GEOLOGICAL SURVEY CIRCULAR 712

Earth Science Data in Urban and Regional Information Systems—A Review

Work done in cooperation with the
U.S. Department of Housing and Urban Development,
Office of Policy Department and Research

Earth Science Data in Urban and Regional Information Systems—A Review

By Victor W. Adams

GEOLOGICAL SURVEY CIRCULAR 712

Work done in cooperation with the
U.S. Department of Housing and Urban Development,
Office of Policy Department and Research

1975

United States Department of the Interior

Geological Survey

V. E. McKelvey, *Director*

Library of Congress Cataloging in Publication Data

Adams, Victor W.
Earth science data in urban and regional information systems–A review.

(Geological Survey Circular 712)
Bibliography: p. 28–29.
1. Information storage and retrieval systems–Earth sciences. I. Title. II. Series: United States. Geological Survey Circular 712.
QE75.C5 no. 712[QE48.8] 557.3'08s[550'.28'5] 75–33226

Free on application to the U.S. Geological Survey, National Center, Reston, Va. 22092

CONTENTS

ILLUSTRATIONS

TABLES

Earth Science Data in Urban and Regional Information Systems—A Review

By Victor W. Adams

ABSTRACT

Computerized information systems for urban land use and regional planning are surveyed in terms of their capability for handling geographically related data, especially earth science data. The need for planners to consider these data stems from the growing necessity to consider (1) the environmental impact of urbanization of open areas and (2) environmental hazards in the development of long-range plans and subsequent zoning and construction ordinances. This study involves earth science data for the nine-county San Francisco Bay region.

Computer information systems are described briefly, and terms used by computer technologists are defined.

Studies made by the U.S. Geological Survey for the benefit of urban planners in the San Francisco Bay region have involved a wide variety of geologic data. Many of the techniques used in deriving the end result can be duplicated in the computer.

Urban and regional information systems that must handle relatively large (multicounty) areas may require a wide range of capabilities. Factors to be considered include the following: (a) interactive and batch mode capability, (b) large capability for data manipulation and computation, (c) multiple type input and output devices, (d) multiterminal support, and (e) support of multiple data bases. The extent to which a particular system incorporates these capabilities depends on regional needs. The specifications of any particular system should be determined by its users for the natural environment of the region. The users, primarily planners, should be aided by environmental scientists and computer specialists. Until such specifications are made, it is impossible to estimate the cost of a system, but the initial cost of the data base will be high.

INTRODUCTION

The San Francisco Bay Region Environment and Resources Planning Study (SFBRS) is a cooperative effort that began in January 1970 between the U.S. Geological Survey of the Department of Interior and the Office of Policy Development and Research of the Department of Housing and Urban Development. It is a study of physical environmental factors, particularly geologic hazards, and their relation to urban and regional planning. The area of study is the nine-county San Francisco Bay region. The need for such a study grew from the recognized need for incorporating physical environmental data in the regional planning effort. The products of the study are maps and reports divided into three series: basic data contributions, the technical series (derived from the data for a technical audience), and the interpretive report series (a final derivation for a nontechnical audience such as planners and government officials). The program elements are grouped into four categories: topographic, geologic and geophysical, hydrologic, and planning program elements. The U.S. Geological Survey and Department of Housing and Urban Development (1971) have published a detailed description of the program elements and products of the bay region study. This report is one of the planning program elements.

Environmental factors are becoming increasingly important to planners in the bay region. Hazards associated with these factors, such as earthquake fault zones and slopes that are potentially unstable, should be recognized and dealt with in the planning phase. The consideration of environmental factors also entails the need to estimate the impact on the environment of various land-use alternatives associated with rapid urbanization. Many environmental factors in the area under study are regional in nature. Earthquake fault patterns are very extensive. Depletion of an underground aquifer has caused land subsidence in adjacent areas as well as in the immediate area. Thus, the amount of data to be considered by planners in studying land-use alternatives can become very great, and use of this data in the traditional form of maps and reports can be very time consuming. Increasingly, this has led to the use of modern electronic computers and to the development of information systems to handle environmental and land-use type data.

The management of data has become a problem in modern society. Almost everyone is familiar with the vast amount of data involved in processing credit card information and in the processing of hundreds of millions of federal and state income tax returns. Government agencies and private companies concerned with the land-use planning process are also deeply involved in this problem. The amount of data handled has grown, but more important has been the growth in requirements for managing and manipulating this data. Now environmental data must become a part of the pool of data used by the computer as a part of a land-use information system.

This study is concerned with problems associated with the inclusion of earth science data in computer-based urban and regional information systems. How, for example, is earthquake data on a map read into a computer? How are map areas with potentially unstable slopes to be stored in a computer data base? This report is a summary of the work of others since the SFBRS is not directly involved in the development of such an operational computer system but rather is interested in the proper and timely use of earth science data by regional planners. The characteristics of urban and regional information systems are determined mainly from a study of two systems in the San Francisco Bay region, that of the Metropolitan Transportation Commission and that of Santa Clara County, which are taken to be typical of regional computer-based information systems. These systems were used to examine the problems involved in making earth science data compatible with existing urban and regional information systems.

This report is primarily for planners, especially those with a limited knowledge of computer technology. It may also be valuable to scientists and others working with planners in this application of earth science technology. In view of the intended audience of this report, a brief section on computer-based information systems provides background information and defines terms for an understanding of the sections that follow.

COMPUTER INFORMATION SYSTEMS

This section presents a brief, nontechnical description of certain computer information systems to provide a frame of reference for later sections. First, digital computer systems in general are discussed to define terms and give a summary of the rapid development of computer technology and what may reasonably be expected in the near future. Next, two special computer systems are defined: management information systems and geographic information systems.

COMPUTER SYSTEMS

A digital computer system can perform the following operations: (1) "read" data or instructions with such devices as card readers or keyboards, (2) perform arithmetic operations on data, (3) determine which operations to perform on the basis of instructions or the result of some previous operations, and (4) "output" the results in a form such as hard copy by a printer or by graphic display on a device similar to a television screen, hereafter referred to as a display tube. The sequence of operations is given by a set of instructions called a program. Within the computer the data and the instructions are in the form of a binary code because the circuit elements operate at two voltage levels that can be called "on" and "off" or "zero" and "one." Each on or off stage represents one binary digit (one bit of information).

A typical system consists of a central processing unit (CPU), various input and output (I/O) devices, and auxiliary storage such as disk storage units and magnetic tapes. The CPU is the heart of the computer system comprising main storage, which holds programs and data, an arithmetic unit, and associated circuitry. The way in which the physical units are put together is called hardware architecture.

Programs are of two types, systems or application, and together constitute what is called software architecture. Systems programs are a semipermanent part of the computer system since they are infrequently replaced. They include support programs for system and program control, programs for hardware testing, and standard mathematical subroutines, for example, and can be thought of as residing in the machine. As such, their instructions are in binary code in what is called machine language. System software relieves the user from the necessity of writing common segments of instructions for each program. Applications programs, as the name implies, depend upon specific user requirements, may be changed frequently, and can be thought of as being read into the machine for each use. Most applications programs are written in a higher level language such as fortran (an acronym for formula translation) in which the statements are quite similar to English. The advantages of a higher level language are that it is independent of the machine and easy to learn and use compared with machine language.

When an application program with all the data it requires is read into the machine at one time and results are outputted after the program has been completed, the computer system is operating in batch mode. If the user reads in a few instructions at a time at a keyboard terminal and waits for results to be outputted (as on a display tube) before reading in more instructions, the computer system is operating in conversational or interactive mode. This is usually done when more than one user shares the computer in a multiple access or time-sharing system. Switching from one terminal (user) to another is so rapid that any one user has the impression that only he is using the system. Special higher level languages are used for the conversational mode.

Table 1 summarizes some information on four generations of computers and generally indicates the improvement in computer technology over the last 20 years. While the magnitude of any particular figure will be meaningful only to specialists, the trend over the 20-year period can be appreciated by all.

MANAGEMENT INFORMATION SYSTEMS

The first business application of the computer was recordkeeping in which the computer replaced clerks. This use was followed by simple control systems such as for inventory or production. For these applications the main function of the computer was to handle a large amount of data. These data, stored in the computer system in auxiliary storage devices, are called the data base of the system. The technical problems involved are those of information storage and retrieval, the computer system requiring a large I/O capability rather than a large capacity for computation.

Within the last several years computer systems have been developed with the goal of providing management with the information it requires for decisionmaking; these systems are called management information systems and incorporate many of the tools and techniques of the mathematical sciences such as statistical analysis of data, simulation and modeling of various parts of the business operation or segments of the economy, and utilization of various methods of optimization. The computing ability of the system has been enlarged without sacrifice of the I/O capability.

GEOGRAPHICAL INFORMATION SYSTEMS

A geographical information system is a computer information system with the capability of handling spatially related data, specifically map-related data such as tract boundaries and stream networks. The spatial characteristics should be preserved in the system so that (1) the data can be indexed by geographical location, (2) the data can be manipulated with relation to these spatial characteristics, and (3) one form of output can be a map. The methods by which these are achieved are discussed in a following section.

EARTH SCIENCE DATA AND THE SAN FRANCISCO BAY REGION

The earth science data involved in this study are restricted to the data generated by the San Francisco Bay Region Study Program. The San Francisco Bay region was chosen for study for the following reasons:

1. There is a wide variety of environmental factors in the area, such as the natural hazards of earthquakes, landslides, and flooding, as well as the environmental effect of the bay. These factors constitute the main reason why the bay region was originally selected for the study program.
2. Many of the land-use problems are regional in nature, such as waste disposal and industrial pollution, water supply and land subsidence,

TABLE 1.—*Characteristics of four generations of computers*

Computer generation	1	2	3	4
Period	1952–58	1958–63	1963–69	1969–?
Basic electronic component	Vacuum tube	Transistors, solid-state diodes.	Integrated circuits.	Large-scale integrated arrays.
Component density/ft^3	10^4	10^6	10^7	10^8 up
Cost, \$/bit: Core	2	.75	.1	.005
Memory	.5	.05	.005	10^{-5}–10^{-6}
Cost (\$ to execute 10^6)	10	1	.05	.001
Relative failure rate	1	.01	.001	.0001

landfill and its effects on the bay, and those problems associated with industrial locations, residential development, and transportation.

3. The bay region is an area of rapid urban growth, and its land-use problems need to be solved soon.

The following statistics (U.S. Geological Survey and Department of Housing and Urban Development, 1971) of the San Francisco Bay region are included for those who may not appreciate its size:

Population (1970)	5,083,549
Dwelling units (1965)	1,404,146
Area (land)	6,952 mi²
San Francisco Bay, original area (1835)	680 mi²
Filled or diked area	280 mi²
Remaining area	400 mi²
Linear shoreline	276 mi
Shoreline, public access	5 mi

Comparisons (1970)	*Area*	*Population*
San Francisco Bay region	7,416 mi²	5,083,549
Israel	7,993	2,910,000
Connecticut	5,009	3,082,217

The various interpretive studies, or program elements, underway on the San Francisco Bay region program were reviewed. These studies, subject to change before completion of the SFBRS, are listed below:

Program element	*Published products*
Slope stability study	Landslide susceptibility maps, four to seven zones, depending on the area.
Bay mud	Maps of land-use categories, three to fifteen zones.
Coastal erosion (San Mateo County).	Coastal strip maps of erosion hazards, three to four zones.
Unconsolidated deposits	Map of land-use categories, five zones.
Active faults	Map of active fault zones.
Engineering behavior	Map with matrix of descriptive information and inferred engineering behavior for 30 to 50 map units (bedrock).
Water-quality studies	Report of water quality at various locations in streams, lakes, and reservoirs.
Erosion and sediment studies.	Map of erosion provinces, six to seven zones (plus cross sections).
Flood inundation maps	Map of flood-prone areas (boundary of 100-year flood zone).
Urban drainage system	Peak discharge for design storms of 2, 5, 10, 25, 50, and 100 years.
Land pollution susceptibility study.	Map of evaluation of site locations for four types of disposal systems.
Coastal flooding	Map of areas of potential inundation by tsunamis.
Mineral commodity study	Commodity location maps, 30 commodities.
Seismicity and ground studies.	Seismic susceptibility map, eight zones.
Earthquakes and their locations.	Computer data bank of recent earthquakes in the San Francisco area.
Ground water studies	Map of aquifer yield, four zones. Map of certain chemical constituents of ground water, four zones. Ground water hazards map. Water storage tables for main alluvial valleys.

Most of these products are derived from basic information such as topography, location of active fault zones, and knowledge of bedrock formations, through the application of scientific and engineering principles. Basic data for the purpose of this report are considered to include:

1. Geologic map
   a. Bedrock (including depth, where available, and strike and dip)
   b. Fault lines (or zones)
   c. Unconsolidated deposits
   d. Landslides
   e. Mineral deposits, sand and gravel
2. Slope map (or slopes derived from topographic map)
3. Topographic map
4. Rainfall map (two types, annual average and maximum 48 hours)
5. Vegetation map
6. Land-use map
7. Land subsidence map
8. Tabular information
   a. Engineering properties of soil, unconsolidated deposits, bedrock
   b. Stream-quality information and rating tables (spot locations)
   c. Temperature and other variables in lakes and reservoirs
   d. Suspended sediment (spot stream locations)
   e. Accumulated sediment (various lakes and reservoirs)
   f. Depth to ground water (spot locations)
9. Earthquake data
10. Soil map
11. Historic flood data

For a geographical information system, this information can be classified as follows:

1. Areal or zonal information, such as land-use zones or bedrock zones, represented on map by boundary lines.
2. Line information such as stream networks or streets.
3. Point information such as well locations.
4. Tabular information such as engineering properties of soil.

Examples of these information categories are given on figure 1. Areal information is illustrated by surficial deposits and rock formations, for example, the bay mud. Line information is illustrated by the San Andreas fault line and also by the bedrock contour lines. An example of point information is the location of small landslide deposits denoted by the small triangles.

METHODS OF EARTH SCIENCE DATA ENTRY

This section examines the methods by which earth science data are entered into the computer data bank and the various kinds of equipment used in the process.

MAP DIGITIZATION AND IMAGE DIGITIZERS

Earth science data was classified in the last section as tabular, point, line, and areal data. The entry of tabular information into the computer data bank presents no special problem. By properly labeling the table and organizing its contents, there is no difficulty in providing software that can locate the table and extract information from it. The same is true for point information because it can be referenced by its coordinates, for example, latitude and longitude. Areal and line information must be handled differently.

Because an area is represented on a map by its boundary line, the problem of entering areal or line information into a computer is the same—to convert a line into a set of discrete values that can then be read in. The process is called map digitization. One way this can be accomplished is by approximating a curved line by a series of straight line segments (fig. 2). A straight line segment can be represented by its end points, so that a line can be represented by a series of points, each given by its coordinates. The closer together the points, the better the approximation to the original line.

An image digitizer (fig. 3) is a device with a reading head free to move in two dimensions over a flat surface on which the map to be digitized is placed. An operator follows the line with the reading head whose position is always digitally known. Coordinates of its position can be read in automatically (a "read in" for a preset time interval or a "read in" when displacement in the X or Y direction exceeds a preset amount) or at operator command. Coordinates are measured from some zero position set at start of digitization.

Speed of digitization depends upon operator skill and map complexity. Boyle (1972, p. 699, 700) gives a range of 120–360 inches (300–900 cm) per hour. Experience within the U.S. Geological Survey gives a speed of 120 inches (300 cm) per hour for digitizing elevation contours on a topographic map.

The cost of image digitizers ranges from about \$15,000 to \$55,000,[1] depending upon precision and auxiliary equipment.

Automatic digitizers are made in which the operator function is performed by an automatic line follower. Experience with these devices for map digitization is limited compared with the use of image digitizers. Boyle (1972, p. 702) gives 3,600 inches (9,000 cm) per hour as an attainable rate. High speeds are possible only for very simple maps. Operator initiative is required where a decision must be made, such as where branching occurs, slowing the process considerably. This equipment is now priced in the range of \$100,000 to \$200,000.

A recent book by Auerbach Publications (1972) gives an excellent description of the equipment as well as price information for both manual and automatic digitizers.

OPTICAL SCANNERS

A map can be digitized by an optical scanning device that records coordinates (explicitly or implicitly as described below) of any information differing from the map background as the map is scanned in a regular way. On a map with line information of only one kind, the device will see background (white) or part of a line (black). It looks at a small square (for example, 0.001×0.001 inch) at a time and sees either predominant white or predominant black since line thickness is much greater than 0.001 inch (if the square is half black and half white, the scanner probably will see

[1] Prices and cost estimates in this circular refer to 1972–73, when this review was made.

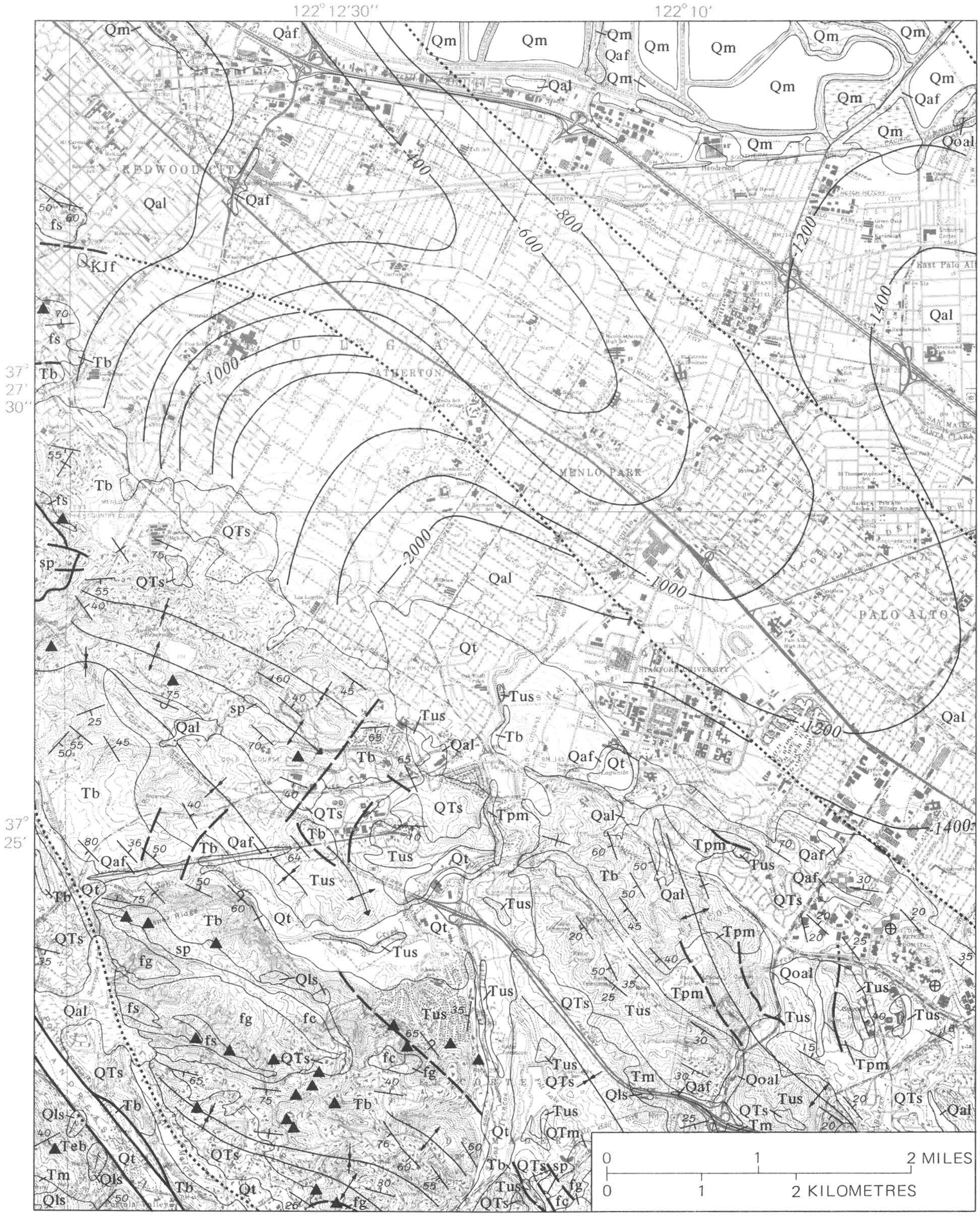


FIGURE 1.—Preliminary geologic map of part of the San Francisco Bay region showing types of land-use information. San Andreas fault zone (heavy lines) is shown in lower left corner. Contacts between rock units shown by thin lines; letter symbols are keyed to brief descriptions of the rock units in the map explanation (for example, Qls = large landslide deposits, Qm = bay mud). Triangles denote landslides too small to show at the scale of this map. Depth to bedrock shown by contours in upper half of map (after Brabb, 1970).

black). If white is seen, a zero is recorded, if black is seen a one is recorded. Thus output of the device is a series of zeros and ones with a coordinate implicit for each since the device scans in a regular way. One way is to scan from left to right in a horizontal line (0.001 in. wide), then move down 0.001 inch and scan the next line. Software processing then removes the zeros and attaches explicit coordinates to each "one." On the computer printout of a line from a small map section, an "x" is printed for each position where a "one" has been recorded (fig. 4*A*). The scale is greatly enlarged because each print position represents the small square the device sees each time it looks at the map—the 0.001 × 0.001 inch square. It is also possible to record not the zeros but instead the explicit position (x, y coordinates) of each "one" as the scanner sees it. In either method, the next step in software processing is to identify lines from the series of "ones" by obtaining the center line of each band of "ones" (fig. 4*B*).

Finally, a program replaces the series of "ones" with a set of straight line segments, each identified by endpoint coordinates and a direction. This produces a file of line segments as if the map had been digitized using an image digitizer (fig. 4*C*).

The use of scanners for the digitization of map information is just emerging from the research and development stage. Users have very little experience on which to base predictions of equipment cost and performance. IBM[2] developed a drum scanner for the Canadian Geographical Information System (Tomlinson, 1968). The material to be scanned is a traced zone boundary sheet. In this system a bit of information (zero or one) is recorded for each spot seen, with subsequent processing as described above. IBM subsequently developed a drum scanner for the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Va. (fig. 5) that can scan maps and separate eight colors from the manuscripts to individual digital codes. For a particular class of information (color), the particular code number is recorded for each spot seen, with subsequent software processing as described above for each class.

A microdensitometer is a device used to convert photometric data, as on film negative, to digital form for computer processing. These are commercially available and are used in research in many fields. Most models can be obtained with an optical aperture as small as 25 micrometres (0.001 in.) square, and some models can take a film as large as 14 × 17 inches (36 × 43 cm) for the input manuscript. They can discriminate between as many as 256 gray levels, each level having its own code for computer processing. Partial success has been achieved in digitizing maps by photographing the map and then using the negative for the input manuscript. This method can work very well if the map contains only one class of data. If the map contains more than one class of data, each class must have its own distinct and consistent gray level for subsequent separation in the software processing. Cost of these devices at present ranges from $22,000 to $25,000 for the basic unit.

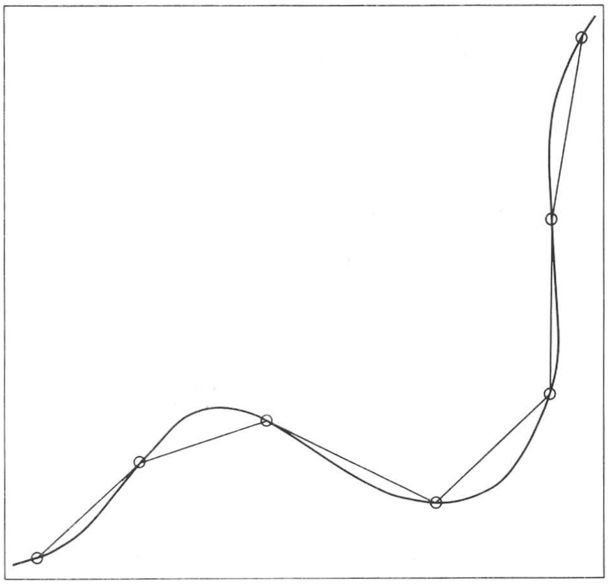


FIGURE 2.—Approximation of a curved line by a series of straight line segments.

The Oak Ridge National Laboratory is digitizing map information in connection with their environmental program.[3] Map overlays are prepared that contain only the information to be digitized. These overlays, or parts of them, are photographed by a 35-mm camera. Each 35-mm frame is then scanned by a Hough-Powell flying spot scanner-digitizer at the University of Pennsylvania. One horizontal line at a time is scanned, with the address (x, y) coordinates recorded each time the background changes from white to black or black to white. This technique is very useful for digitizing areal information. The areas to be recorded are

[2] Use of specific corporation or trade names is for descriptive purposes only and is not an endorsement of these firms or products by the U.S. Geological Survey.

[3] ORNL–NSF Environmental Program under NSF Interagency Agreement No. AAA-R-4-79.

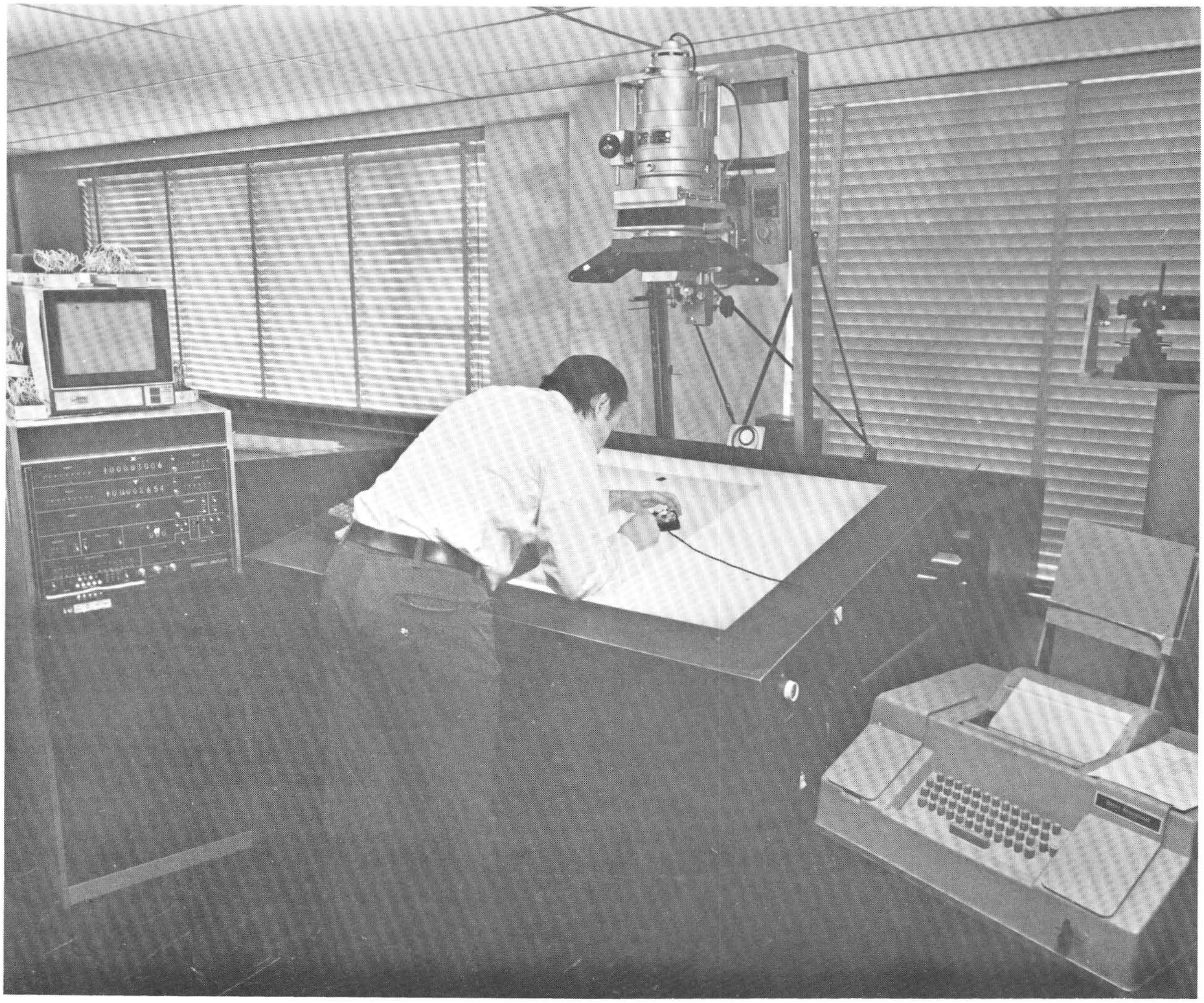


FIGURE 3.—Image digitizer used by the U.S. Geological Survey.

blacked in on the overlay. The scanner then records the boundary points of each zone for each scan line.

PROBLEMS ASSOCIATED WITH MAP DIGITIZATION

The process of map digitization requires a quality control procedure to ensure that the digitized data represent the map data. This procedure consists of outputting the digitized map data to a plotting device, reproducing the digitized lines and points on transparent paper. This can be overlaid on the map that was digitized to provide a visual check of the accuracy of the process. The use of a display tube on line with an image digitizer enables gross errors to be detected and corrected before the entire map is completed.

Ideally, for digitization, a map should have only one class of information—for example, elevation contour lines. Most maps, however, are complex, containing more than one class of information and cluttered with various symbols. These data must be identified before or during digitization.

With the use of image digitizers or automatic line followers, the operator involved is generally not familiar with the map being digitized. The process of digitization can be slowed considerably where lines serve multiple purposes (for example, where a fault line is also part of the boundary of a bedrock formation), where several classes of data must be digitized, or where there are gaps in the data or illegible data. Pre-editing the map manuscript may speed up the digitizer processes; however, for some maps it may be less costly to edit after digitization and correct the digitized data.

When new maps are to be made for digitization, it is obviously better to keep them uncomplicated. If digitization is performed with an optical scanner that distinguishes only between map background and information, it is necessary to eliminate data that are not to be digitized. Map overlays, one for each class of data to be digitized, can be used to extract the desired data. A plot of the digitized data can easily be compared with the overlay as a quality control check.

DIRECT ENTRY OF FIELD DATA

Most geographically related data will be entered into the computer by digitization of maps. However, the possibility of direct entry of field data into the computer should be considered for purposes of correcting existing entries or adding data where there are gaps in the record.

Smith and Berg (1973) have developed field forms for recording data that are specially designed to be read by a keypunch operator without further coding (fig.6). Only data entered in the boxes are keypunched. Similar forms have been designed for other kinds of data.

Certain optical scanners can recognize and read in typed symbols and some handprinted symbols. For example, the Geological Survey has one that recognizes the following handprint characters: "A C T X 1 2 3 4 5 6 7 8 9 0 + /" provided they are carefully made and of a certain size. If a field form could be designed requiring only these symbols and with boxes of the correct size, the scanner could read the sheets directly, thus bypassing the keypunch operation. Such a form should be field tested to see if it is practical. Alexander-Marrock, Friend, and Yeats (1970) have described a similar approach utilizing a mark-sensing field form.

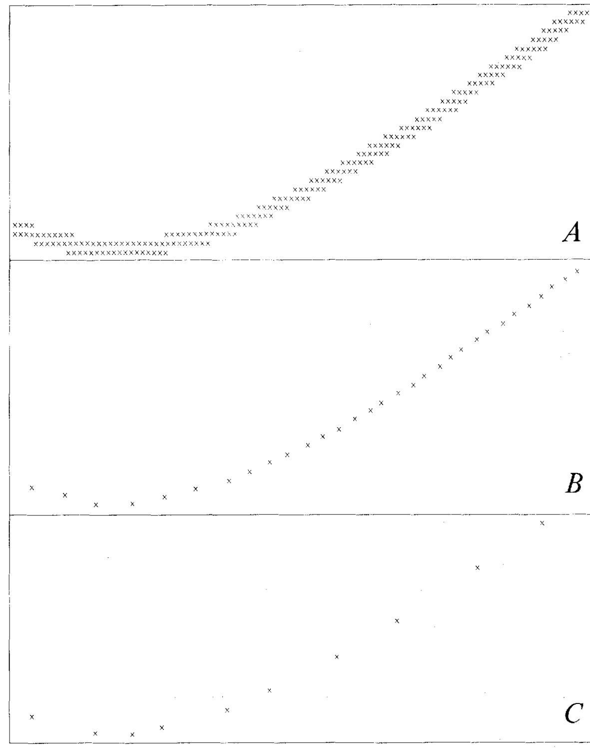


FIGURE 4.—*A*, Computer printout of scanned line. *B*, Center line of the scanned line. *C*, Scanned line represented by end points of straight line segments.

URBAN AND REGIONAL INFORMATION SYSTEMS

A study conducted by the U.S. Department of Housing and Urban Development (1968) distinguishes between planning- and operations-oriented information systems. Operations personnel are concerned with the present, planners with the future. Therefore, planners are interested in present data only to the extent that it can be used to predict the future. The fundamental difference between the two stems from the planning process. Planners deal with aggregates and are concerned with relationships between variables, while operations personnel must deal with individual entities. At present (1968, when the report was published) information systems mainly retrieve data and produce various summaries and statistical analyses; in the future they will estimate current states, predict future states, and develop alternatives. To accomplish this, they must use simulation and mathematical modeling.

It is possible for one computer information system to have both capabilities. The operations-oriented information system corresponds to the older business information systems discussed previously in the section on computer information systems, and the planning-oriented information system to the newer, evolving management information systems.

An urban and regional information system involves socio-economic data (for example, land use, population, income, employment, assessed valuation) that, to be effectively used, must reflect their geographical location. Therefore, such a system should combine capabilities of a management information system with the capabilities of a geographic information system. The specifications of any particular system will vary depending on user requirements.

DATA BASE ORGANIZATION FOR GEOGRAPHICAL INFORMATION SYSTEMS

The fact that a system is a geographical information system is reflected in its data base because the information regarding the geographical distribution of data resides there. The data base of such a system can be organized in several ways. In the following discussion, a record is defined as a collection of related items of data treated as a unit. A data file is a collection of related records. A data item is an individual item of information in a record. Data structure refers to the arrangement and interrelation of records in a file.

AREAL TYPE DATA

POLYGON OR AREA BOUNDARY

Consider a land-use map (fig. 7) in which the boundaries of the various land-use zones have been digitized as described in the previous section on earth science data entry. The set of line segments enclosing a zone constitutes a record of the area boundary enclosing that particular zone. Each enclosed land-use zone on the map has a separate record. All the land-use records from this map (and other maps, if more than one is needed to cover the region) constitute a file, the land-use file for the region. The set of line segments in a record is ordered so that a zone is always defined by going around it in the same direction, clockwise or counterclockwise. Each record contains other information regarding the zone—at least, a symbol representing its land-use category. For example, the zones numbered 1 and 6 might both be agricultural. For such a map, there would be other polygon files, one for each class of areal data (for example, soils or bedrock geology).

Two data files can be combined in the computer by an overlay process, which is analogous to overlaying two maps, both drawn on transparent material, so that the polygon boundaries on both maps may be seen and some particular combination of results studied. For example, a land-use map with four categories L1, L2, L3, and L4 and a

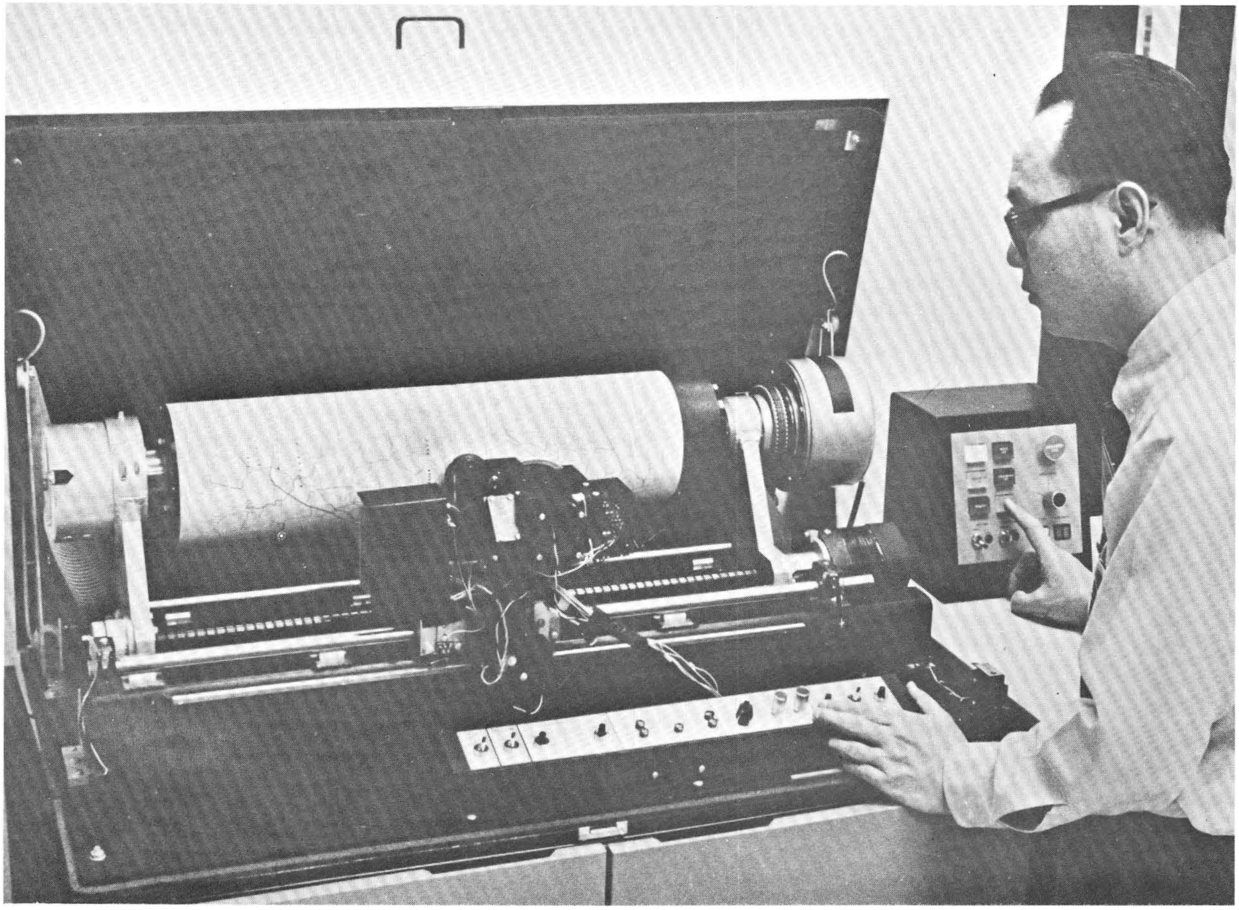
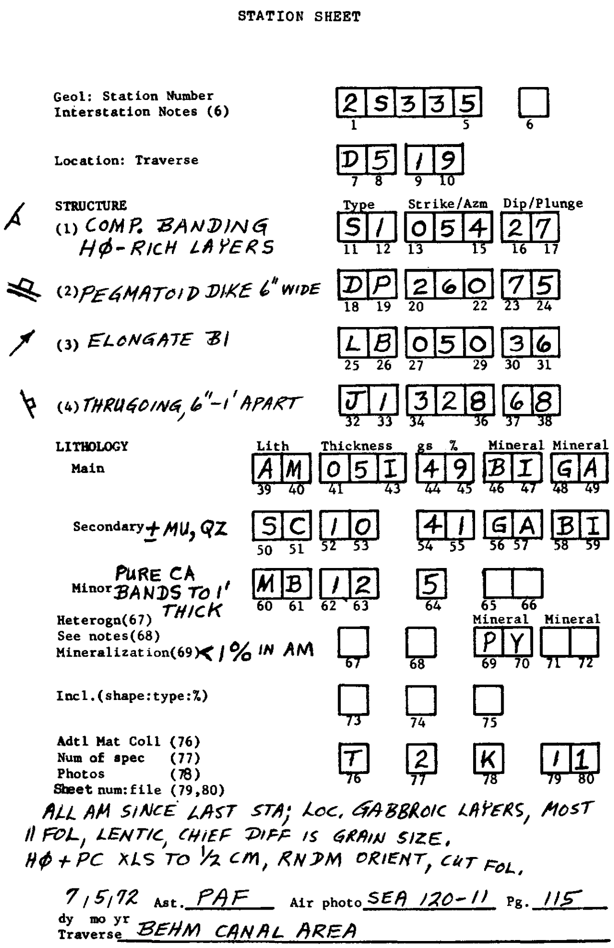


FIGURE 5.—IBM optical scanner.


STATION SHEET

Geol: Station Number 2 5 3 3 5 (1–5)
Interstation Notes (6)

Location: Traverse D 5 (7–8) / 9 (9–10)

STRUCTURE — Type, Strike/Azm, Dip/Plunge
(1) COMP. BANDING HØ-RICH LAYERS — S / (11–12) 0 5 4 (13–15) 2 7 (16–17)
(2) PEGMATOID DIKE 6" WIDE — D P (18–19) 2 6 0 (20–22) 7 5 (23–24)
(3) ELONGATE BI — L B (25–26) 0 5 0 (27–29) 3 6 (30–31)
(4) THRUGOING, 6"–1' APART — J / (32–33) 3 2 8 (34–36) 6 8 (37–38)

LITHOLOGY — Lith, Thickness, gs, %, Mineral, Mineral
Main — A M (39–40) 0 5 I (41–43) 4 9 (44–45) B I (46–47) G A (48–49)
Secondary + MU, QZ — S C (50–51) / 0 (52–53) 4 / (54–55) G A (56–57) B I (58–59)
Minor PURE CA BANDS TO 1' THICK — M B (60–61) / 2 (62–63) 5 (64) (65–66)
Heterogn(67)
See notes(68)
Mineralization(69) < 1% IN AM — (67) (68) Mineral P Y (69–70) Mineral (71–72)

Incl.(shape:type:%) — (73) (74) (75)

Adtl Mat Coll (76) T
Num of spec (77) 2
Photos (78) K
Sheet num:file (79,80) / 1

ALL AM SINCE LAST STA; LOC. GABBROIC LAYERS, MOST
// FOL, LENTIC, CHIEF DIFF IS GRAIN SIZE.
HØ + PC XLS TO 1/2 CM, RNDM ORIENT, CUT FOL.

7/5/72 dy mo yr Ast. PAF Air photo SEA 120-11 Pg. 115
Traverse BEHM CANAL AREA

FIGURE 6.—Station sheet for encoding field data to be keypunched into a computer.

soil map of four types of soil S1, S2, S3, and S4 (fig. 8) can be overlaid to show areas of soil type S1 and land-use category L1 (fig. 9*A*). All area within either S1 or L1 forms the boundary of their union (fig. 9*B*). All area common to both S1 and L1 is called their intersection (fig. 9*C*). This overlay process, determining the boundaries of a new set of polygons on the basis of unions or intersections, can be done in the computer. Any number of files can be combined by cascading the results, the third file being overlaid with the set resulting from overlaying the first and second files, and so on.

GRID STRUCTURE

A map of areal data having four categories A through D can be overlaid with a grid of north-south and east-west lines, forming a uniform cell structure (fig. 10). Each cell has a record, and the attribute of the predominant zone within the cell is assigned to the record for that cell. For example, the cell in row four, column one is B; the one in row two, column two is C. The cell structure need not be uniform. Quarter-quarter sections could be used, which are nonuniform in some places because of surveying inaccuracies.

If the cell size is large relative to the region, the areal maps could be hand encoded. If, however, the cell size is small relative to the region, hand encoding may be impractical. The grid file can then be created by first creating a polygon file and then, in the computer using suitable software, overlaying the grid structure and recording the data in each cell. When map digitization of areal data is done using the technique of the Oak Ridge National Laboratory as described in a previous section, it is possible to write software to assign the data directly to cells without first digitizing the area boundary.

If an areal unit is large relative to the cell size, many contiguous cells will have the same information, thus wasting file space if they are all stored (fig. 11). In such a situation a type of data compression can be used. Not all the cells overlaying the large polygon would be used. Only the first cell in each row wholly within the polygon could be recorded, with a notation as to the number of cells in that row within the polygon. Amidon and Akin (1971) discussed more sophisticated techniques of data compression.

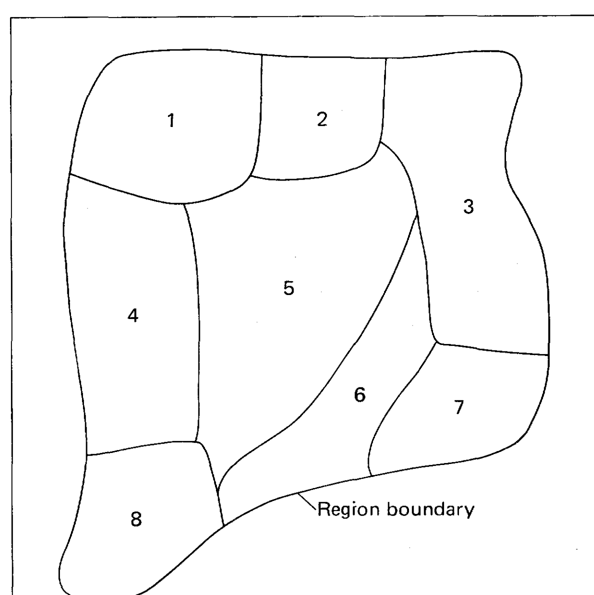


FIGURE 7.—Sample map of study region containing eight digitized land-use zones.

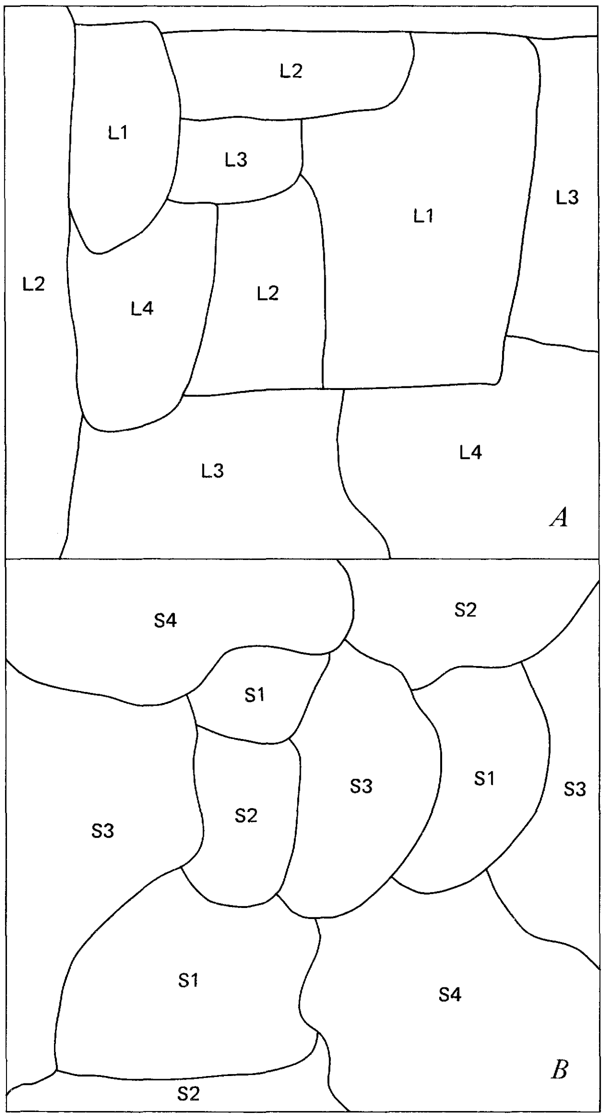


FIGURE 8.—*A*, Map of land-use zones. *B*, Map of soil types.

Thus far, each class of areal data has been considered as constituting a separate file. For the polygon area boundary file, there is one record for each polygon, one set of polygons for each file. For the grid structure, there is one record per cell, one set of cells (covering the region) for each file. However, with the grid structure, files of areal data can be combined by assigning all the data to a single file of cells, one record for each cell with each record containing data on all classes of areal data for that cell location. It is obvious that the more numerous the classes of data, the less frequent the opportunity for data compression, since it will be increasingly rare that all classes of data will have

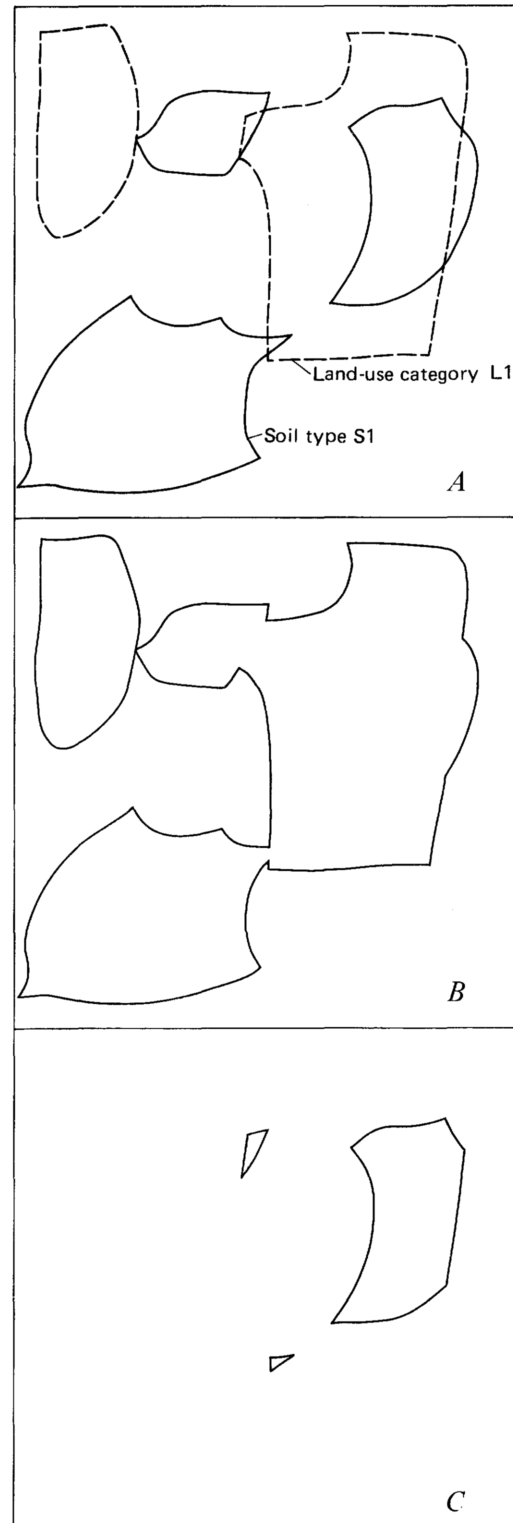


FIGURE 9.—Overlay of soil type S1 and land-use category L1. *A*, The overlay. *B*, The union. *C*, The intersection.

polygons of large area in the same geographical location.

INCORPORATING LINE AND POINT DATA

There are two kinds of line data, isopleths and networks. Isopleths are equal-value lines, such as elevation contour lines. In one sense isopleths are area boundary lines since they separate areas of higher from areas of lower value. To the extent that they are used as such they can be handled as described in the preceding paragraph. However, an isopleth has a definite value associated with it (for example, a 2,000-ft elevation contour line), whereas area boundary lines, in general, have none.

Networks are lines representing areas, such as streets or streams, that are too narrow to be represented as areas on a map. In a polygon type file, networks can be stored as area boundaries are, but the definitions of a file and a record are much more arbitrary. The storage of network information in a grid basis is much more difficult. The simplest method would be a code indicating the presence or absence of a particular class of data in a cell. For example, the cell record could have a field for stream data in which a "1" would indicate the presence and a "2" the absence of a stream in that particular cell. If this information were inadequate, the cell record could be enlarged to include more detailed information. A separate file for that particular network could also be maintained, organized as it would be in a polygon file, thus creating a hybrid data base.

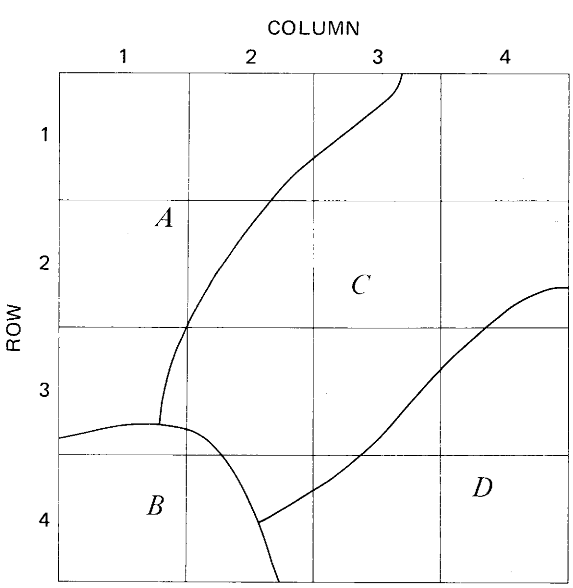


FIGURE 10.—Map with grid overlay.

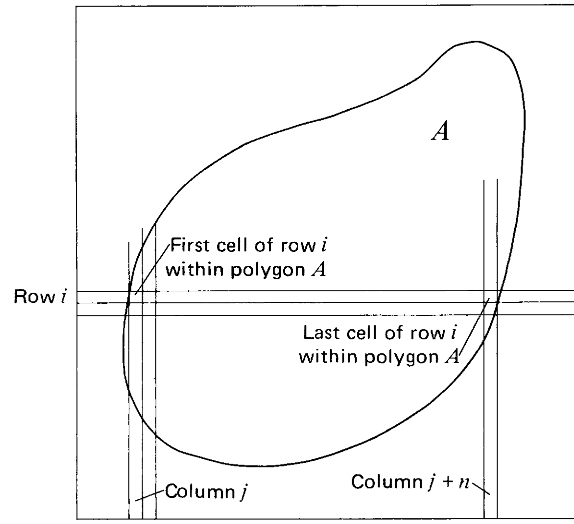


FIGURE 11.—Grid size small relative to polygon size.

The inclusion of point data in a geographical information system data base is not difficult. Some information can be added to a cell record, and a separate file for each class of point data, referenced by coordinates, is easy to create, use, and maintain.

DATA STRUCTURES FOR INFORMATION SYSTEMS

Thus far, only those aspects of data base organization resulting from the geographical nature of data have been discussed. Other aspects of data base organization relating to computer information systems in general should be mentioned. For example, suppose that information on all land in agricultural use were to be retrieved. If the data structure were organized on a polygon basis, the system would retrieve only the land-use file. Then it would search sequentially through the records of each polygon for those coded, for example, "OZ," for agricultural use; this could be quite time consuming. The retrieval time could be lessened considerably by inverting the file; an inverted file is a cross-indexed file in which a key word identifies a record. An index would list all polygons in the land-use file, grouped by land-use classification, with addresses of (pointers to) each polygon. Or, alternatively, an index table at the beginning of the land-use file would contain a list of pointers, one for each land-use classification. The pointers would indicate the first polygon of each land-use class, "OZ" in this example. The first "OZ" polygon would contain a pointer indicating the location in the file of the second "OZ" polygon, and so on. The file could also be inverted on other data items. An inverted file, or a similar technique is even more necessary for a cell-type data base structure since there will probably be many more cells in a region than there are polygons in any given file.

It is not possible in this paper to give a discussion of data base organization. Dodd (1969) has written an excellent introductory, nontechnical article on data base organization. The design of a computer data base can be complex and costly and should be undertaken by experienced computer scientists.

SOME SELECTED SYSTEMS

Two bay region systems were used as study models in determining the characteristics of land-use information systems: (1) the system of the Metropolitan Transportation Commission (MTC) and the Association of Bay Area Governments (ABAG) and (2) the system of Santa Clara County (SCC).

DATA BASE ORGANIZATION

On a diagrammatic representation of the computer information system of the MTC, the sections labeled "Urban growth and location" and "Transportation" represent modeling capabilities (fig. 12). The part labeled "Data base" consists of the system software and files of the four types of records:

1. Transportation facilities—organized by networks
2. Trip zones—organized by a table (290 × 290 matrix)
3. Records organized by map zones:
   a. One file for each class of information (dwelling unit, employment, population, and land use)
   b. A record for a particular file contains geographical information and class information. The geographical information is given first by county (1–9), then by map zone (1–290), then by tract (1–1040), with further refinement as necessary.
4. Unique locators. These are special cases organized as in 3, above.

Most of the data are from census tracts. The 1970 census for the nine-county bay region is divided

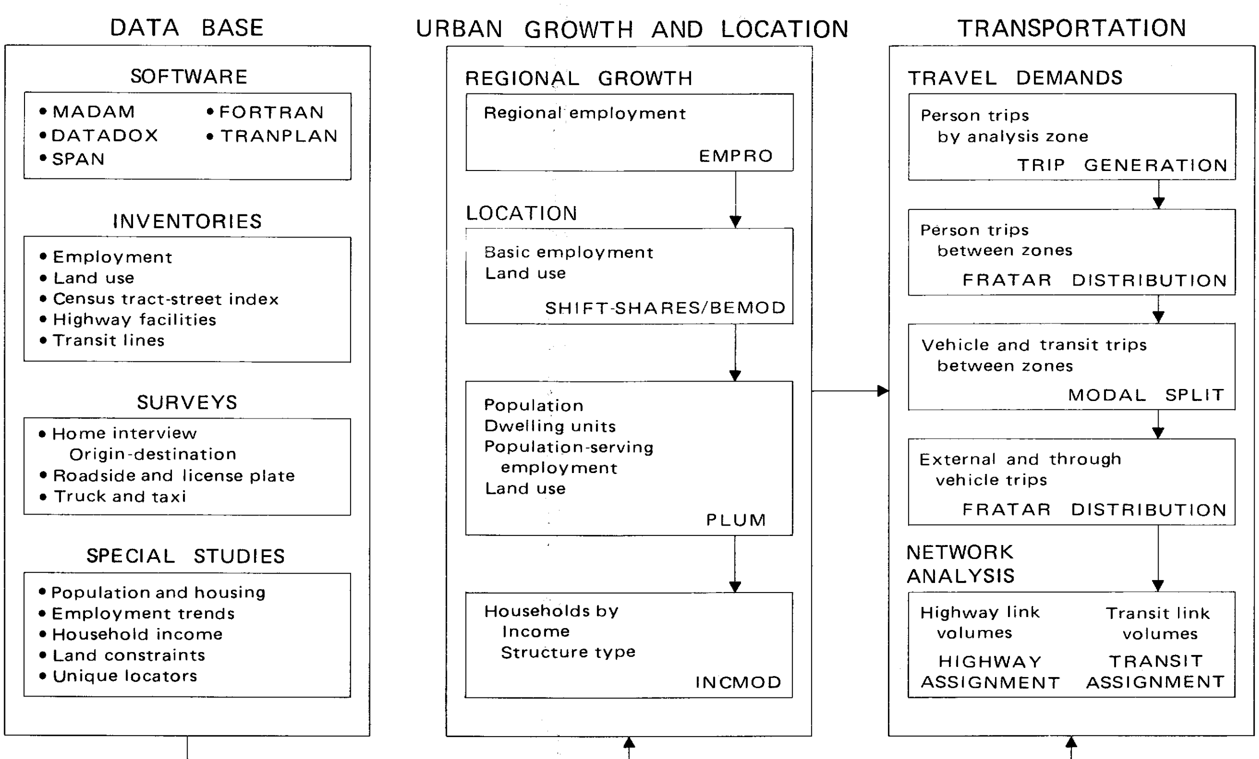


FIGURE 12.—Bay Area Transportation Study Commission modeling system.

into 1,040 tracts grouped into 290 tract zones (fig. 13).

The computer information systems of Santa Clara County use three types of files:

1. Property files (tax assessor) organized by a property location index. Each parcel of land has an assessor's parcel number and the coordinate of its estimated centroid (x, y) in the California state coordinate system.
2. Census files (1970) organized as follows: First count—by census block group with (x, y) of centroid in the California state coordinate system. Second and fourth count—by census tract with (x, y) of centroid in latitude and

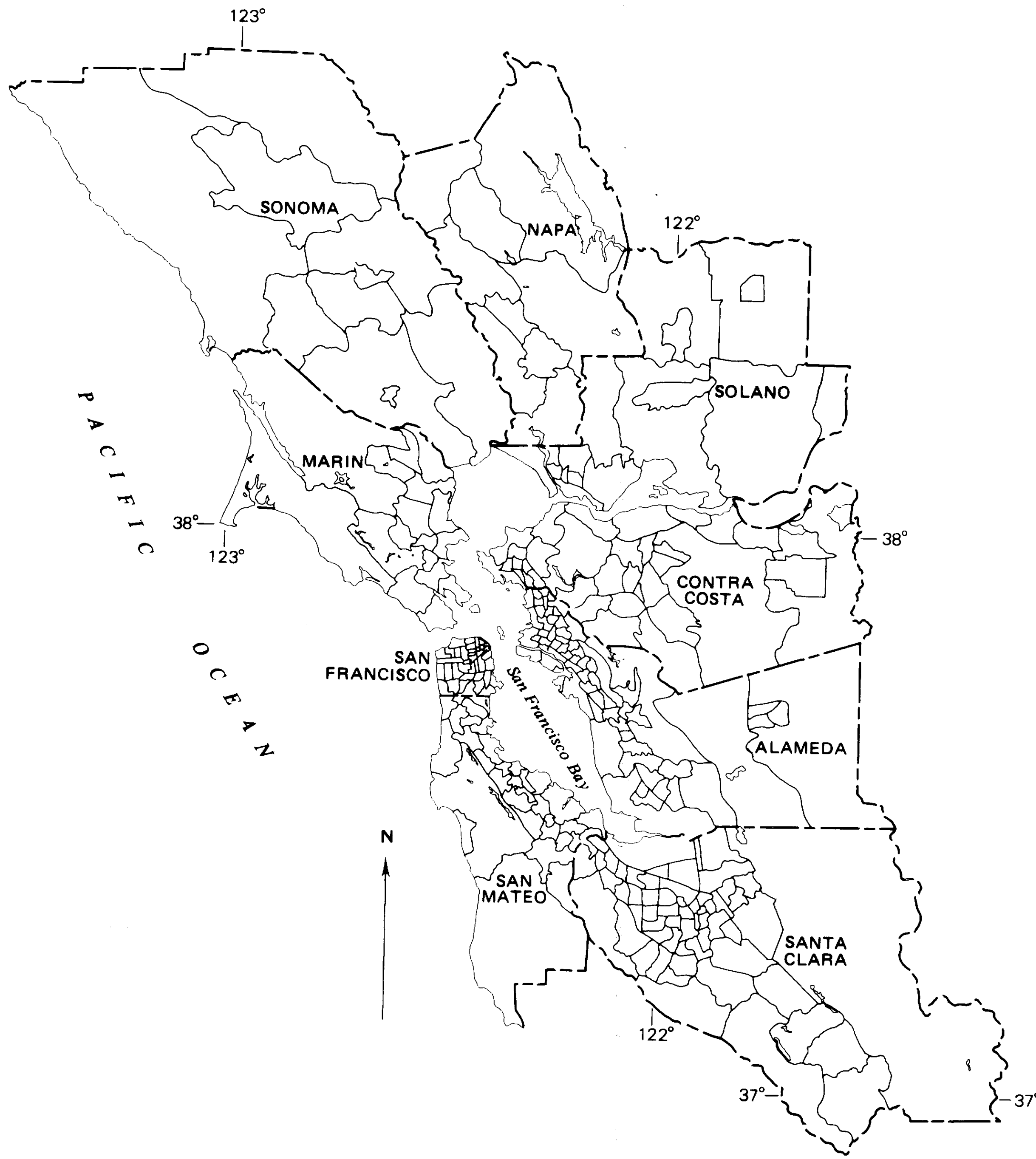


FIGURE 13.—Tract zones for the nine-county San Francisco Bay region.

longitude. Third count—by census block.

3. Files organized on the ACG–DIME system.

The ACG–DIME system was developed by the U.S. Census Bureau for use in metropolitan areas as a computerized geographic system for the 1970 census. ACG stands for address coding guide, DIME for dual independent map encoding. Each street intersection is a node, and each node is uniquely numbered. A street segment is that part of the street between nodes. To avoid unusually long or curved segments, nodes are created between street intersections where necessary. A DIME record is created for each street segment containing, among other data, the identifying nodes and, for each street, the address and the block and tract numbers to which that side of the street belongs. The (x, y) coordinate of the nodes can be added to the record. For more details the reader is referred to U.S. Bureau of the Census (1970). The census tract is one areal unit common to both the MTC/ABAG and the SCC system.

The Oak Ridge National Laboratory is developing a regional information system as a part of their environmental program, ORRMIS (Oak Ridge Regional Modeling Information System), sponsored by the National Science Foundation, Research Applied to National Needs Program. Its data base consists of nested cells referenced to a geodetic coordinate system (using latitude and longitude). The largest cell, 30 seconds latitude × 30 seconds longitude, is the reference cell. Nested within it are cells 15 seconds × 15 seconds, and within these are cells 3.75 seconds × 3.75 seconds. The size of these cells can be appreciated by comparing them with familiar units. At the latitude of the San Francisco Bay region, the reference cell is about 165 acres (67 ha) and the 3.75-second cells are about one hectare (10,000 m^2). Different data elements are assigned to the various cells depending upon resolution, accuracy, and type of the data. A slope map is not digitized, but elevations are stored on an approximately 10 × 10 m grid, and slopes are computed as required.

SOFTWARE CAPABILITY

MTC uses a general-purpose software package developed by System Development Corporation for large-scale data management and analysis. It consists of several individual processor packages and is able to handle data retrieval, file manipulation and data reduction, matrix operation, statistical analysis, report generation, and graphic display. One of the processors, ICORR (Internal Correspondence), is used for the point-in-polygon evaluation whereby entities with geographical coordinates (for example, street address associated with census data) are assigned to study areas such as census tracts. The MTC system is described in detail by Kevany (1968).

The regional growth forecasts and land-use projection part of the MTC system uses five models (fig. 12):

1. EMPRO—forecasts regional employment growth in industrial categories, given certain trend data and assumptions about the future.
2. SHIFT–SHARES—allocates regional employment estimates in the basic employment classes to individual counties.
3. BEMOD—allocates basic employment to analysis zones within each county.
4. PLUM—(a) allocates residential population in households to analysis zones, relative to the location of employment and other factors, and (b) locates population-serving employment by zone, relative to the location of residential population, basic employment, and other factors.
5. INCMOD—projects households into income categories, then into dwelling-unit structure classes. The output of these models is at the tract zone level. The transportation demand and network analysis models shown in figure 13 (based on results of the region growth models listed above) also output at the tract zone level. These models are discussed by the Bay Area Transportation Study Commission (1969).

Santa Clara County is currently developing an interactive facility (Cristiani and others, 1973). Geographically related data are referenced to 334 traffic zones. These are related to census tracts so that each traffic zone is wholly within a census tract and each tract is made up of one or more traffic zones. Files can be displayed on a display tube in either map or tabular form. They can be manipulated by use of a light pen and the keyboard. Modeling is performed using a household and commercial model with a regional forecasting model (economic growth model, demographic model, industrial location model) as the driving function. The user can change data relating to a policy decision, rerun the model, and check results. Output from the model is stored as data for reruns at 5-year intervals.

The Oak Ridge National Laboratory is engaged in a regional modeling effort that began in 1970 (Oak Ridge National Laboratory, 1971; Meyers, 1971). First, a county demonstration model was developed made up of a series of linked submodels—socioeconomic, land-use simulation, intrinsic land-use suitability, air diffusion, and soil loss prediction. This model is being used to study Campbell County, Tenn. A comprehensive regional model is being developed for a 16-county, 6,500-mi^2 (17,000 km^2) region surrounding Knoxville, Tenn. The data base for this larger model is being constructed.

UTILIZATION OF EARTH SCIENCE DATA

NATURAL RESOURCE INFORMATION SYSTEMS

A few geographic information systems include only natural resource and land-use data. One of the oldest is the Canadian Geographic Information System, a very ambitious system for all of Canada, which was started in 1963 and implemented 1965–71. The data base, organized on the area boundary or polygon system, has files on present land use, capacity of the land for four categories (agriculture, forestry, recreation and wildlife), and watersheds, and has data manipulation capabilities to calculate areas and perform overlay operation. It operates on the batch mode, and experienced programmers are required (Boyle, 1972; Tomlinson, 1968).

GRIDS (Guided Resource Inventory Data System) was developed by the Department of Natural Resources of the State of Washington. The objective of the system is to maintain a current land inventory for use in making operational plans for land managed by the department. The data base is organized as a regular grid structure with a cell 660 feet square (10 acres, ¼ of a ¼ × ¼ section; 4.05 ha). There is one record per cell with the following information: land cover, slope, aspect, elevation, soil, watershed, road access, land use, forest cover (tree type) and nonforest type (R. A. Harding, written commun., 1972).

NARIS (Natural Resources Information System) was developed by the Center for Advanced Computation, University of Illinois at Urbana-Champaign; it covers parts of eight counties in northeast Illinois. Its data base is an irregular grid structure, each cell being a ¼ × ¼ mile section (40 acres, 16 ha). To date, each tract (cell) contains 15 classes of information under five major headings:

Geology:
1. Interpretation of waste disposal
2. Interpretation of construction
3. Water resources
4. Sand and gravel resources
5. Surficial deposits

Land use:
6. U.S. Department of Housing and Urban Development codes
7. Northeastern Illinois Planning Commission codes

Forestry:
8. Native woody vegetation
9. Planted woody vegetation

Soil:
10. U.S. Dept. of Agriculture Soil Conservation Service soil characteristic

Water:
11. Watershed
12. Wells
13. Present impoundments
14. Future impoundments
15. Streams

Each of these 15 classes is, in turn, made up of data elements. The values of these data elements are the attributes of each tract. For example, soil information is collected by plots of various size within each tract. Each plot is described by the following data elements:

NUMBER—the type of soil in the plot;
SLOPE—the slope of the soil in the plot;
EROSION—the current erosion of the soil in the plot;
ACRES—area of the plot in acres;
OVERLAP—denotes whether or not the plot extends into an adjacent tract.

The objective of the system is to provide natural resources information to public agencies and to the private sector. It can be user operated in a conversational mode with its own language (Center for Advanced Computation, 1972b).

The Raytheon Company (1973) is currently developing a natural resources information system for the Department of the Interior (Bureau of Indian Affairs and Bureau of Land Management). It is to include natural resources and land-use information and will be used as an aid in the management of public lands. System specifications call for the ability to store and retrieve map data and for the capability of data manipulation such as overlaying maps to obtain unions or intersections of specified types of data and making area and perimeter calculations. The data base currently being developed is a variation of the area boundary structure. However, instead of storing complete polygons, the system stores only arcs, which are parts of an area boundary between two branch points. For example, polygon I consists of arcs ab, bc, cd, de, and ea, and polygon II consists of dc, cf, fg, and gd (fig. 14). An index table in the data base gives (for each file) the arcs making up each polygon. When maps (files) are overlaid, new arcs are created whenever an arc from one file crosses an arc of another file. Also new polygons are created, and a new table of polygon arcs is written. It remains to be seen how effective this data base organization is.

Boyle (1972) and Center for Advanced Computation (1972a) give additional information on other natural resources information systems.

PROGRAM INVENTORY AND GEOLOGIC MODELING

Earth science information in its most commonly available form consists of maps and tables of physical, chemical and engineering properties, which are the basic data on page 4. They are developed from raw data (field observations and actual measurements of various physical and chemical properties) with the benefit of scientific interpretation. These basic data are still very technical and generally require further interpretation by scientists or engineers to be useful to the planner. This interpretation of basic data has been one of the main features of the San Francisco Bay Region Study Program. The inventory of these studies listed on page 4 is now presented in a slightly different form (table 2). The left-hand column contains basic data. Most of these data are generated routinely by the U.S. Geological Survey, although not uniformly for all areas of the country. The right-hand column lists results of the interpretive studies. The center column lists the basic data that would, ideally, be used in the interpretive study on the same row to the right. In some instances all of the basic data so listed is used. In others, only part of it is used, depending on the area being considered, the availability of the basic data for that area, and the current methods involved in the particular study. Most of the items in the right-hand column are generally not developed routinely by the survey but are the outgrowth of special requirements developed on the San Francisco Bay Region Study Program. The data in the right-hand column plus some items in the left-hand column can be used for

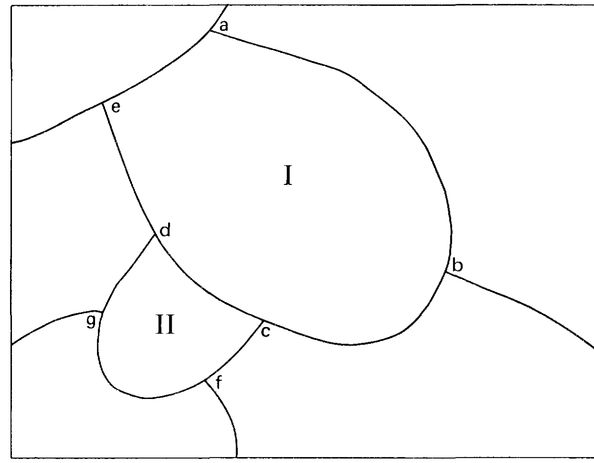


FIGURE 14.—Polygons made up of arcs.

the evaluation of land-use plans, especially to avoid hazards.

The development of most interpretive data from basic data is, in a sense, a form of mathematical modeling or simulation. Some of the interpretive studies consist mainly of overlaying basic data and interpreting the overlays (McHarg, 1969), which can be performed in the computer. To the extent that interpretation of the overlays can be reduced to a table look-up procedure or a mathematical formula, this can also be performed in the computer.

On the landslide susceptibility map prepared for

TABLE 2.—*Program inventory*

Basic data	Data used in interpretive studies	Interpretive studies
1. Bedrock	1, 3, 4, 6, 15, 17, 18, 19	A. Landslide susceptibility.
2. Unconsolidated deposits	1, 2, 4, 5, 7, 8, 12, 19, I	B. Land use implications—region of bay mud.
3. Landslides		
4. Active faults	1–6, 9, 11–15, 17–20, I	C. Coastal geologic processes.
5. Topographic map		
6. Slope map	2, 4, 5, 7–10, 12, 14, 17, 18, 19, I	D. Physical properties of unconsolidated deposits and land-use implications.
7. Bay mud		
8. Land subsidence	1–4, 6, 11, 12, 17	E. Hillside materials—inferred engineering behavior and land-use implications.
9. Ground-water		
10. (a) Stream networks (b) rating tables.	5, 10, 14, 15, 19	F. Flood-prone areas and land-use implications (boundary of 100-year flood).
11. Water bodies		
12. Earthquake locations	5, 10, 14, 15	G. Urban drainage system (technical report).
13. Mineral commodities		
14. Historic flood data	1, 6, 9, 11, 15, 16, 17, 19, A, F	H. Land pollution susceptibility.
15. Rainfall (isohyetal map)		
16. Evaporation data map	1–4, 7, 12	I. Seismic response map.
17. Soils		
18. Vegetation	1–5, 9, 10, 11, 15, 16, 18, 19	J. Ground water studies.
19. Land use		
20. Water quality and suspended sediment.	1, 2, 3, 5, 6, 10, 11, 14, 15 17–20	K. Erosion and sediment studies.

San Mateo County, only three basic data items are used: bedrock, landslides, and slope categories. Seven categories of susceptibility to landslides are defined ranging from lowest to highest, according to the following list (Brabb and others, 1972):

I. Areas least susceptible to landsliding. Very few small landslides have formed in these areas. Formation of large landslides is possible but unlikely, except during earthquakes. Slopes generally less than 15 percent but may include small areas of steep slopes that could have higher susceptibility. Includes some areas with 30 percent to more than 70 percent slopes that seem to be underlain by stable rock units.

II. Low susceptibility to landsliding. Several small landslides have formed in these areas, and some of these have caused extensive damage to homes and roads. A few large landslides may occur. Slopes vary from 5 to 5½ percent for unstable rock units, to more than 70 percent for rock units that seem to be stable.

III. Moderate susceptibility to landsliding. Many small landslides have formed in these areas, and several of these have caused extensive damage to homes and roads. Some large landslides likely. Slopes generally greater than 30 percent but includes some 15–30 percent slopes in areas underlain by unstable rock units.

IV. Moderately high susceptibility to landsliding. Slopes all greater than 30 percent. Several large landslides likely. These areas are mostly in undeveloped parts of the county.

V. High susceptibility to landsliding. Slopes all greater than 30 percent. Many large and small landslides may form. These are mostly in undeveloped parts of the country.

VI. Very high susceptibility to landsliding. Slopes all greater than 30 percent. Development of many large and small landslides is likely. These areas are mainly in undeveloped parts of the county.

L. Highest susceptibility to landsliding. Consists of landslide and possible landslide deposits. No small deposits are shown. Some of these areas may be relatively stable and suitable for development, whereas others are active and causing damage to roads, houses, and other cultural features.

In developing the susceptibility map, the map of landslide deposits is first overlaid on the map of bedrock geology. Those areas with landslide deposits are zones in category "L." Next, the result is overlaid with a slope map of six slope categories. Those areas not categorized as zone "L" are categorized according to table 3 (from Brabb and others, 1972).

If it were possible to generate the results of all the interpretive studies by simulation or modeling, it would be feasible to have the computer data bank consist mainly of basic data. Interpretations could be generated as required. This approach has several advantages. Owing to the pressure of urbanization and its impact on the environments, interest in the type of data generated by the interpretive studies will grow, fostering the development of better geologic models so that a data base of interpretive data will probably have to be updated frequently. However, the basic data are relatively stable, requiring much less updating. As areas of application grow, the amount of data generated by models will probably result in a greatly enlarged data base if it consists of interpretive data. Finally, there is a great advantage if, instead of a digitized slope map, the data bank contains elevations at grid intersections of suitable size so that slopes can be computed as needed. Then models requiring slope information as an input are not locked into preselected slope categories.

EARTH SCIENCE DATA FOR INFORMATION AND PREDICTIVE LAND USE

Large-scale regional land-use models are developing a need for earth science data owing to the growing requirements to study the environmental impact of various land-use alternatives. The Oak Ridge National Laboratory, in developing the models for their environmental program, are providing for this need by including the necessary data, mostly assigned to the level of the 30 × 30 second cell. The data include geology (bedrock and surficial, mineral deposits, aquifers), surface water information, and soil types. Tables of physical and chemical properties are included in the data base (C.R. Meyers, Jr., oral commun., 1972.) The Center for Advanced Computation of the University of Illinois included the following data in the data base of the recommended Illinois Resources Information System Center for Advanced Computation, 1972a):

Geology	
Surface deposits	Digitized area.
Oil, gas, and water wells	Point.
Oil and gas fields and storage	Section.
Oil and gas statistics	Township.
Coal statistics	Point.
Soil	
Soil associations and interpretations	Section.
Soil types	County.
Soil productivity	County.
Soil and water	County.
Conservation needs	Watershed.
Land use	
Aerial photography	1:24,000.
Land-use interpretation	¼ × ¼ section.
Water	
Stream network digitization	Network.
Water quality standards	Network.
Pollutant emissions	Point.
Surface water quality	Point.
Ground water quality	Point.
Public water-supply quality	Point.
Flood plans	Digitized area.
Flow duration and low flow frequencies	Point.
Time of travel	Network.
Public water-supply adequacy	County, municipality.

EVALUATION AND RECOMMENDATIONS

The need for user participation in the design of a computer information system is well recognized. In the design of an urban and regional land-use information system, participation of users must be preceded by identification of the using community, which may include not only various government bodies but also the private sector. The larger the region, the more varied the potential users of the system. Totschek, Almendinger, Kevany, and Needham (1970) discussed the need for a common information center for the nine-county bay region, its relation to the various components of the user community, and a recognition of the differing data requirements of the user. However, the report is concerned only with socioeconomic information and does not consider earth science data. The Center for Advanced Computation of the University of Illinois (1972a) recommends a statewide information system. A survey of user needs indicated five potential user groups: (1) social service agencies, (2) the economic and industrial sector, (3) the physical services sector (for example, transportation), (4) the natural resources sector, and (5) agricultural agencies.

Few land-use decisionmakers have much knowledge of earth science or the role it should play in land-use planning; many have little knowledge of computer technology. Therefore, in the design of a specific system, after the user community has been identified, considerable interchange of knowledge may be necessary between users, computer scientists, and specialists in the field of earth science technology, primarily geologists and civil engineers. Any design effort that bypasses or weakens this step may waste much time and money.

An urban and regional land-use information system with a large and diverse user community will require a wide range of capabilities. The main characteristics of such a system are:

1. *Interactive and batch mode capability*—Many of the demands on the system can be met by using programs run in the batch mode, such as routine reports and simulation studies where no outside decisions are required. However, the ability to interact with the system so as to make decisions during a simulation program enables planners to get a better feel for the results of various decisions. This is especially true where geographical data can be shown on a display tube in its spatially distributed form. If a storage-type display tube is used, overlays can be performed on the screen of the tube without the necessity of performing the calculations in the computer. It should be mentioned that with an interactive system, special user languages should be available so that the user will not require a programmer in order to communicate with the system.
2. *Large data manipulation and computational capabilities*—Large-scale urban and regional information systems will be required to do much more than provide fast retrieval of large amounts of data for printout or display. The system must also be capable of sorting and recombining the data in various ways for analysis and reporting. Closely allied with this is the requirement for statistical analysis. Finally, a capacity to do large-scale simulation and modeling is necessary.
3. *Multiple-type input and output devices* — Programs run in batch mode imply a hard copy output on a printer or a plotter or both. Work in the interactive mode implies the use of

TABLE 3.—*Landslide failure record for rock units in San Mateo County*

[Modified from Brabb and others, 1972]

Rock unit on geologic map in order of increasing proportion of surface having failed by landsliding	Map symbol	Approx. area in county (mi²)	Approx. area that has failed (mi²)	Relative suscep-tibility numbers	Susceptibility numbers in each slope interval					
					0–5°	5–15°	15–30°	30–50°	50–70°	>70°
(No data for surficial deposits, undivided, Qu; alluvium, Qal; San Francisco Bay mud, Qm; windblown sand, Qd; beach deposits, Qb; artificial fill, Qaf; terrace deposits, Qt; Page Mill Basalt, Tpm; unnamed volcanic rocks, KJv; marble, m; shale near Palo Alto, Ksh; conglomerate, fcg; or metamorphic rocks, fm, but extent of landsliding probably small)				I						
Limestone	fl	.30	.00		I	I	I	I	I	I
Colma Formation	Qc	11.11	.01		I	I	I	I	I	I
Sandstone at San Bruno Mountain	KJs	4.77	.10		I	I	I	I	I	II
Butano(?) Sandstone	Tb?	10.56	.19		I	I	II	II	II	II
Unnamed sandstone	Tus	1.81	.04		I	I	I	I	I	II
Granitic rocks	Kgr	24.61	.90		I	I	I	I	I	II
Serpentine	sp	4.76	.09	II	I	I	I	II	II	II
Sandstone of Franciscan assemblage	fs	22.19	.74		I	I	I	II	II	II
Slope wash and ravine fill	Qsr	4.51	.18		I	I	I	II	II	II
Greenstone of Franciscan assemblage	fg	11.70	.61		I	I	I	II	II	II
Chert of Franciscan assemblage	fc	1.43	.10		I	I	I	II	II	II
Lompico Sandstone	Tlo	.40	.03		I	I	I	I	I	II
Sheared rocks of Franciscan assemblage	fsr	9.99	.83		I	I	II	III	III	III
Pigeon Point Formation	Kpp	7.77	.84		I	I	II	II	III	III
San Lorenzo Formation, undivided	Tsl	1.00	.11		I	I	I	II	II	III
Merced Formation	QTm	7.91	1.01	III	I	I	II	II	III	III
Sandstone, shale and conglomerate	Tss	3.34	.51		I	I	I	I	II	III
Butano Sandstone, Sky Londa area	Tb	22.55	4.33		I	I	II	III	III	III
Santa Clara Formation	QTs	6.67	1.85		I	I	III	IV	IV	IV
Rices Mudstone Member of San Lorenzo Formation	Tsr	1.37	.43		I	I	II	III	IV	IV
Vaqueros Sandstone	Tvq	7.60	2.41		I	I	II	III	III	IV
Monterey Shale	Tm	5.11	1.76		I	I	II	IV	IV	IV
Purisima Formation, undivided	Tp	23.06	7.81	IV	I	II	III	IV	IV	IV
Lambert Shale	Tla	19.95	7.25		I	I	II	III	III	IV
Mindego Basalt and other volcanic rocks	Tmb	10.80	4.01		I	I	II	III	III	IV
Butano Sandstone along Butano Ridge	Tb	20.18	7.66		I	I	II	III	IV	IV
Santa Cruz Mudstone	Tsc	19.25	7.98		I	I	I	III	IV	IV
San Gregorio Sandstone Member of Purisima Formation	Tpsg	2.41	1.06		I	I	I	IV	V	V
Tunitas Sandstone Member of Purisima Formation	Tptu	2.76	1.24		I	I	III	V	V	V
Tahana Member of Purisima Formation	Tpt	33.46	16.08	V	I	II	III	V	V	V
Pomponio Member of Purisima Formation	Tpp	11.97	5.76		I	I	II	V	V	V
Twobar Shale Member of San Lorenzo Formation	Tst	.80	.42		I	I	II	III	IV	V
Santa Margarita Sandstone	Tsm	.65	.41		I	I	I	III	III	VI
San Lorenzo Formation and Lambert Shale, undivided	Tls	6.83	4.56	VI	I	I	III	V	VI	VI
Lobitos Mudstone Member of Purisima	Tpl	3.71	2.57		I	II	II	VI	VI	VI
Landslide deposits	Qls	83.88	83.88	L	L	L	L	L	L	L

either a keyboard, both for data entry and system replies, or possibly a display tube for output. Data and program input may be by keyboard, the conventional keypunch and card reader, or optical scanners.

4. *Support of many terminals*—Terminals for both updating the data bank and obtaining system output may be spread out within the region. This does not necessarily imply time-sharing, although that may be most efficient.
5. *Support of multiple data bases*—Some data bases will be geographical in nature, others will have no spatial distribution. Among those with spatial distribution, some may be organized on a node and segment basis, such as an ACG–DIME census file or a stream network, while areal data may exist in a file of cells. Software must provide for handling and combining all of these data bases.

Special user languages were mentioned above. An example is included at this point for the benefit of those who have not used one. The language is DIRAC, described by Askevold and Vallee (1973). A file of all mineral deposits in Montana has been restricted to a subset of 25 properties. This subset is interrogated by a series of commands to find properties that satisfy the following criteria: (1) contains zinc, (2) contains silver, (3) occurs in an igneous or igneous-metamorphic area, (4) contains gold, (5) contains copper, (6) occurs along east-west structures, and (7) located in Jefferson County. Figure 15 shows how the interrogation would appear either typed on teletypewriter or shown on a display tube. Another excellent example is given by the Center for Advanced Computation (1972b).

A particular system will have a mixture of

```
   query
      25 ENTRIES
? find ELEM Zn!
      18 ENTRIES
? hold  1
? file   0
      25 ENTRIES
? find   ELEM Ag!
      22 ENTRIES
? hold  2
? file   0
      25 ENTRIES
? find  ROCK IG
      15 ENTRIES
? or ROCK IM
      18 ENTRIES
? and  ELEM Au!
      18 ENTRIES
? and  ELEM Cu!
      16 ENTRIES
? and  GEOL EW
       2 ENTRIES
? and  LOCA JF
       1 ENTRIES
? type
EW Carbonate Chief.  Alhambra-Warm Springs.  Jefferson County
             <Sec.  29, T. 8 N., R. 2 W.>
              Au 8,055 oz.
              Ag 66,037 oz.
              Cu 76,364 lb.
              Pb 403,584 lb.
              Ore minerals: galena, chalcopyrite, arsenopyrite, pyrite
              Gangue:    quartz
              Host rock:    quartz monzonite
              Alteration:
? file  0
      25 ENTRIES
?
```

FIGURE 15.—Example of an interactive language (part 1). A series of commands and responses in which two subfiles (1 and 2) are retained for later use (see fig. 16) and data are typed out for one property that satisfies seven criteria specified through a series of logical commands (actual terminal listing) (G. V. Askevold and J. F. Vallee, 1971, unpub. data).

capabilities determined by its own user community under the influence of the natural environment of the region. When the recommended system for a region is too expensive for immediate consideration, which is most likely, it should be designed to evolve in a planned fashion. Boyle and Sharp (1972) give three methods of developing equipment systems for spatial data processing, which are similar to regional information systems:

1. Use a large computer on a service basis with normal I/O equipment.
2. Purchase low-cost equipment as an interim measure and add to or replace with a larger system later.
3. Purchase expensive high-quality equipment at the start.

```
      file  1
          18 ENTRIES
  ?  corr  DIST DIST
  ?  draw 0
              DIST
                        - - -  *                                                        %
     3      16.6    WINS   ********                                                     %
     2      11.1    CAMO  *****                                                         %
     1       5.5    COCL   **                                                           %
     1       5.5    NIMI    **                                                          %
     1       5.5    CURL   **                                                           %
     2      11.1    ALHA  *****                                                         %
     2      11.1    AMAZ  *****                                                         %
     6      33.3    BAS1    ****************                                            %

  ?  file  2
          22 ENTRIES
  ?  corr  ROCK  ELEM
  ?  type
                           ROCK                          ELEM

                              ----      Au       Ag       Pb       Zn       Cu       Fe
                          ------------------------------------------------------------------
   5     4.5%     ----          0        1        1        1        1        1        0
  70    63.6%     IG            3       14       14       14       11       14        0
  10     9.1%     SD            1        1        2        2        2        2        0
  15    13.6%     IM            3        3        3        2        2        2        0
   5     4.5%     IS            2        0        1        0        0        1        1
   5     4.5%     CX            0        1        1        1        1        1        0
                          ------------------------------------------------------------------
   TOTAL=   110.                9       20       22       20       17       21        1
                              8.2     18.2     20.0     18.2     15.5     19.1      0.9
   ?                      ------------------------------------------------------------------
```

FIGURE 16.—Example of an interactive language (part 2). (Top) A histogram showing the distribution of zinc-producing mines (file 1) in the various mining districts. (Below) A correlation matrix that relates elements to host rock type for the silver-producing properties (file 2) (G. V. Askevold and J. F. Vallee, 1971, unpub. data).

The rapid advances in hardware technology are leading to rapid declines in the cost of certain hardware items and to rapid changes in ideas of urban and regional system design (table 1). The development of low-cost disc storage devices, with more dramatic developments promised for the future, are radically changing our concepts of what constitutes a large data bank for practical use. More recently, the development of the minicomputer is making possible the design of many systems requiring a small, inexpensive computer. A minicomputer can be used with several image digitizers and display tubes for online editing of the map digitization process. When the amount of calculation is small, a minicomputer with a storage display tube can form a small system for display and analysis of geographical data, including overlaying of areal data files.

While these developments make large-scale urban and regional information systems more economically feasible than was imagined 10 years ago, they also make if difficult to estimate costs for a system unless it is currently under development. Also, a cost estimate cannot be made in general, but only for a specific system. The Canadian Geographic Information System cost over $3,000,000 with data acquisition costing over $20,000,000 (Center for Advanced Computation, 1972a). Cost of the IRIS system is estimated at $4,000,000 for the computer system and $1,500,000 for the initial data base.

The original cost of the data base may be very great. Obviously, the greater the detail, the greater the costs. Taylor (1971) distinguishes four types of planning as regional, monitor, community, and ordinance administration and estimates their data needs as follows:

Characteristic	Region	Monitor and early warning	Community	Ordinance administration
Study area size	Total region	Total region	100 mi² (64,000 acres)	1,000 acres.
Largest subarea	200 acres	200 acres	1 acre	Parcel.
Preferred subarea	40 acres	40 acres	Subparcel	Subparcel.
Typical activity code levels.	2 digits	2 digits	3 digits	4 digits.
Updating	Annual	Monthly	Monthly	Weekly.

He then relates accuracy requirements to these systems (fig. 17). The cost of establishing and maintaining a computer data base to support a geographical information system is not linearly related to accuracy but rather rises very rapidly as accuracy approaches 100 percent.

DATA BASE ORGANIZATION

One of the major problems in the design of a geographical information system or an urban and regional information system is the design of the data base, especially with regard to area type data. The two main types are the polygon and the cell, both previously described. The polygon type has the advantage of storing the boundaries of an area to within system tolerance. This may be useful where boundaries are precisely known, such as political or land ownership boundaries. It may be unnecessary where the boundaries are not precisely known, such as in many bedrock formations.

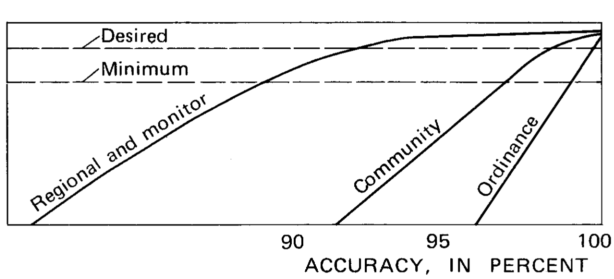
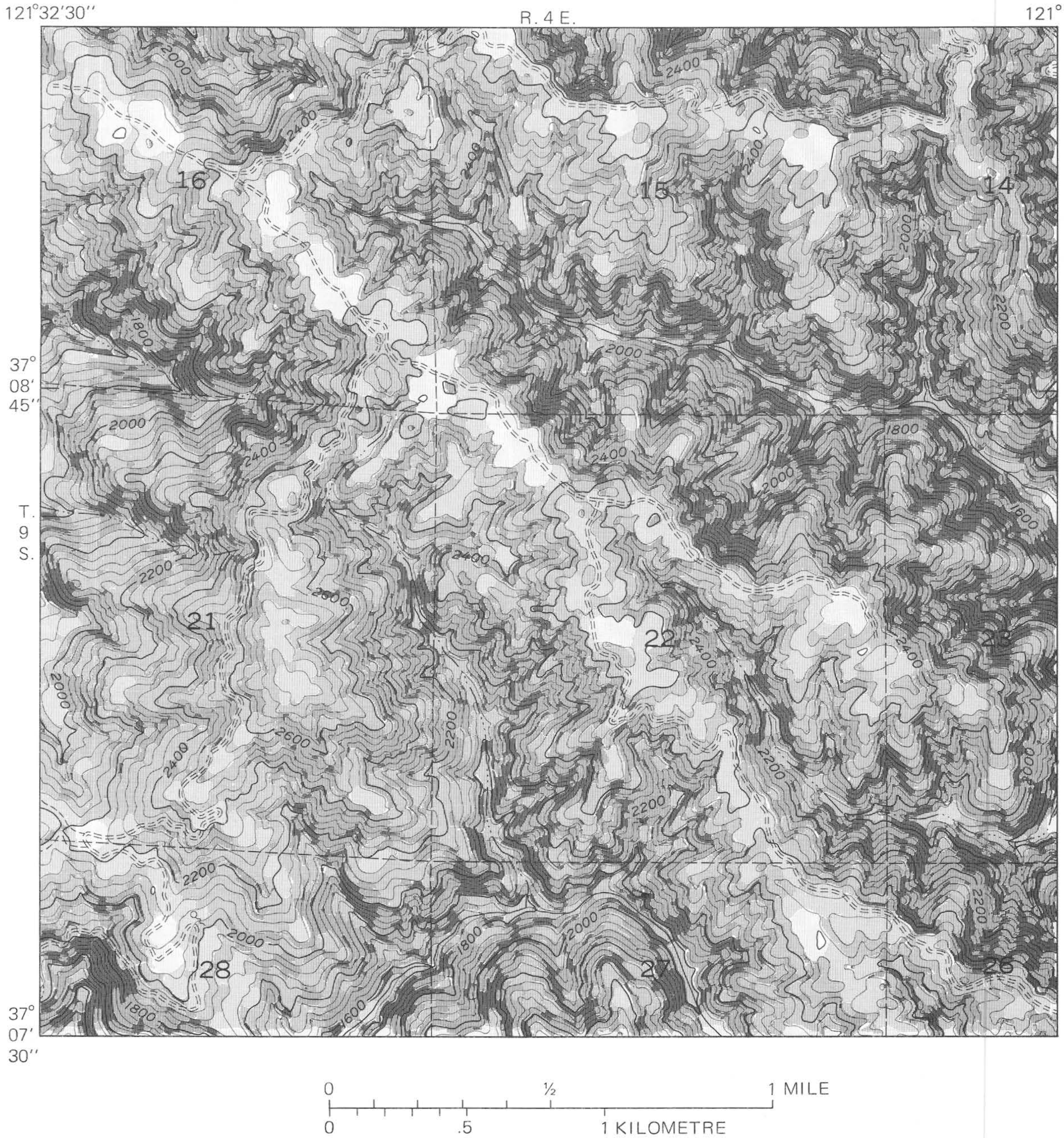


FIGURE 17.—Error tolerance by type of planning (from Taylor, 1971).

If overlaying polygons must be done in the computer, it can be very expensive. The polygon type, therefore, is not conducive to mathematical modeling where a large amount of overlaying is required. The properties of the cell structure are just the reverse. The overlaying has already been accomplished in establishing the data base. The main drawback of the cell type is its imprecise representation of areal boundaries.

At present, a cell structure is recommended for earth science areal type data for a region such as the San Francisco Bay region, especially if a large amount of simulation and mathematical modeling is to be done. Imprecise representation of the boundary is not a serious problem for these purposes. A cell size of about 150 × 150 m (or its equivalent in geodetic cell size) is required for some geologic areal data; for the remainder, a 300 × 300 m cell is satisfactory. Although it may be advisable to include line and point information with the cell record, it probably will be necessary to have separate network files of some line information for modeling purposes.

The handling of slope information depends on its use and on the terrain of the region. In relatively flat country, a slope map of about five zones may be amenable to digitization, since it is no more complex than many land-use maps. However, a similar slope map for mountainous terrain can be so complex as to be virtually impossible to digitize with an image digitizer or a line follower. For example, the slope map of the Mount Sizer area, near San Jose in the San Francisco Bay region, has six zones: 0–5, 5–15, 15–30, 30–50, 50–70, and over 70 percent (fig. 18). If such a slope map must be digitized, it might best be done using a scanner for the intermediate slope map products, the so-called "choke negatives" used in making the map. The first choke negative has only zones with slope over 70 percent and would be digitized giving the sixth zone. The second choke negative has zones with slope over 50 percent and would be digitized. Suitable software would allow the results of the first digitization to be subtracted from the results of the second (in the computer) yielding zones with slope between 50 and 70 percent. The process could be continued until the six slope zones were digitized. If slopes are not to be used in modeling, this process seems an unwarranted expense. If slopes are to be used for modeling, reliance on a slope map restricts the user to the slope categories of the map. For



PERCENT OF SLOPE

Water	0-5	5-15	15-30	30-50	50-70	70 Plus

This slope information was developed at 1:24,000 scale using the contour color-separation plate for the 7.5-minute quadrangle map covering the area. Each slope category is the result of photomechanical manipulation of the contour lines. The slope zones have been edited to remove the effects of contour values, elevation figures, and nonfactual images produced by the process.

FIGURE 18.—Part of slope map of Mount Sizer quadrangle (U.S. Geological Survey, 1970).

these reasons, it is recommended that, in mountainous areas such as in the San Francisco Bay region, elevation on a small grid rather than a slope map be digitized.

The compatibility of earth science data with other geographically distributed data depends on use of the data. An example is census data, generally organized on the ACG–DIME system or assigned to census tracts. Some systems include or provide for a limited amount of land-use information and transportation networks. The census tracts are small in urban areas and large in rural areas. However, much of the planning—hence much of the modeling—will be concerned with areas not presently urbanized. Census information should be assigned to cells in rural areas, cells much smaller than the census tracts and whose boundaries coincide with boundaries of square groups of earth science data cells as in the ORNL organization.

If there is much demand for the use of an interactive system, for example, using a storage display tube and not requiring much computation in the computer, it may be cheaper to maintain a separate set of polygon files, in addition to the cell file, for just such applications. Depending on the method of digitization, these files may be archival files derived from digitization of maps and used in the generation of the cell file.

Finally, the data should be organized on the basis of geodetic coordinates, and the metric system should be used for linear measure, since national adoption of the metric system is near. The difficulty of mathematical modeling across zones of a projective system, such as the Universal Transverse Mercator (UTM) (fig. 19), is largely removed by the use of geodetic coordinates. Although cell size becomes smaller toward the north, the variation is small from one row to the next and is easily handled in a mathematical model. Finally, maps can be plotted in any of the commonly used projections with existing software routines.

MAP DIGITIZATION

The cost of digitizing the earth science data generated by the San Francisco Bay Region Program has been estimated. As a preliminary step, the amount of data to be digitized was estimated. From

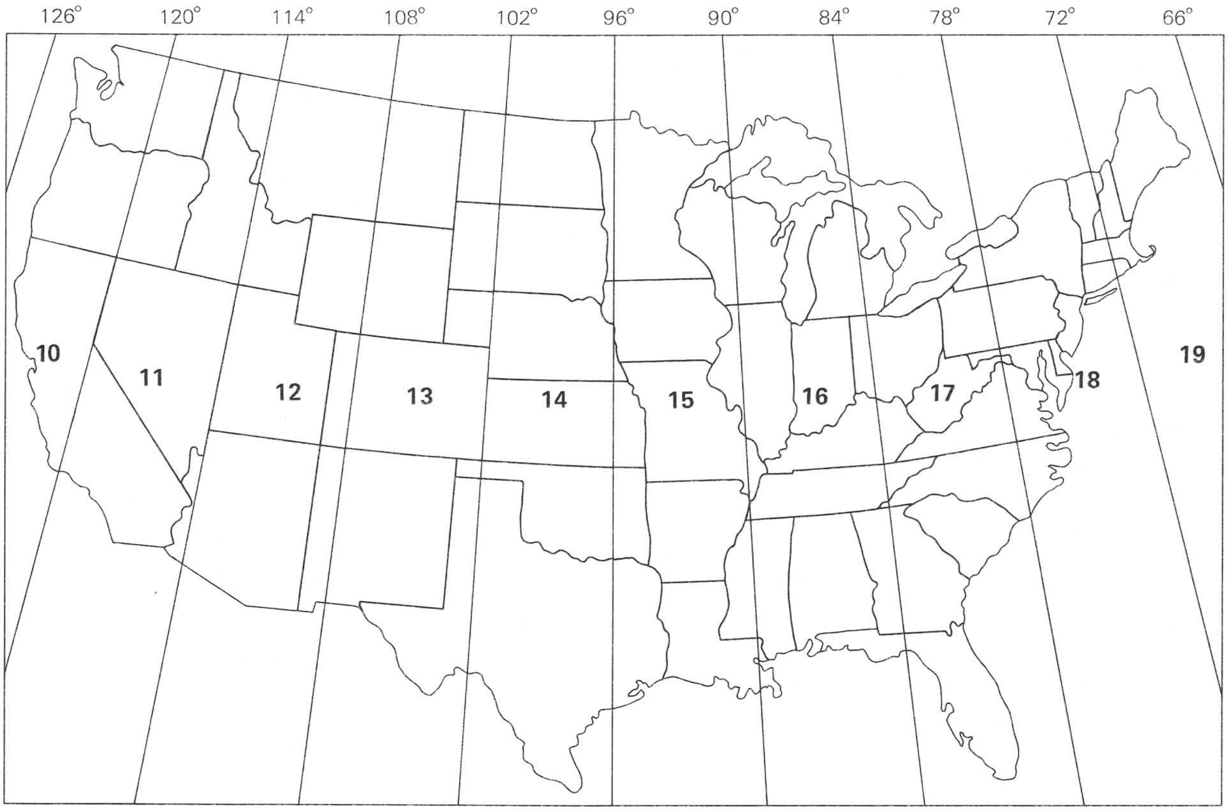


FIGURE 19.—Map zones for the Universal Transverse Mercator projection in the United States.

this, an estimate of the time of digitization was made, and the results then converted to cost. Such estimates are difficult to make and not very reliable except for very simple maps. Many companies in the map digitizing business will only bid on a "cost per point" basis. This particular estimate was made more difficult for two additional reasons. The region covered is very large so that for each data element, such as bedrock geology, the estimate for the region is based on an extrapolation of measurements made for a small part of the region taken as representative or average. This may lead to errors if the part selected is not truly average. Second, an average rate of digitization was used. Actually, the rate will vary considerably from map to map.

The estimate provides for digitization of elevation contours and not slope maps. It is assumed that the elevations will be gridded subsequently and stored in the data bank, with slopes being computered as needed.

The estimate is based on the use of manual image digitizers. Most of the maps are too complex to consider use of automatic line followers. No estimate of the cost of digitization by the use of scanners is included because (1) scanners used for map digitization are still in the research and development stage, (2) software development is incomplete, and (3) the amount and kind of preediting or preparation of special manuscripts for scanning depend on the scanner and the data being scanned and are, at present, difficult to estimate.

The following table is based on extrapolation of measurements made on small parts of the bay region:

Data type	*Data element*	*Map scale*	*Total line length (inches)*
Basic data.	Bedrock	1:62,500	39,500 (plus 33,500 symbols)
	Surficial deposits	1:62,500	16,000
	Landslides	1:62,500	8,500
	do	1:24,000	145,000
	Active faults	1:250,000	200
	Annual rainfall	1:500,000	1,000
	Soil	1:62,500	38,000
	Vegetation	1:62,500	82,000
	Land use	1:62,500	97,000
Interpretive.	Landslide susceptibility	1:62,500	155,000
	Bay mud	1:125,000	500
	Land-use categories	1:62,500	500
	Flood inundation maps	1:24,000	13,700
	Aquifer yields	1:125,000	2,500
	Hillside materials and inferred engineering behavior.	1:62,500	39,500 (identical to bedrock)

Data are not yet available for other program elements that may be digitized.

Total line length is now summed to give two grand totals, one for a data base of basic data and one for a data base for interpretive data. The total for interpretive data also includes active faults and land use since these data elements will probably be required in both sets.

	Basic data	*Interpretive data*
Total line length (inches)	427,000	309,000
Time to digitize (hours)	4,270	3,090

The time is based on a digitization rate of 100 inches per hour, a conservative estimate based on the experience of the U.S. Geological Survey in digitizing elevation contours from topographic maps. Their experience also calls for allowing about 40 percent of total time for editing, which means that the time shown above is 60 percent of total time. Therefore, total time including allowance for editing is:

	Basic data	*Interpretive data*
Total time (hours)	7,150	5,150

To these totals should be added the time to digitize elevation contours, 25,000 hours. This estimate (including allowance for editing) is based on estimates furnished by the U.S. Geological Survey. With elevation added, the totals are:

	Basic data	*Interpretive data*	
		With elevations	*Without elevations*
Total hours	32,150	30,150	5,150

Converting this to cost depends on local conditions. Taking a cost factor that should include operator salary, cost of supervision, and amortization of equipment, the following costs are obtained for a range of cost factors:

Cost factor	*Basic data*	*Interpretive data*	
		With elevations	*Without elevations*
$10/hr	$321,500	$301,500	$ 51,500
15/hr	482,250	482,250	77,250
20/hr	643,000	603,000	$103,000

SUMMARY

The focus of this study is urban and regional information systems that must handle relatively large (multicounty) areas including urbanized parts that are expanding. The main concern has been with the incorporation of earth science data, whose main attribute, from the point of view of a computer system designer, is its spatial quality. This attribute must be preserved in the computer system.

The design of such an information system should recognize the following factors:

1. System specifications should be developed considering the requirements of the user community. The users will need the aid of geologists, hydrologists, and engineers regarding the use of earth science data for their particular region.
2. A system can have a wide range of capabilities. Some of the more important ones are:
   a. Interactive and batch mode capability.
   b. Large data manipulation and computational capability.
   c. Multiple type input and output devices.
   d. Multiterminal support.
   e. Support of multiple data bases.

The extent to which a particular system incorporates these will depend on regional needs.

3. The initial cost of the data base will be high. The greater the accuracy, the higher the cost, in a rapidly increasing, nonlinear fashion.

When considering specifically an information system for the San Francisco Bay region, it may not be necessary to build from scratch but only to modify the system of the MTC to make it truly regional. It must be decided whether to incorporate earth science data as interpretive data or as basic data plus the necessary software to develop interpretations as required.

Decisions must be made on two additional items. The data base may be organized on a cell basis or as an area boundary file. At present a cell-organized data base seems best if there will be much modeling and simulation. The area boundary file is best for a simple storage and retrieval system. If overlaying of files is required, it may prove very costly with an area boundary file, unless the Raytheon arc approach is a significant improvement.

Second, the operation of map digitization should consider the hardware used, map standards, and editing procedures together as a whole. Use of an image digitizer is probably most economical at present. Scanning devices may be better for very simple maps. However, for scanning, the production of overlays from complex maps may not be much different than digitizing with an image digitizer. For the San Francisco Bay region in particular, (1) a cell-type data base is recommended for the earth science data, (2) the slope maps for the area should not be digitized, and (3) the cost to digitize the earth science data can be only roughly estimated.

REFERENCES CITED

Alexander-Marrock, P. D., Friend, P. F., and Yeats, A. K., 1970, Mark sensing for recording and analysis of sedimentological data: Systematics Assoc. Spec. Vol. 3, 16 p.

Amidon, E. L., and Akin, E. S., 1971, Algorithm selection of the best method for compressing map data strings: Assoc. Computing Machinery Commun., v. 14, p. 769–774.

Askevold, G. L., and Vallee, J. F., 1973, Information organization for interactive use—design implications in data-base systems: Am. Soc. Inf. Sci. Jour., v. 24, no. 4, p. 287–299.

Auerbach Publications, 1972, Auerbach on digital plotters and image digitizers: Princeton, N. J., Auerbach, 211 p.

Bay Area Transportation Study Commission, 1969, Bay area transportation report, 85 p.

Boyle, A. R., 1972, Geographic information systems: Geographical Data Handling, Internat. Geog. Union, v. 2, p. 1151–1200.

Boyle, A. R., and Sharp, J., 1972, Basis of assessment of equipment selection: Geographical Data Handling, Internat. Geog. Union, v. 2, p. 668–709.

Brabb, E. E., compiler, 1970, Preliminary geologic map of the central Santa Cruz Mountains, California: U.S. Geol. Survey open-file map, scale 1:62,500.

Brabb, E. E. Pampeyan, E. H., and Bonilla, M. G., 1972, Landslide susceptibility in San Mateo County, California: U.S. Geol. Survey Misc. Field Studies Map MF-360, scale 1:62,500.

Center for Advanced Computation, 1972a, IRIS—Illinois resource information system feasibility study, final report: Urbana, Univ. Illinois, 204 p. 1972b, NARIS—A natural resource information system: Urbana, Univ. Illinois, Doc. No. 35, 15 p.

Cristiani, E. J., Evey, R. J., Goldman, R. E., Mantey, P. E., 1973, An interactive system for aiding evaluation of local government policies: Electrical and Electronics Eng. Trans. on Systems, Man and Cybernetics, v. 2, p. 141–146.

Dodd, G. G., 1969, Elements of data management systems: Assoc. Computing Machinery Surveys, v. 1, no. 2, p. 117—133.

Kevany, M. J., 1968, An information system for urban transportation planning: Santa Monica, Calif., System Development Corp., 56 p.

McHarg, I. L., 1969, Design with nature: Garden City, N.Y., Natural History Press, 197 p.

Meyers, C. R., Jr., 1971, New tools for regional planning: Am. Inst. Architects Jour., v. 56, no. 4, p. 37–40.

Oak Ridge National Laboratory, 1971, The environment and technology assessment progress report, June to December 1970: U.S. Atomic Energy Comm., p. 178–201.

Raytheon Company, 1973, Program summary report: The development of a natural resource information system, v. 1, 25 p.

Smith, J. G., and Berg, H. C., 1973, Use of machine-processable field notes in a wilderness mapping project (Granite Fiords area), southeastern Alaska: U.S. Geol. Survey Jour. Research, v. 1, no. 5, p. 579–585.

Taylor, L. R., 1971, Proposed information for planning: San Diego County Comprehensive Planning Organization, v. 1, 85 p.

Tomlinson, R. F., 1968, A geographic information system for regional planning: CSLRO Symposium, Canberra, Australia, p. 200–210.

Totschek, R., Almendinger, V., Kevany, M., and Needham, K., 1970, Work program for a bay region information support center: System Development Corp., 117 p.
U.S. Bureau of the Census, 1970, The DIME geocoding system: Census Use Study Rept. no. 4, 46 p.
U.S. Department of Housing and Urban Development, 1968, Urban and regional information systems: Washington, D. C., 333 p.
U.S. Geological Survey and Department of Housing and Urban Development, 1971, Program design, 1971: San Francisco Bay Region Environment and Resources Planning Study, 123 p.

☆U.S. GOVERNMENT PRINTING OFFICE: 1975-0-690-036/31