,A7,M4	S2,A7,M4	S2,A7,M4		S3,A7,M4 A8,	
Impact Assessment	S1,A2,M4 M5	———	S1,A4,M4 M5	———	S2,A2,M4 A3,M5	S2,A7,M4 M5			S3,A7,M4 A8,M5	
Land Capability		———	S2,A5,M4	———		S2,A7,M4			S3,A7,M4	
Regulation										S3,A7,M4 A8,M6

Key:

SYSTEM REQUIREMENTS

- S1 General purpose computer
  Card and tape files
  General data management software
  No specialized peripheral equipment
- S2 General purpose computer
  Digitizer, plotter
  Specialized software for data capture
- S3 Dedicated computer with digitizer, CRT, plotter
  Specialized software for data capture

LEVEL OF AGGREGATION

A0	State
A1	County
A2	Minor Civil Division/Tract
A3	Enumeration District/Block Group
A4	Square mile/square kilometer
A5	40 acre/9 hectare
A6	one acre/one hectare
A7	natural areas
A8	ownership parcel

ANALYTICAL TASKS

M1	Trend projection models
M2	Optimal location/allocation models
M3	Spatial interaction models
M4	Spatial association, measurement, and display techniques
M5	Diffusion models
M6	Record keeping and monitoring systems

Inspection of Table 2 does allow some tentative conclusions. These are:

1. Socio-economic data aggregated to the Minor Civil Division/Tract or higher level can be employed for many sectoral planning needs at the statewide level without a need for detailed physical data.

2. Initial land inventory and monitoring needs, may be met at coarse grid or polygon levels of aggregation.

3. Critical area designation data needs require finer grid and/or polygon encoding.

4. The data needs for impact assessment, land capability, and regulation of land are more difficult to determine at this time. The extent to which a common system can meet those needs, and whether the technology is available for fine grain systems needs further investigation. This implies the development of pilot systems for each of the last three data use/application areas to determine whether a common system might be developed and at what level of governmental jurisdiction; because it is at this end of the scale of analysis that statewide systems must interface closely with regional and local systems.

This discussion of interrelationships implies a number of issues that warrant exploration.

ISSUE SET #2 IDENTIFY SYSTEM BENEFITS AND COSTS FOR EACH CATEGORY OF DATA USE/APPLICATION.

1. ESTIMATE FREQUENCY AND IMPORTANCE OF DATA USE/APPLICATION CATEGORIES.

2. ESTIMATE COST OF SYSTEMS, FOR DEVELOPMENT, DATA COLLECTION, IMPLEMENTATION, AND MAINTENANCE.

CLARIFY THE COMBINATION OF SYSTEM REQUIREMENTS, LEVEL OF AGGREGATION, ANALYTICAL TASKS, AND GEOCODING OPTIONS APPROPRIATE FOR EACH DATA USE/APPLICATION.

IDENTIFY THE COMMONALITIES BETWEEN SYSTEMS SERVING DIFFERENT DATA USE/APPLICATION THAT WOULD ALLOW DESIGN OF A SINGLE SYSTEM SERVING SEVERAL DATA USE/APPLICATION CATEGORIES.

IDENTIFY THE STATE, REGIONAL, AND LOCAL INTERRELATIONSHIPS AND RESPONSIBILITIES FOR SYSTEM ELEMENTS.

Exploration of these issues can begin by:

RESEARCH SHOWING THE EFFECT OF SPATIAL AGGREGATION (GRID SIZE) UPON DATA USE/APPLICATION AREAS, AND THE RELATIONSHIPS BETWEEN SPATIAL AGGREGATION AND NUMBER OF DATA CATEGORIES FOR A COVERAGE.

MORE CLEARLY DEFINE THE APPLICATIONS POTENTIAL FOR EACH GEOCODING OPTION. DEVISE COMPARATIVE TESTS FOR APPLICATIONS PROBLEMS.

CONSTRUCTION OF COST MODELS FOR GEOCODING OPTIONS, e.g., FOR DIGITIZING:

Cost (Polygon) = f (No. of map mountings, No. of polygons, No. of points per polygon)

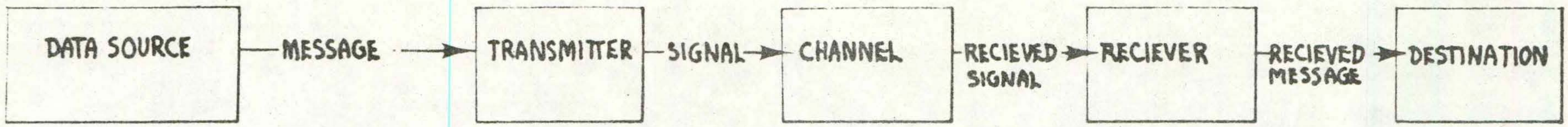
Cost (Point) = f (No. of map mountings, No. of points)

Cost (Line) = f (No. of map mountings, No. of lines, No. of points per line).

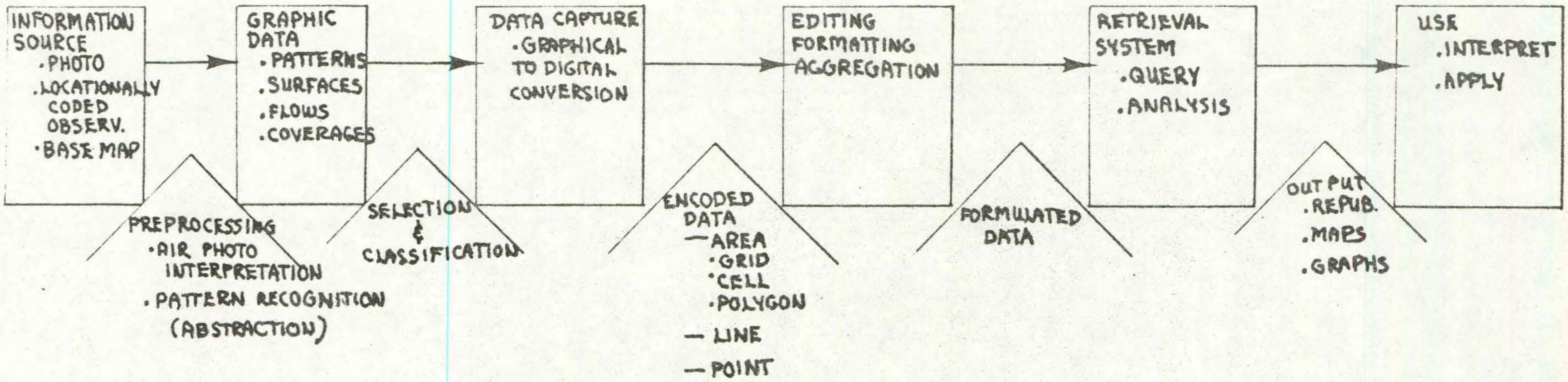
A COMMUNICATIONS SYSTEM FOR GEOGRAPHIC DATA

A generalized communication system is comprised of elements: a data source from which a message is encoded, a transmitter from which a signal is emitted, a channel for communicating the signal, a receiver for receiving the signal and converting it back to a message to the final destination. Figure 1(a) provides a diagram of this generalized communication system. This problem of sending and receiving messages through the use of the system which is constrained by channel capacity and the presence of perturbances (noise, and and distortion) is the general case from which this discussion of encoding geographic data is an application.

A geographic information system also has elements of a sender, receiver, message, signal and channel. The problems of channel capacity, and transmission cost may be considered as analogous to computer storage size and machine processing cost. Of particular interest in this paper is the encoding problem. In the generalized communication system, information theory is used to measure the amount of information (in units called "bits") that is contained in the data being processed and this theory aids in the evaluation of alternative encoding schemes to eliminate redundancy through efficient coding. Information theory is employed here to consider means of reducing redundancy in the storage



a) A Generalized Communications System



b) A Communications System for Geographical Data

Figure 1. A Communications System

and processing of geographic data. Figure 1(b) illustrates the adaptation of the generalized communication system model to geographic data which demonstrates that the terminology, or organization concepts for information theory is useful in the analysis of geographic data handling problems. However, the main utility of information theory is in the area of efficiency in encoding data.

In applying information theory, a digital coverage consisting of a quantized arbitrary matrix is considered a set of messages. The gray level of intensity for each cell is a "message." If there are m gray levels or intensity categories, the total amount of information in a n-by-n digital coverage (which is average amount per element times the number of elements) can be as high as $n^2 \log_2 m$ bits. The actual information content depends upon the probabilities with which the gray levels occur. Physical pictorial media can be used to store information at extremely high densities. However, pictures encountered in practice (television images, line drawings, printed pages, etc.) have information content that fall appreciably short of their potential capacities, often by a factor of two or more. The difference between potential and actual information content is called redundancy. (Rosenfeld, 1969, page 9.) Efficient encoding of pictures or coverages is possible if redundancy is minimized, and there has been considerable effort made in devising coding schemes to represent a picture or coverage as compactly as possible.

This process of approximating the picture acceptably (where the standards of acceptably may be either objective or subjective) by another picture that has lower information content has been directed largely toward the goal of television bandwidth compression. However, the manual abstraction of coverages from pictures and other sources provides an initial abstraction of the information from which further approximation is possible. The object is to reduce redundancy and not content.

The process of creating coverages from images or pictures is the first step in efficient coding of pictorial information. Cells within a single map segment of a coverage by definition have the same message or value. Hence, within a map segment a message of the next cell is a predetermined value rather than a probability of that element being the same as the preceding element message. Thus, the source document is an abstraction of reality, which when encoded enables the reduction of redundancy. Consequently, encoding of coverages reduces to a case of coding long "runs" of repeated messages. Therefore, it is then economical to encode the first message of each run and then the length of the run rather than encoding each message or cell in a sequence, and all of the detail that appears in a picture is replaced by a simpler coverage that looks like the original but that has a lower redundancy. The degree of "compression" that can be obtained by approximation methods for generating coverages is generally greater than that obtainable by encoding techniques alone.

There are two basic methods used to approximate pictures; these are sampling and quantization. Sampling consists of taking values at a finite set of points, and approximates the surface by interpolating analytically simple functions through these values. In quantization, one allows the function or picture to take on only a finite set of values or quantization levels, (replacing the actual value at each point by the quantization level closest to it).

In approximating a function or surface from sampling methods one can sample from equally spaced points or a rectangular array, though it is sometimes desirable or necessary to use unequally spaced points. Contouring routines such as SYMAP or trend surface analysis, using polynomial interpolation, can be used to approximate the value at any point in an n-by-n matrix representing an n-by-n digital picture.*

Figure 2 represents efficiencies in coding geographic data. Figure 2(a) represents a n-by-n matrix of fine cells or pixels with m gray tone levels. With no predictable pattern the information content is $n^2 \log_2 m$ bits. A complex urban land use scene would have a lower level of randomness or entropy and a rural land use scene would even be more orderly or less complex a pattern. Regular sampling to approximate the image implies a larger cell size (see Figure 2(b) where $n > n'$) and quantization (where $m > m'$), imposes a further ordering to the image.

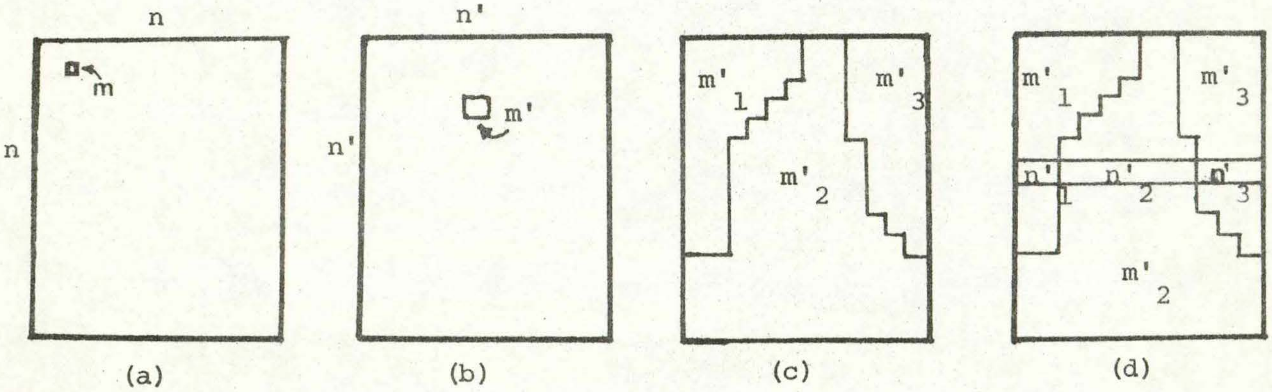
Figure 2(c) suggests boundaries are drawn around contiguous sample cells of like category, thereby creating the coverage. Figure 2(d) represents one type of encoding where row i consists of n'_1 columns of m'_1, n'_2 columns of m'_2 and n'_3 columns of m'_3. In the case of (b) the information loss can be estimated and in (d) the coding efficiency can be estimated.**

The utility of information theory is that it forces system designers to ask important questions about information loss, redundancy, and coding efficiencies.

---

[*] Most grid systems record the predominant use, primary and secondary categories, or actually measure the area of each quantization level within the grid. Nevertheless, grid units are a means of spatial sampling and only the attributes of the sample points differ.

[**] Run length coding, as this row record with a multiple for sequences of repetitious cells is called, is only one type of encoding. Polygon or line segment encodings are often used. The run length coding illustrates best for the purpose of this discussion, the potential for compression or efficiency of encoding.



(a) (b) (c) (d)

Figure 2. Information Content and Efficiency of Coding

ISSUE SET #3 WHAT ARE THE FIDELITY REQUIREMENTS OF THE RECEIVED MESSAGE AT THE DESTINATION?

WHAT DISTORTION RATE CAN BE TOLERATED?

WHAT ABSTRACTION IN PREPARING COVERAGES IS TOLERABLE?

WHAT EFFICIENCY IN CODING CAN BE ACHIEVED?

Exploration of these issues can begin by:

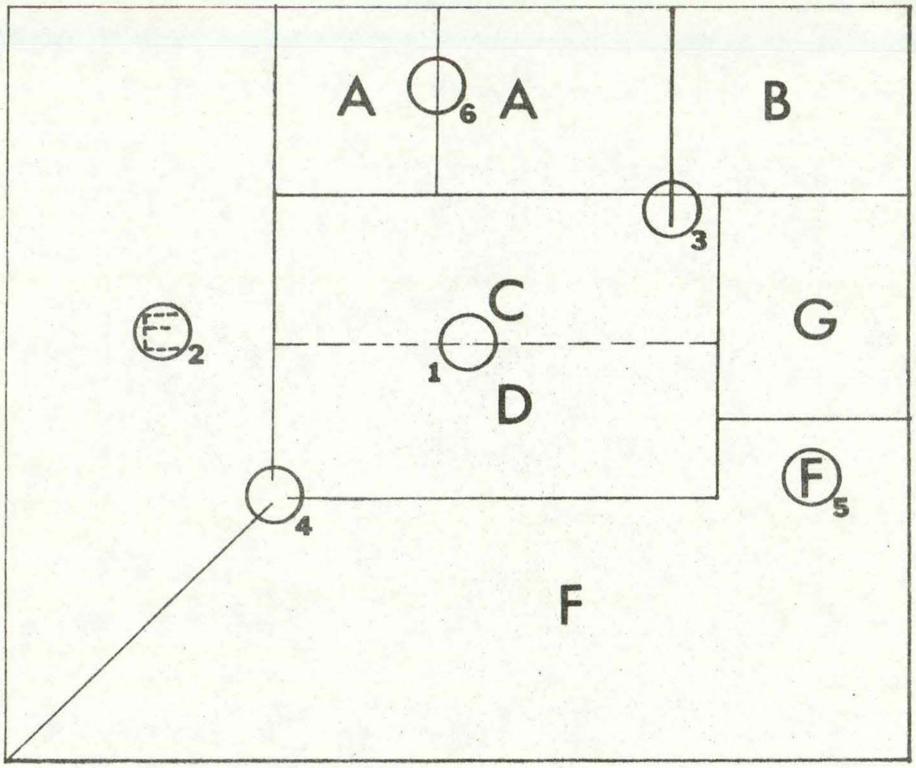
EXPERIMENTS TO MEASURE THE INFORMATION CONTENT OF ORIGINAL MAPS AND PICTURES, THE INFORMATION LOSS IN GENERATING COVERAGES, AND THE CODING EFFICIENCY OF CAPTURING DATA IN GRID AND POLYGON DATA STRUCTURES

Although the efficiency of coding source documents is of major importance, another important issue is to select encoding schemes which minimize the potential for error. The next section deals with errors in data capture.

ERRORS IN CAPTURE AND ENCODING

The capture, encoding and processing of coverages provides considerable "opportunity" for the introduction of error. Although source document errors are not within the scope of this discussion several of the more basic source document error will be identified. Most attention is given to the encoding errors that occur in the capture of coverages. Finally, some attention is given to logical errors associated with the processing of data, particularly that processing associated with editing the file to insure completeness and error detection.

Figure 3 identifies six common source document errors that should be removed prior to data capture, but which if not, should be detected for correction during or following digitizing. The errors discussed here are confined to those that affect processing, and do not include compilation errors, that might place a line erroneously, but does include shrinkage of the source document material.



1. Missing line segment
2. Missing center identifier
3. Drafting error - overshoot -
4. Drafting error - undershoot
5. Redundant center identifier
6. Redundant line segment

Figure 3. Source Document Errors

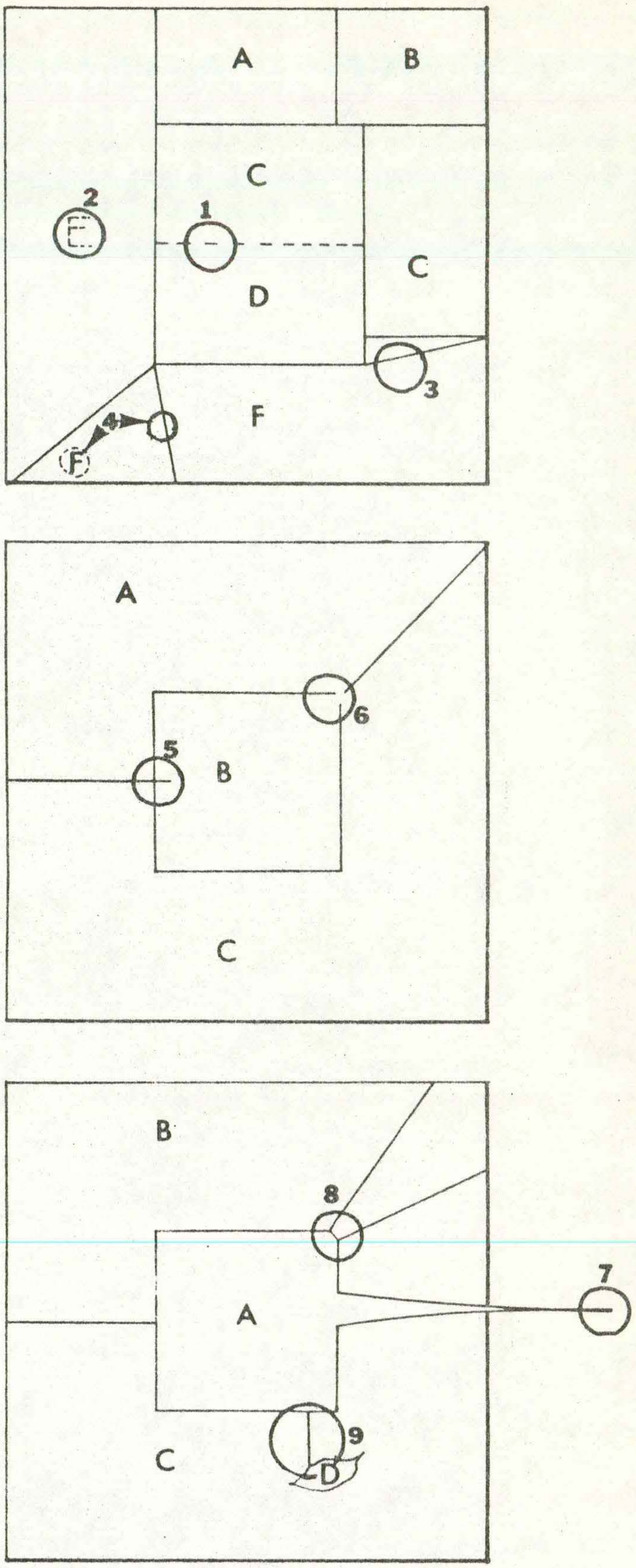
The first source document error is a missing line segment which is detectable by noting that there are two center identifiers within the same polygon. On the other hand, this could be a redundant center identifier. Source document error type two is a missing center identifier which is detectable by identifying polygons without a center identifier. Source document error type three is a drafting error consisting of an overshoot or stub where a line extends beyond its junction with another line and into a polygon. Source document error type four is again a drafting error but this time consisting of an undershoot or gap between line segments that should join. Source document error type five is a redundant center identifier and error type six is a redundant line segment.

Although it would be desirable to assume error free source documents, it is unlikely that error free source documents can be achieved. Consequently, the data capture and processing system must be capable of detecting and identifying these errors for subsequent remedy.

Figure 4 identifies various kinds of encoding errors which can occur in spite of error free source documents. Essentially the same kind of errors are possible, although in this case they are a result of faulty encoding rather than faulty source documents. Encoding error type one is a failure to encode a line segment; type two, a failure to encode a center; type three, a failure to end a line segment; and type four, a redundant encoding of a center, or a segment.

Encoding errors, type five and six, consist of the overshoot and undershoot problem resulting from difficulties of digitizing line segments to meet at a common junction point.

Encoding errors specifically with digitizing are represented by encoding



(1) Failure to encode a line segment,
(2) A center,
(3) Failure to end line segment
(4) Redundant encoding of center or segment

(5) Overshoot,
(6) Undershoot at junction points

(7) Digitizer error,
(8) Superfluous polygon because of narrow isthmus
(9) Failure to identify contained and containing polygons.

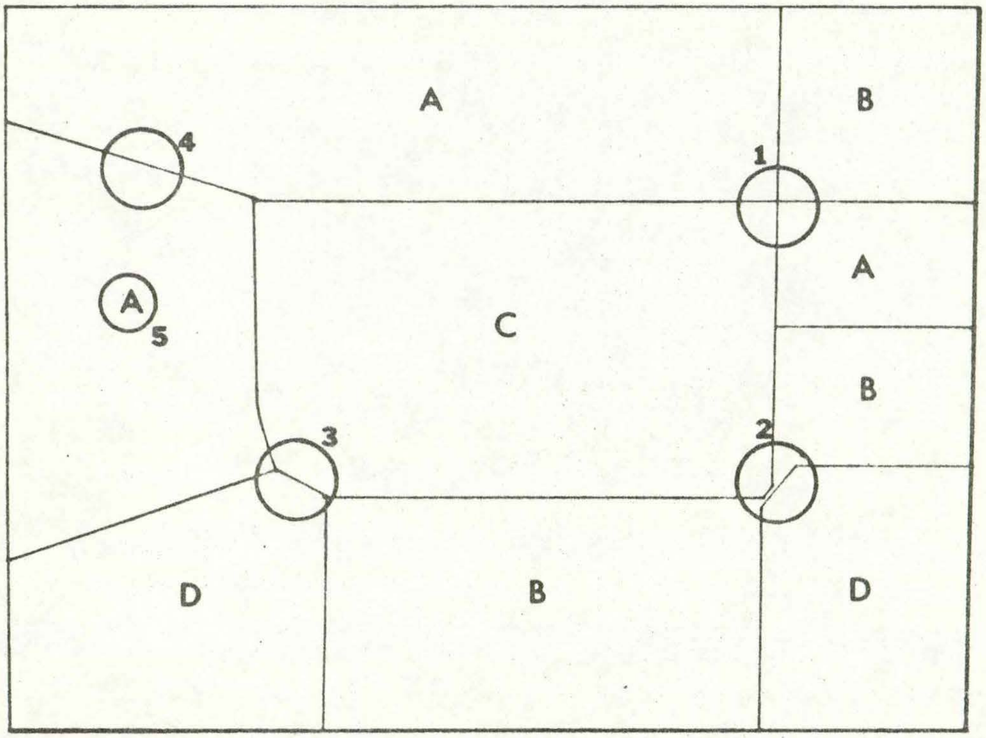
Source: Goodchild (1974 a)
(Modified)

Figure 4. Encoding Errors

errors seven, eight and nine. Error type seven is a pure digitizer error caused by the movement to another point than intended; type eight is where a narrow isthmus was not maintained and the digitizer created two polygons where one was intended; and error type nine, consists of failure to identify contained and containing polygons.

Figure 5 identifies logical errors that exist within files because of source document errors and/or encoding errors and must be detectable or corrected. Error type one consists of junctions with diagonal non-unique center identifiers which result in logical error when edit routines attempt chaining around the polygon to insure correct encoding. This problem exists because many coverages such as land use have non-unique center identifiers. To avoid this problem one can assign unique sequence numbers to every polygon. But then superfluous center identifiers may be created which results in a polygon having two different center identifiers. Error type two is a gap, or isthmus, which is intended but for which the threshold for detecting undershoots has been exceeded and the program logic thereby closes the gap. Error type three is a short line segment which is treated as a junction because it too falls within the tolerance of an undershoot making area center polygon A and polygon B diagonal, when they are really not. Logical error types 4 and 5 occur if superfluous line segments or centers are encoded. Edits that chain around polygons and junction points will identify these problems.

Although this discussion of capture and encoding error pertain directly to digitizers, a scanner which detects the presence of cells that are part of a line segment and converts these line segments to polygon records must deal with the same problem, as well as the additional problem of converting cells that indicate the presence of a line segment to polygon records. Alternatively,



1. Junctions with diagonal non-unique and identifiers
2. Gap or isthmus
3. Short line segment
4. Superfluous line segment
5. Superfluous center
6. Slivers or superfluous polygons that occur when double encoding polygon line segments fall beyond tolerance limits for assuming they are the same segment.

Figure 5. Logical Errors

scanning the coverage to create grid data is a problem of filling the cells with the appropriate polygon center identifiers.

Assuming a clean source document, scanning for the presence and absence of line segments consists of smoothing small scan cells into line segments, identifying junctions, and describing polygons and associating the correct center identifiers with each polygon. Such is done in the CGIS system (Tomlinson, 1967).

Direct scan to grid units is an aggregation problem; one of combining small scan cells into grid units for storage and use. ORRMIS (Durfee, 1974) performed separate scans for each category or classification of a coverage. For coverages consisting of land use or soil type, a manageable number of scans enabled encoding a coverage, whereas a large number of uniquely identified areas requires an inordinate number of scans to encode a coverage for say census tract identifiers.

Scanning or automatic line following technology is more likely to encounter difficulty with source document errors, whereas a digitizer operator can recognize and correct many source document errors. Consequently, line gaps, uneven line widths, intensity variations of patterns on source documents may cause considerable problems with fully automated data capture systems. In addition, the logic of creating line segments and polygons from scan cells may create error situations.

Whether data capture is accomplished by wholly manual methods, such as overlaying a grid on a map, by man-machine interaction, such as digitizing, or completely by machine as in scanning, the potential for error must be anticipated and procedures developed to insure an adequate level of quality of the encoded data to meet the purposes of the intended application.

In digitizing the degree to which interactive editing is employed may be a function of the level of digitizer operators. High level operators may be capable of judgements to interactively edit data; whereas if lower level digitizer operators are employed, post edit of their work may be more advisable.

ISSUE SET #4 WHAT DEGREE OF CLEANLINESS OF THE SOURCE DOCUMENT IS ESSENTIAL?

DOES ENCODING FROM DIGITIZERS REQUIRE AN INTERACTIVE MODE FOR ERROR DETECTION, OR CAN LOGICAL EDITING PROCEDURES IN BATCH MODE OPERATION BE EMPLOYED TO PURGE DATA FILES OF ERRORS?

IS THE FEASIBILITY OF SCANNING A FUNCTION OF DATA DENSITY?

WILL RASTER SCANNING USING INTERACTIVE EDITING COMPENSATE FOR UNCLEAN SOURCE DOCUMENTS?

Exploration of these issues can begin by:

DIGITIZING AND SCANNING EXPERIMENTS USING SOURCE DOCUMENTS OF VARYING DEGREES OF CLEANLINESS.

EFFECTIVENESS MEASURES

One of the major problems in encoding geographic data is the lack of measures by which to assess the effectiveness of the encoding. This section of the paper attempts to identify possible effectiveness measures that would allow comparison of data capture technology and encoding methods. On one hand, there are a set of effectiveness measures that relate to the ability to replicate the source documents in map form and there is a second set of effectiveness measures that are more user oriented. These latter measures relate to the marginal utility of additional precision of data with respect to decision making, which at this time can only be approached by setting standards or requirements related to the degree of aggregation necessary for different classes of problems or applications.

With respect to the former problem, that of replicating coverages or overlays of coverages, several effectiveness measures are suggested. If a coverage is assumed to consist of a large number of pixels or scan cells, an effectiveness measure would be the proportion of pixels that are correctly classified through the polygon and grid coding. Similarly, an overlay of two or more coverages could also be assessed by calculating the number or percent of pixels correctly classified. With respect to area measurement, the area estimates from polygon and grid coding could be compared to the pixel count, to estimate the percent error in area measurement. Similarly, although beyond the scope of this investigation, are effectiveness measures generated from cartographic or map standards, for cartographic applications of the coded data.

In addition, there needs to be comparative measures developed with respect to quality control that encompasses the error rate from the source document to the capture to the editing. What is needed here are per record, per cell, per polygon, per frame, per hour error rates for different system types by coverage type. One such measure would be to regress digitized points for a line of a known function.

Finally, there needs to be effectiveness measures that relate to the applications of updating, edge matching and retrieving data.

McAlpine and Cook (1971) have performed an analysis of overlaying coverage to determine the degree of agreement between initial and derived map segment (polygon) descriptions, and concluded that there is a considerably increasing risk of faulty description of known points as size of map segment (polygon) decreases. See Appendix A.

ISSUE SET #5 WHAT EXISTING EFFECTIVENESS MEASURES CAN BE EMPLOYED?

WHAT NEW MEASURES NEED DEVELOPMENT?

WHAT MEASURES OF COST, OR ERROR ARE APPROPRIATE?

TEST COVERAGES

Figures 6 and 7 are source documents for portions of two hypothetical coverages of the same USGS 7½' quadrangle, which are constructed to possess the source document errors that were discussed earlier, the opportunity for the digitizing errors, and the situations which will result in logical errors. One source document contains non-unique center identifiers, whereas the other contains sequential or unique center identifiers. In addition to the above mentioned situations, there are some known sizes and shapes within the coverage to provide a basis for comparison in terms of area measurement. Assuming that the two coverages will be overlayed, there are cases where lines should coincide exactly and there are cases where lines should intersect in pre-specified ways. Coverages of adjacent quadrangles should be included to test edge matching capability of systems.

Appendix B contains the Statement of Work for a Request for Proposal (RFP) issued by the Corps of Engineers (1975) to demonstrate, digitize, manipulate and display geographic data. Their test coverages consisted of a land use coverage, a floodplain coverage, a topographic contour coverage, an update coverage, and a coordinate listing of point data. Contractors were requested to deliver products generated from digitizing these coverages and the Corps is performing a comparison and evaluation of the results.

The results of this experiment should provide valuable information for the design of test coverages for further and more comprehensive comparative testing of systems.

Legend for Following Figures

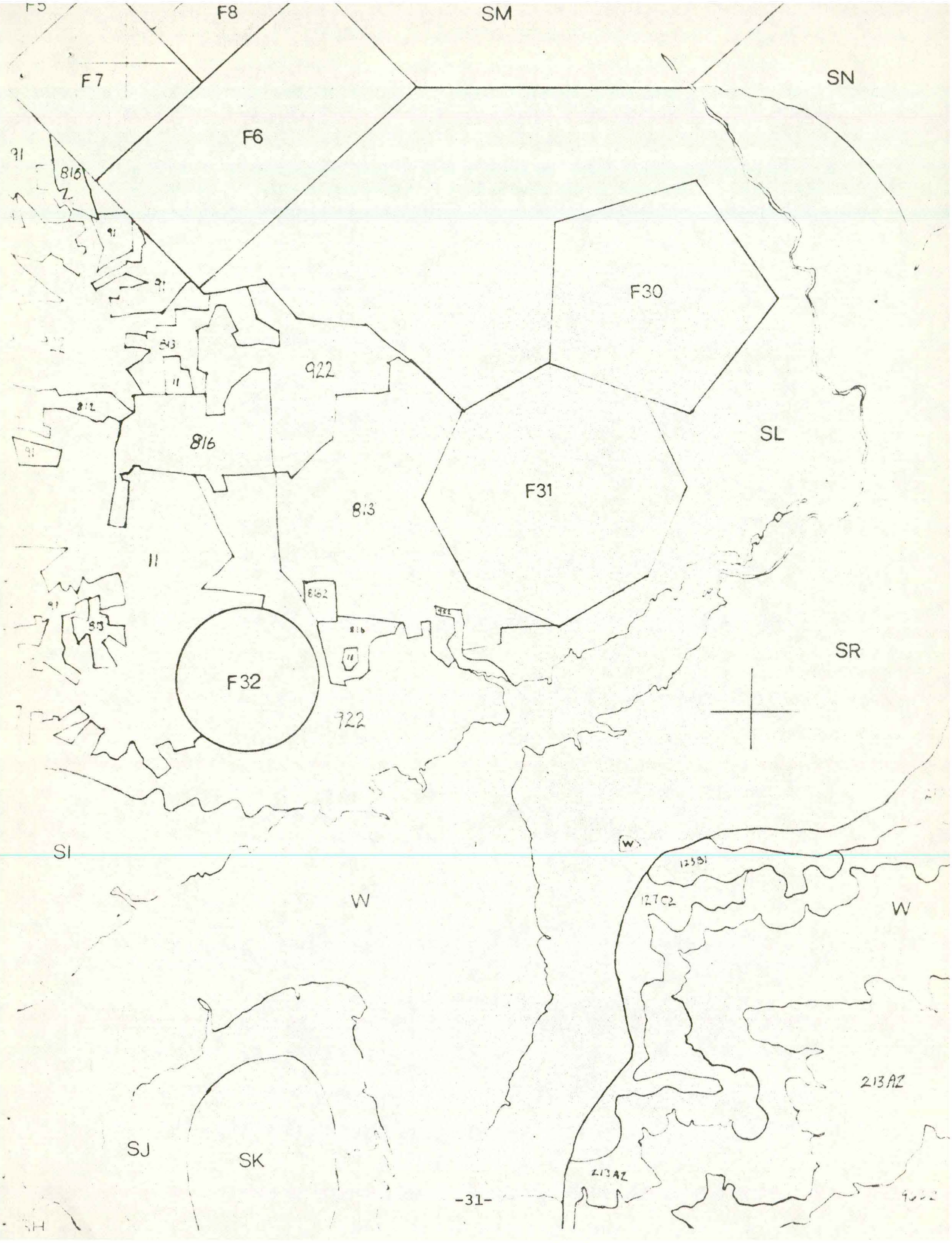
Figure 6. Portion of Test Coverage I. Regular and irregular shapes with non-unique descriptors; areas of some polygons known. Data capture problems described in text are present.

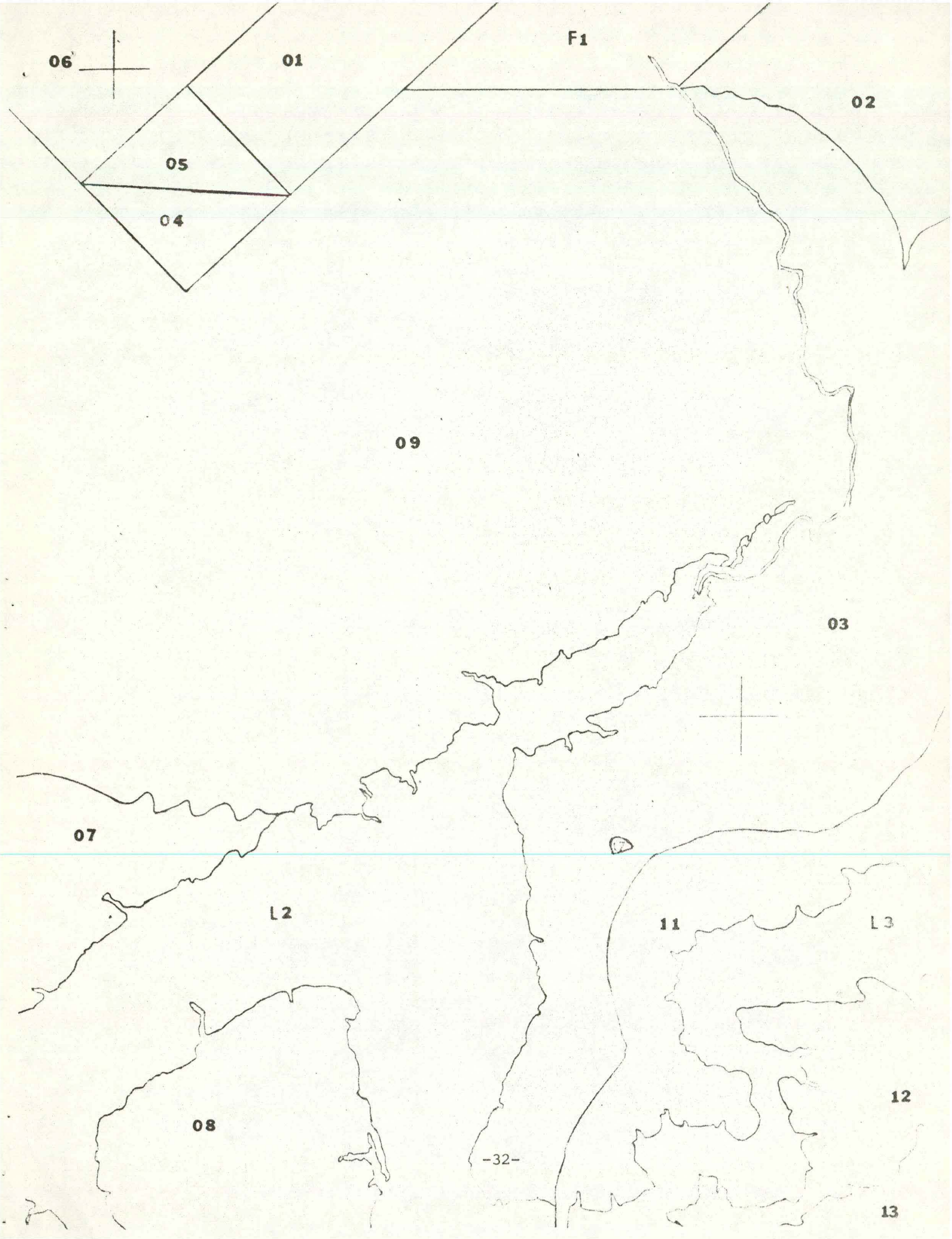
Figure 7. Portion of Test Coverage II. Regular and irregular shapes with unique descriptors. Overlay on Test Coverage I to show correspondence of classificatory units of the two coverages.

Notes:

Test Coverage I includes non-unique descriptors diagonal to one another at junctions, contained polygons, line segments of known length and orientation, areas of known size, narrow isthmus, undershoots, overshoots, and superfluous and missing center descriptors.

Test Coverage II contains line segments that should coincide with line segments of Test Coverage I, segments that should intersect line segments of Test Coverage I in prespecified ways, and resultant areas of known measure.


F5
F8
SM
SN
F7
F6
91
816
91
91
813
11
922
812
816
91
F30
F31
SL
813
11
862
816
11
91
813
F32
922
SR
SI
W
123B1
127C2
W
213AZ
SJ
SK
213AZ
922
SH


06
01
F1
02
05
04
09
03
07
L2
11
L3
08
12
13

ISSUE SET #6 WHAT KIND OF SOURCE DOCUMENT PROVIDES A FAIR AND COMPARATIVE TEST FOR DIFFERENT SYSTEMS, PARTICULARLY FOR SYSTEMS DESIGNED FOR DIFFERENT SCALES AND OBJECTIVES?*

SHOULD TEST COVERAGES INCLUDE SOURCE DOCUMENT ERRORS? CAN SOURCE DOCUMENT ERRORS BE SYSTEMATICALLY VARIED TO DETERMINE EFFECTIVENESS OF SYSTEMS TO DETECT AND CORRECT?

ARE THERE OTHER SITUATIONS OR TESTS THAT SHOULD BE INCORPORATED INTO THE TEST COVERAGES?

SHOULD THE TEST COVERAGES INCLUDE DIFFERENT LINE SYMBOLIZATION (E.G., DASHED AND SOLID LINES)?

BENCHMARK TESTS

In computer applications, benchmark testing is used to compare performance and efficiency. There are essentially three ways to benchmark test a system. One way is to take a standardized job mix and run it through various computers to compare the central processing units. Time factors are the major concern. The second way is to take different software programs and run the standard job mix to compare efficiency in running the programs with the set hardware system. This essentially compares how well different software packages utilize the same hardware. The third class of benchmark testing varies the input/output devices. This would depend on whether the user required, for example, large computation abilities but minor input/output requirements versus a primarily storage operation requiring extensive input/output capabilities.

In benchmark testing, one needs to define the quality attributes sought and express them in some metric form. These measures assess with mathematical and logical accuracy the performance of various operations.

---

*The test coverages developed here do not include variations in line-width, discrimination between colors, or discrimination between gray tones as this is a test of encoding coverages, i.e., areas of like characteristics separated by networks of lines.

With respect to data encoding, one objective is to compare the accuracy of the final product with the original input. This kind of benchmark test checks the efficiency of the hardware system in the functioning of the software programs. This requires use of standardized input in the comparison of the outputs utilizing statistical tests.

Thus the effectiveness measures in the test coverages discussed in the prior two sections provide a basis for constructing a benchmark test.

ISSUE SET #7 HOW BROAD OR HOW NARROW SHOULD ALL ENCODING TESTS BE? SHOULD IT ONLY TEST THE CARTOGRAPHIC REPLICATION OF SOURCE DOCUMENTS? SHOULD IT INCLUDE OVERLAY ANALYSIS? SHOULD IT INCLUDE OTHER ANALYSIS, E.G., GENERATING SLOPE MAPS?

CAN ONE GENERAL BENCHMARK TEST BE CONSTRUCTED FOR BOTH DIGITIZERS AND SCANNERS?

SHOULD THERE BE SEPARATE BENCHMARK TESTS FOR EVALUATING SOURCE DOCUMENTS? FOR ENCODING ERRORS? FOR LOGICAL ERRORS?

IN CONCLUSION

This paper has identified a series of issues in encoding coverages which will require attention within the next few years. The major unresolved issue is the breadth of the tests to compare encoding processes. Geographic data handling for statewide land use applications requires addressing the issues raised in this paper. Although systems will have to be designed before these issues are resolved fully, the issue identification process alerts system designers to potential problem areas.

Until some of these issues are resolved, system designers should caution planners of statewide land use information systems of the potential for errors, and delays and cost overruns when attempting to encode and replicate a large number of complex coverages. Presently a coarser statewide system is more appropriate, while at the same time undertaking prototype developments in smaller study areas to test more sophisticated encoding tehcniques, and to develop staff capabilities.

Acknowledgements

The author wishes to acknowledge John Milligan, Steven Schmidt, and Richard Talcott, Research Assistants, Institute of Urban and Regional Research, University of Iowa for assistance in preparing this issue paper. Jack Dangermond, Michael Goodchild, David Pinick, Scott Sollers provided valuable comments and suggestions in a meeting to review a draft of the paper. Hugh Calkins, Charles Meyers, and David Sinton also reviewed draft materials and provided valuable comments and suggestions.

REFERENCES

Corps of Engineers, 1975, "Request for Proposal to Demonstrate, Digitize, Manipulate and Display Geographic Data", Washington, D.C.: Department of the Army, Office of the Chief of Engineers, Directorate of Civil Works, DAEN-CWA-C (RFP NO. DACW73-75-R-0003.)

Durfee, R.C., 1974, "ORRMIS Oak Ridge Regional Modeling Information Systems", Oak Ridge, Tenn.: Oak Ridge National Laboratory, ORNL-NSF-EP-73.

Goodchild, Michael F., 1974a, "PLUS: Technical User's Guide", London, Ontario: Department of Geography, University of Western Ontario, (mimeo).

Goodchild, Michael F., 1974b, "Error Estimation for Geographic Data Structures", Department of Geography, University of Western Ontario, (mimeo), 19 pp.

"Land Management Information in Minnesota: The Beginning of a Statewide System", 1972, Minneapolis: Center for Urban and Regional Affairs, University of Minnesota and Saint Paul: Minnesota State Planning Agency.

McAlpine, J.R. and Cook, B.G., 1971, "Data Reliability from Map Overlay", Canberra, Australia: Division of Land Research, CSIRO, presented at Australian and New Zealand Association for the Advancement of Science, 43rd Congress, Brisbane, May 1971.

Rosenfeld, Azriel, 1969, Picture Processing by Computer, New York, Academic Press.

Shelton, R.L. and Liang, Ta, 1975, "Land Inventory Systems: The Cornell Experience", Ithaca: Cornell University.

Tomlinson, R.F., 1967, "An Introduction to the Geographic Information System of Canada Land Inventory", Geographic Information Systems ASP-ACSM Annual Convention, Washington, D.C.

APPENDIX A

DATA RELIABILITY FROM MAP OVERLAY

Source:

McAlpine and Cook (1971)

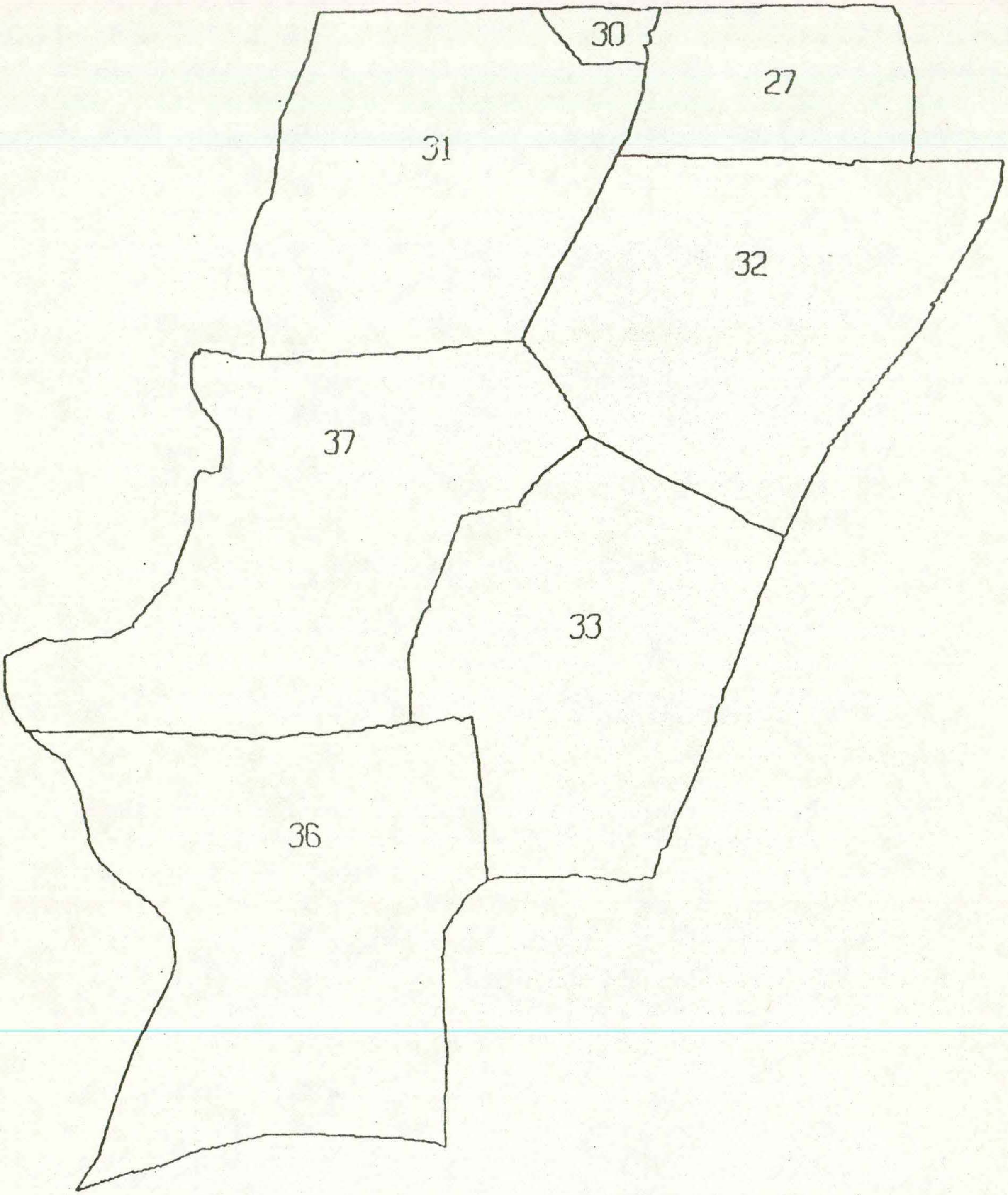


Fig. 4.- Census Division initial map segments.

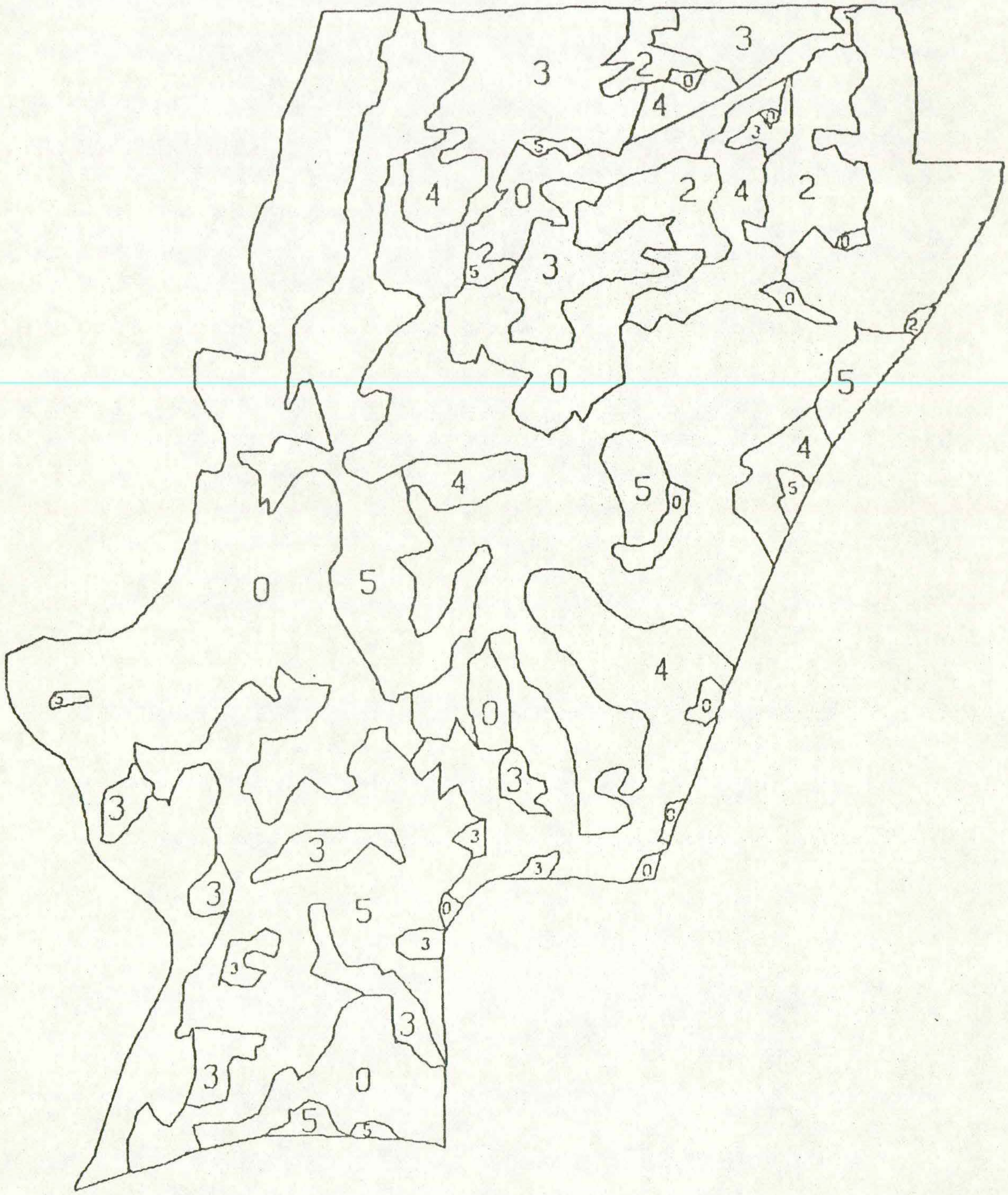


Fig. 5.- Land use intensity class initial map segments.

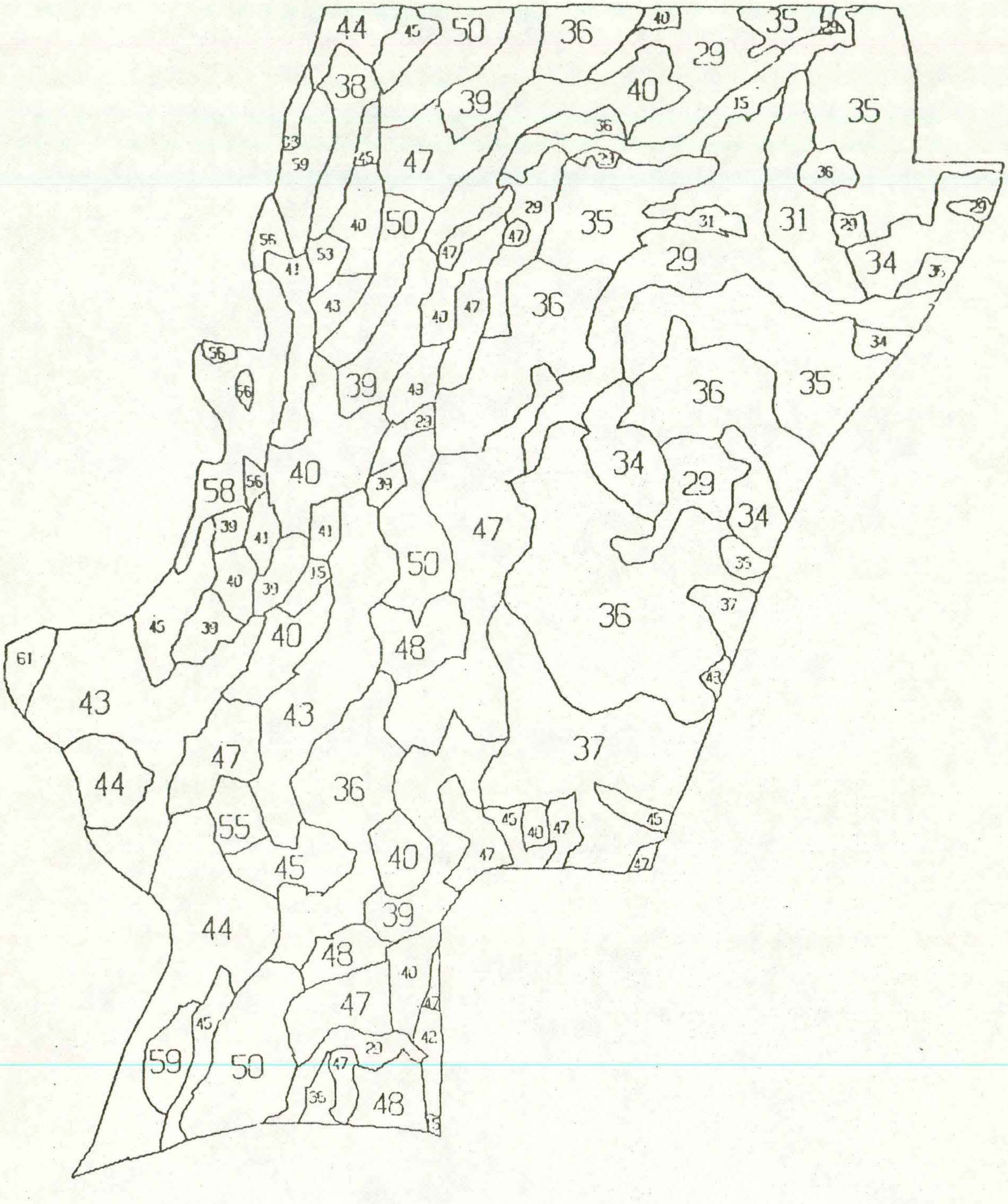


Fig. 6.- Land system initial map segments.

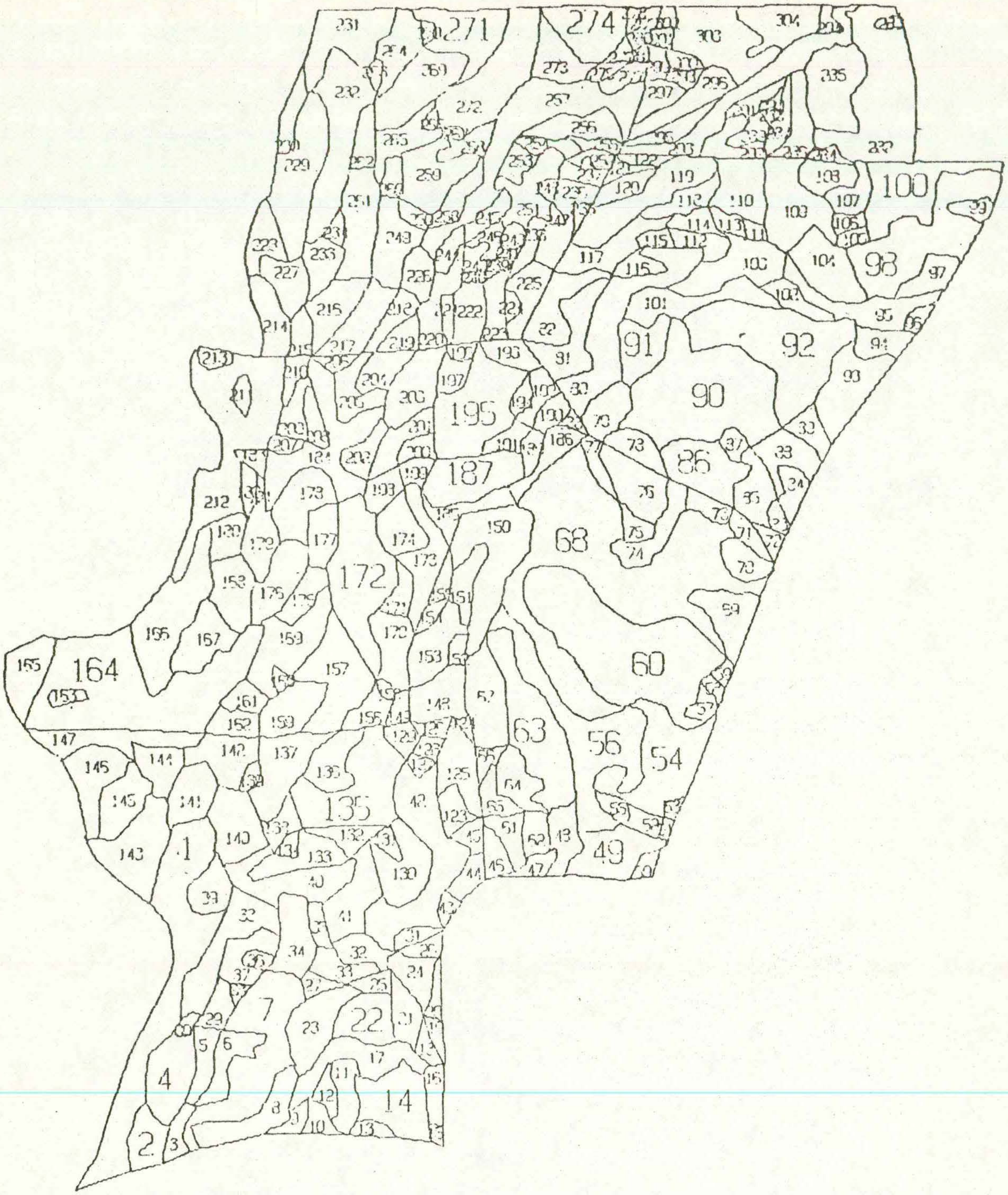


Fig. 7.- Derived map segments from overlay of Figs. 4, 5, and 6.

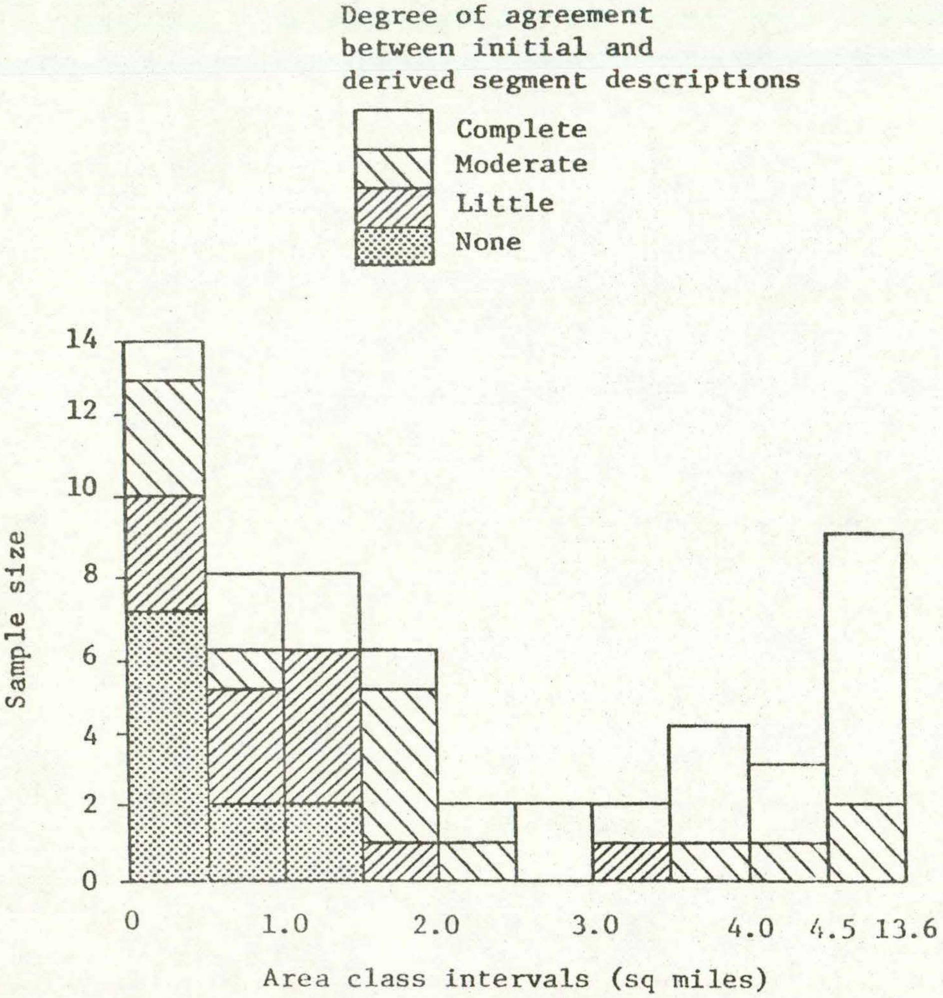


Fig. 8.- Results of case study test.

APPENDIX B

DEMONSTRATE, DIGITIZE, MANIPULATE, AND DISPLAY GEOGRAPHIC DATA

Source:

Corps of Engineers (1975)

STATEMENT OF WORK

1. PURPOSE OF PROGRAM. The purpose of this program is to enable the commercial concerns to demonstrate their present capabilities by a comparative project. The utilization of geographic data can be considered to consist of the five general phases of:

a. Collection of Raw Data

b. Interpretation and Extraction of Information

c. Data Processing and Manipulation

d. Modeling and Analysis

e. Decision Making

For the purpose of this program only the third phase, data processing and manipulation, are to be demonstrated.

2. GOVERNMENT FURNISHED DATA. The following items (data products) will be furnished to the contractor by the Government for use in the demonstration:

a. Item a. 1:24,000 U.S.G.S. Topographic Quadrangle sheet, Healdsburg, California.

b. Item b. Land Use Overlay to the northwest quarter of the Healdsburg Quadrangle.

c. Item c. Land use overlay coding legend.

d. Item d. 100 year floodplain, update information and point data overlay to the northwest quarter of the Healdsburg Quadrangle.

e. Item e. Topographic Contour Overlay to the northwest quarter of the Healdsburg Quadrangle.

f. Item f. Coordinate listing of point data (Vehicular Traffic Intersections in State Plane Coordinate). Note: Overlay in item 2d contains 6 point data markings, item 2f refers to 6 additional point data.

3. SERVICES REQUIRED TO BE DEMONSTRATED. The contractor shall provide all personnel, facilities equipment and material necessary to perform the following services:

a. Digitize and encode the graphic products listed as Items b, d, e, and f in above Article 2. With respect to Item 2e, Topographic Contour Overlay, digitize only index contours (100 foot contour interval lines).

3. SERVICES REQUIRED TO BE DEMONSTRATED. (Continued)

With respect to Items 2d and 2f, point data, digitize total of 12 vehicular traffic intersections.

b. Demonstrate overlay manipulation by identifying the individual land use areas and point data within the floodplain and below the 100 foot contour line. (See Fig. 1, p. S-4).

c. Demonstrate an update capability without redigitizing the basic data by identifying the individual land use areas and point data affected by the change presented in Article 2d (Update overlay).

4. DELIVERABLE PRODUCTS. The contractor shall furnish to the Government at the address shown in block 14 of DD Form 1155 the following:

a. Separate line graphic products on white paper produced from the digitized land use and contour overlays at the same scale as the input graphics.

b. A tabular listing of the acreage to the nearest hundredths of an acre for each land use polygon depicted on the overlay described in above Article 4a. The listing should include the land use codes and polygon coordinates and acreages. The contractor will determine State Plane coordinates for each polygon by using the middle letter of each land use code corresponding to the appropriate polygon as the locational point.

c. A graphic product on white paper displaying the land use areas and point data defined in Article 3b at the same scale and using the same coordinate system as the input data. (See Fig. 1, p. S-4).

d. A tabular listing of the acreage to the nearest hundredths of an acre for each land use polygon depicted on the overlay described in above Article 4c. The listing should include the land use codes and polygon coordinates and acreages. The contractor will determine State Plane coordinates for each polygon by using the middle letter of each land use code corresponding to the appropriate polygon as the locational point.

e. A graphic product on white paper displaying the land use areas and point data defined in Article 3c at the same scale and using the same coordinate system as the input data.

f. A tabular listing of the acreage to the nearest hundredths of an acre for each land use polygon depicted on the overlay described in above Article 4e. The listing should include the land use codes and polygon coordinates and acreages. The contractor will determine State Plane coordinates for each polygon by using the middle letter of each land use code corresponding to the appropriate polygon as the locational point.

g. A report which includes a brief description of:

4. DELIVERABLE PRODUCTS. (Continued)

(1) The identity and general characteristics of all major items of equipment such as digitizers, computers and display devices.

(2) Precision specifications of digitizing and plotting equipment.

(3) The processes used with particular emphasis on those techniques or processes considered to be unique or innovative.

(4) Edit procedures used to locate and correct errors in the digitization process.

(5) Basic word length and core, disk or tape storage requirements for the processes used in the program.

(6) The contractor's capability to input and display geographical data in reference to various coordinate systems (eq. Universal Transverse Mercater, Geographic Coordinate Systems or State Plane systems) and various scales (e.g. 1:24,000, 1:50,000 etc).

(7) Equipment, software or techniques that would be considered to be proprietary in a production environment.

(8) Contractor's capability to merge data geographically from more than one map sheet and ability to perform Boolean manipulations on an irregularly bounded area covering portions of more than one quadrangle. (e.g. the user may desire land use statistics for an area defined by a watershed which includes terrain mapped on portions of 2 or more Quadrangle sheets).

(9) Dollar estimates presented for digitizing each type overlay considering a total of 50, 100 and 150 of each type overlay to entire 1:24,000 U.S.G.S. quadrangle sheets assuming similar relative data complexity per overlay.

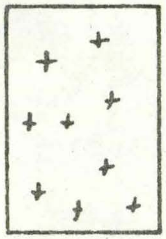
(10) Dollar estimates presented per task (Article 3b, 3c) for 50, 100 and 150 of each type overlay to entire 1:24,000 U.S.G.S. quadrangle sheets assuming similar data complexity per overlay.

(11) Relationship between geographic precision and areal accuracy of results vs actual and projected production costs.

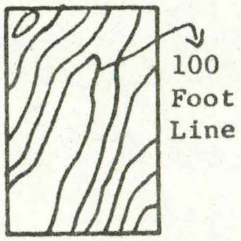
5. SCHEDULE OF WORK. The contractor shall commence work within one calendar day after date of receipt of this order and shall prosecute the work and services in accordance with this statement of work. The contract shall be completed upon receipt and approval of the deliverable products by the Government or 30 calendar days after commencement of work, whichever is later.

FIGURE 1. Example of Services required as described in Article 3b and Deliverable Products as described in Articles 4c and 4d.

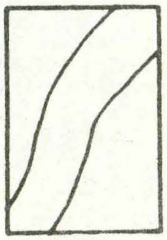
Land Use



Point Data



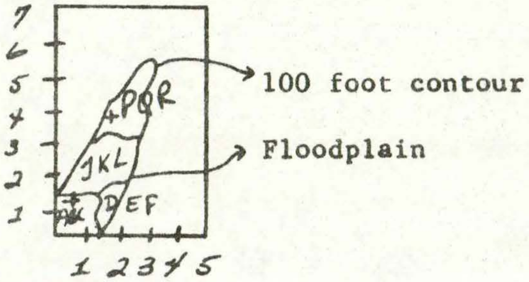
Contours



Floodplain

(1) Digitize and encode Land Use, 12 point data, 100 foot contours and the Floodplain.

(2) Produce graphic of land use areas and point data in the floodplain and below the 100 foot contour line.*



Composite Graphic

(3) Produce tabular listing of land use areas in the floodplain below the 100 foot contour line.*

CODE	COORDINATES (State Plane)	ACREAGE
ABC	0.50x0.80	42.39
DEF	2.32x1.45	26.24
.	.	.
.	.	.
.	.	.

*NOTE: Only portions of many land use polygons will meet the criteria - The graphics and statistics should be developed for those land use polygons and portions thereof that remain after the manipulation or update.

RFP NO. DACW73-75-R-0003
Sheet 1 of 2 Sheets

FORMAT-PRICE PROPOSAL BREAKDOWN

ITEMS BREAKDOWN: (Attach extra sheets if additional space is needed)

(1) DIRECT LABOR COST:

TYPE OF PERSONNEL

____________	________ M.H.	$________	AV $________
____________	________	________	________
____________	________	________	________
____________	________	________	________
____________	________	________	________
____________	________	________	________

SUB-TOTAL DIRECT LABOR $__________

(2) OVERHEAD ON DIRECT LABOR (%) __________

(3) GENERAL & ADMIN. OVERHEAD(%) __________

(4) MATERIALS, SUPPLIES __________

(5) OTHERS (Renderings, Reproduction, Consultants, Computer time, etc.) (Describe these items in detail) __________

SUB-TOTAL COST $__________

TOTAL (All options included) $__________

Instructions to Offerors

The purpose of this form is to provide a uniform document on which the Offeror submits to the Government a summary of proposal costs (and attached supporting information) suitable for detailed review and analysis. The above format is not intended as a rigid requirement. Data may be presented in another form, if required for a more effective and efficient presentation of cost.

APPENDIX C

NOTES:

ITEM (1) Direct Labor Cost. Enter the hourly salary computed on the basis of the annual salary rate reportable on the individual's TD Form W2 for income tax purposes plus a factor, if any, for raises contemplated during the contract period. Provide a separate breakdown of labor by job category and furnish basis for cost estimate. Identify subcontracted labor and indicate basis of establishing source and reasonableness of cost.

ITEMS (2) & (3) Overhead. Provide the method of computation and application of your overhead expense, including cost breakdown, and showing trends and budgetary data as necessary to provide a basis for evaluation of the reasonableness of proposed rates.

Has the Department of Defense, National Aeronautics and Space Administration, or the Atomic Energy Commission performed any review of your accounts or records in connection with any other Government prime contract or subcontract within the past twelve months?

☐ YES ☐ NO. IF YES, IDENTIFY BELOW.

NAME AND ADDRESS OF REVIEWING OFFICE	TELEPHONE NO.
______________________________	____________
______________________________	
______________________________	

****