GEOGRAPHIC INFORMATION SYSTEM DEVELOPMENT IN THE CARETS PROJECT

By William B. Mitchell, Robin G. Fegeas, Katherine A. Fitzpatrick,
and Cheryl A. Hallam

U.S. Geological Survey

(E78-10128) GEOGRAPHIC INFORMATION SYSTEM DEVELOPMENT IN THE CARETS PROJECT (CENTRAL ATLANTIC REGIONAL ECOLOGICAL TEST SITE), VOLUME 4 Final Report (Geological Survey, Reston, Va.) 78 p HC A05/MF A01 CSCL 08F G3/43 N78-24592 Unclas 00128

Contents

	Page
Preface	v
Abstract	1
Background	2
The management subsystem of a GIS	3
Survey and evaluation of existing GIS approaches in the United States	9
Advisory group of the IGU commission on geographical data sensing and processing	12
Conclusions and recommendations resulting from the IGU case studies	17
Canada Geographic Information System	25
Data input process	25
Data reduction	30
Retrieval subsystem	31
Comparison of the accuracy of area calculations by manual methods and the CGIS	34
Geography Program geographic information system	45
Digitization of polygon map data	49
Editing process	55
Interactive digitizing and cartographic editing system	61
Conclusions	66
References	71

Illustrations

		Page
Figure 1.	Geographic information system development	4
2.	Manual procedures of data input process	26
3.	Map image data as processed by the drum scanner	27
4.	Computer operations of the data input process	29
5.	Example of land use plot and tabular summaries	33
6.	Geographic information system operations flow diagram	46
7.	Graphic input procedure flow diagram	57

Tables

Page

Table 1. Selected geographic information systems and related encoding methods ---------- 18

2. Area determinations for CGIS map sheet 11L03 ---------- 35

3. Area determinations for CGIS map sheet 82J ---------- 37

4. Area validation for CARETS map sheet JRG3W ---------- 39

5. Area determination for a single polygon (No. 209) in acres -- 40

6. Area determination for a single polygon (No. 319) in acres -- 41

7. Area validation of randomly selected polygons values in acres ---------- 42

8. Area validation of randomly selected polygons values in acres ---------- 43

9. Area validation of randomly selected polygons values in acres ---------- 44

PREFACE

This report chronicles the problems, decision points, activities, and research involved in designing and implementing a geographic information system for the Central Atlantic Regional Ecological Test Site (CARETS) Project and later for the U.S. Geological Survey (USGS). It is not possible to identify a clear point of demarcation between activities and efforts devoted solely to the CARETS Project and those expended for the USGS. This report covers activities from the initiation of the CARETS Project in July 1972 to the cutoff date of this report, June 1976. The development thrusts described in this report had their genesis in the CARETS Project and their expansion and extension into the Geographic Information Retrieval and Analysis System under the USGS are testimony to the critical importance that has been assigned to the development of an operational system for handling geographic information.

Any geographic information system must include at least six subsystems in order to facilitate the encoding, storage, manipulation, retrieval, and use of spatial data. These six subsystems are: (1) the management subsystem; (2) the data acquisition subsystem; (3) the data input subsystem; (4) the data retrieval and analysis subsystem: (5) the information output subsystem; and (6) the information use subsystem. The data acquisition subsystem is considered in another of the CARETS final reports. The information use subsystem of the geographic information system is insufficiently developed at this time and also will not be discussed. This report describes actions and research undertaken to develop the management, data input, data retrieval and analysis, and information output subsystems.

William B. Mitchell

GEOGRAPHIC INFORMATION SYSTEM DEVELOPMENT IN THE CARETS PROJECT

By William B. Mitchell, Robin G. Fegeas, Katherine A. Fitzpatrick, and Cheryl A. Hallam

ABSTRACT

Experience in the development of a geographic information system to support the CARETS Project has confirmed the considerable advantages that may accrue by paralleling the system development with a rational and balanced system production effort which permits the integration of the education and training of users with interim deliverable products to them. Those advantages include support for a long-term staff plan that recognizes substantial staff changes through system development and implementation, a fiscal plan that provides continuity in resources necessary for total system development, and a feedback system which allows the user to communicate his experiences in using the system. Thus far balance between system development and system production has not been achieved because of continuing large-scale spatial data processing requirements coupled with strong and insistent demands from users for immediately deliverable products from the system. That imbalance has refocused staffing and fiscal plans from long-term system development to short- and near-term production requirements, continuously extends total system development time, and increases the possibility that later system development may reduce the usefulness of current interim products.

BACKGROUND

The Central Atlantic Regional Ecological Test Site (CARETS) is one of the sites designated in 1970 by the National Aeronautics and Space Administration (NASA) where detailed environmental evaluations of the Earth Resources Technology Satellite (ERTS, later changed to Landsat) and correlative aircraft and ground data were to be conducted by multidisciplinary teams. Sponsored jointly by NASA and the U.S. Geological Survey (USGS), the CARETS project was formulated during 1970 and 1971 in the USGS. The project formally came into being as a NASA-sponsored ERTS experiment on July 1, 1972.

The primary objective of the CARETS demonstration was to test the extent to which ERTS and related remotely sensed data could be used as input to a regional land resources information system. The rationale and structure of the CARETS experimental information system were chiefly concerned with ways of gathering, processing, packaging, and calibrating information and making it available to users and decisionmakers. Development of a geographic information system was one of the four major aspects of the CARETS project, the others being compilation of land use maps from the remotely sensed data, assessment of environmental impact, and evaluation of resulting data products by users. The goals for the geographic information system included the automatic measurement and summation of data sets, in addition to the retrieval of updated and composited data sets.

Following the start of the CARETS project, investigators made a number of key data-acquisition decisions that influenced the development of the geographic information system. These decisions were to: (1) compile Level II land use information based on 1970 data obtained from high-altitude aircraft,

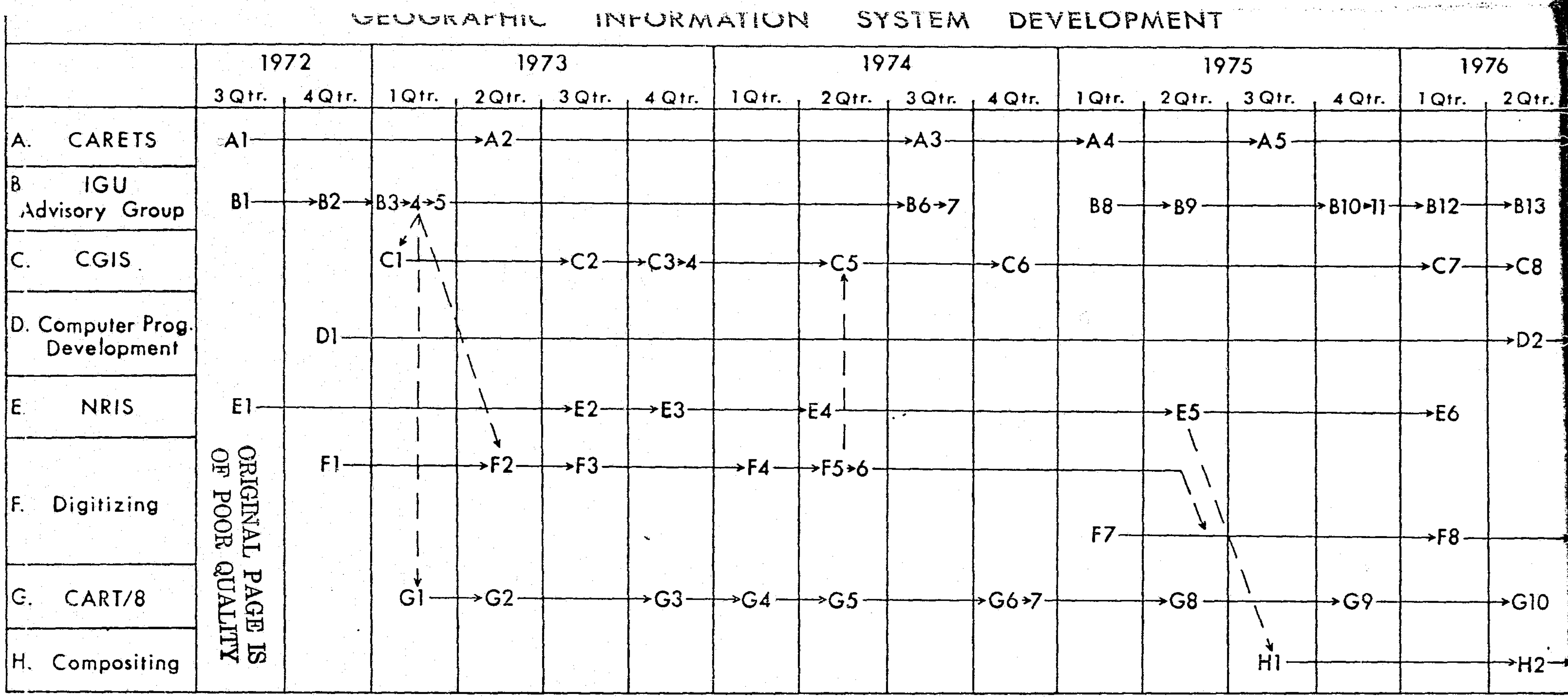
(2) update Level II land use data to 1972, (3) provide census county subdivisions, drainage basin, cultural feature, and geology overlays and, (4) compile Level I land use and land cover data from ERTS images. These data sets were then to be digitized as the basic data base for the CARETS information system. To retrieve, manipulate, analyze, display, and summarize that data base, investigators decided to adapt and incorporate procedures, programs, and capabilities of geographic information systems in operation or under development at the time the CARETS project was initiated.

THE MANAGEMENT SUBSYSTEM OF A GIS

Tomlinson, Calkins, and Marble (1976) have noted that any geographic information system must comprise at least six major subsystems in order to facilitate the encoding, storage, manipulation, retrieval, and use of spatial data. These six subsystems are: (1) the management subsystem, (2) the data acquisition subsystem, (3) the data input subsystem, (4) the data retrieval and analysis subsystem, (5) the information output subsystem, and (6) the information use subsystem. They point out that "there are just as many problems, possibly more, on the management side of an information system as there are on the technical side." One of the major management and administrative problems in implementing the system is to design an information system that minimizes the adverse effects produced by time delays in obtaining project approval, acquiring the necessary equipment, and hiring and training the staff.

This report describes the problems, activities, and research involved in developing a geographic information system (GIS) for the CARETS project. The report identifies critical decision points for the development of a GIS capability. Some of those decision points have been diagrammed in figure 1, which shows the development of the GIS. A brief description of the GIS development follows, including a discussion of the major directions of development. Finally, a brief description of the basic characteristics of the geographic information system being used for land use and land cover mapping and data compilation in the USGS in 1977 is presented.

The sequence of decision points and subsequent events and activities in the development of the CARETS Geographic Information System that are diagrammed in figure 1 emphasize the critical role that timing, lead time, and interrelationships between the main components of a GIS play. The first five events (events A1 through A5), shown along the top line, indicate the controlling time sequence in CARETS. The development of a GIS for CARETS, however, could not keep pace with this time line. The documentation of this inability and the lessons to be learned in planning the development of a GIS are the major contributions of this report.



CARETS

A1 Award of CARETS/ERTS contract.

A2 CARETS maps completed.

A3 CARETS Land Use and Census Tract maps open filed.

A4 Delivery of data summaries derived from ERTS: land use by county at 1:250,000.

A5 Delivery of tape data on land use by county and drainage basin at 1:250,000.

IGU Advisory Group

B1 Decision to ask IGU to advise CARETS and GAP.

B2 Request for proposal to IGU.

B3 IGU grant award.

B4 IGU system recommendations on GIS development.

B5 IGU case studies and data encoding experiments.

B6 Preliminary system reports and encoding experiment results.

B7 IGU research report draft submitted to USGS.

B8 Geography Program initiated procurement for IGU grant for analysis of spatial data handling in USGS.

B9 IGU grant award: analysis of spatial data handling in USGS.

B10 IGU first interim report on status of USGS spatial data systems.

B11 IGU software inventory (version 4).

B12 USGS/IGU working seminar series begun.

B13 IGU seminars concluded.

CGIS

C1 IGU advisory group recommends CGIS test by CARETS.

C2 CGIS preliminary test on Norfolk sheet.

C3 CGIS preliminary test submitted to CARETS.

C4 Evaluation and offer of CGIS to do volume test.

C5 USGS/CARETS agreement with CGIS.

C6 CGIS digitizing of CARETS maps (volume test) initiated.

C7 Completion of CGIS volume test on CARETS maps and delivery of data.

C8 Subcontract for graphic display of CGIS digitized data.

Computer Program Development

D1 In-house programming of graphic input procedure and manipulative capabilities.

D2 GIRAS PHASE I: development and current status of Geographic Information Retrieval and Analysis System.

NRIS

E1 Raytheon NRIS contract awarded.

E2 Demonstration of NRIS.

E3 NRIS software available and testing of program begun.

E4 Twenty-seven of thirty-five NRIS programs tested.

E5 Contract to share costs with Bureau of Indian Affairs for NRIS compositing program optimization by Washington State University.

E6 Optimization attempt unsuccessful.

Digitizing

F1 In-house digitizing for CARETS.

F2 RFP for digitizing begun.

F3 Final draft of digitizing RFP.

F4 Digitizing RFP responses received.

F5 Evaluation of responses shows none qualified.

F6 Digitizing of Geography Program land use and other data sets by applied Physics Lab of Johns Hopkins University started.

F7 Start of contract with I/O Metrics Corporation to digitize land use and other data sets for Geography Program.

F8 I/O Metrics contract expired March 1 and extended to October 1, 1976.

CART/8

G1 IGU hardware specifications.

G2 Preliminary RFP for interactive cartographic system.

G3 Purchase orders issued for equipment.

G4 Digital Equipment Corporation minicomputer received.

G5 Tektronix graphic display equipment received.

G6 Instronics digitizer received.

G7 Dynamap interface and software received. Interactive cartographic system installed and tested. Operational use of system begun.

G8 DYNAMAP CART/8 Version 1 software installed. Operational use of system continued.

G9 Improved display capability for interactive cartographic system and CART/8 Version 2 software requisitioned.

G10 CART/8 display hardware and software purchase order issued.

Compositing

H1 CALSPAN overlay tape compositing contract begun.

H2 CALSPAN contract extended to December, 1976.

Figure 1--Geographic information system development

By the second quarter of 1973, (about 1 year after the initiation of the CARETS ERTS project), land use and land cover maps were ready to be digitized. Shortly after the start of the project, investigators decided to negotiate with the Commission on Data Sensing and Data Processing of the International Geographical Union (IGU) for a grant and to request consultation concerning how best to attack the problems of developing a geographic information system (figure 1, event B1). In particular, it was felt that the IGU Advisory Group could expedite the developmental process by recommending existing GIS systems or components of systems that could be immediately utilized. Although the IGU group ultimately contributed valuable insights and advice to the CARETS project, the lead time that elapsed between the decision to ask the IGU to form an advisory group and the actual grant award (figure 1, event B3) was nearly half a year. The major recommendations of the IGU group for GIS development in CARETS are indicated on the diagram by dashed lines (figure 1, events C1, F2, and G1).

After consulation between the CARETS project staff, USGS staff, and the IGU Advisory Group, steps were taken to obtain a test of the Canada Geographic Information System (CGIS) using CARETS data from the Norfolk map sheet (figure 1, events C1 and C2). A competitive request for proposal (RFP) was also used for digitizing services from commercial firms in the United States (figure 1, event F2). The necessary paperwork was initiated to procure computer hardware for an interactive cartographic system (figure 1, event G2). In addition to these measures, the GIS staff in CARETS continued digitizing and programming experiments with USGS digitizing equipment (figure 1, event D1 through D2 and F1 through F6) and continued to monitor the Raytheon Corporation contract with the Bureau of Land Management and the Bureau of Indian Affairs, which promised to lead to the development of the Natural Resources Information System (NRIS) (figure 1, event E1).

Had the CARETS project not encountered unforeseen lengthy delays, any of these activities could have met the interpretation and compilation production schedule of CARETS and would have permitted the timely digitizing and computer processing of spatial data necessary for a geographic information system. None of the activities met the schedule, however, and some of the approaches have not even yet proved operationally viable. None of the six proposals received from the 60 vendors who requested the RFP for digitizing provided acceptable error-free digitizing services at the time of the evaluation (figure 1, event F5).

The Canada Geographic Information System effort (figure 1, events C1 through C8) provided the digitizing, manipulative, and analytic operations required on actual CARETS spatial data. From the standpoint of the acquisition of an in-house geographic information system capability, however, the CGIS was not considered because of the infeasibility of importing a geographic information system from Canada on a sole-source basis, and other factors. A general description of the digitizing process of the CGIS is given later in this report. The CGIS has conducted an accuracy evaluation of its system, which is described in the same section.

Although the adaptation of existing GIS capabilities in the United States did not prove to be a timely or operationally viable solution to the problem of processing, manipulating, or retrieving CARETS land use and land cover data within the time frame originally set for the project, the CARETS objective of participating in the development of an operational geographic information system has been achieved. All of the elements of the geographic information system that are now major operational components of the USGS GIS were the direct result of in-house or contractual activities initially undertaken or funded by the CARETS Project. A system description of the GIS is included in this report, but special note must also be made of the graphic input procedures developed for post-digitizing processing by the USGS, and a utilization of the interactive cartographic system (INTERMAP) as an integral part of the GIS, which is believed to be unique in the United States. Finally, the studies and evaluation provided by the IGU Advisory Group (figure 1, events B1 through B7) were initiated during the CARETS project and constitute probably the only existing body of consolidated information on the development of geographic information systems. Followup studies by the IGU Commission will expand this critical and valuable source of information on operational geographic information systems (Tomlinson, Calkins, and Marble, 1976).

SURVEY AND EVALUATION OF EXISTING GIS APPROACHES IN THE UNITED STATES

In the development of a geographic information system for the CARETS project, considerable effort was made to locate, acquire, and evaluate computer programs already developed in support of other operational geographic information systems in the United States. The most useful systems appeared to be the Polygon Information Overlay System (PIOS) developed for the County of San Diego (Dangermond, 1971); and the MAP/MODEL System developed for the Columbia Region Association of Governments, Portland, Oregon (Arms, 1968). Descriptions of those systems were analyzed and evaluated for their applicability to the CARETS GIS problem. Neither system, however, provided an operable solution to the critical problem of editing and correcting the digital record resulting from the digitizing process. Without such a procedure, for the graphic input, all errors resulting from the interpretation process, compilation of other map sources, and digitizing process would be retained and compounded in the subsequent processing, analysis, and manipulation of the digital record. In addition, both systems digitized each polygon separately with the result that all lines on the map were digitized twice. Since it is impossible in manual digitizing for an operator to redigitize a given line with exactly the same precision, the many slivers of areas and false polygons created by this method introduced additional errors into the digital record. Therefore, the two systems were not considered further.

During this initial period of evaluation, another system which was found to show promise was the Natural Resource Information System[1/] (NRIS), then under development (Raytheon Corpoation, 1973) (figure 1, event E1). The NRIS

1/ Any use of trade names and trademarks in this publication is for descriptive purposes only and does not constitute endorsement by the U.S. Geological Survey.

was to be developed on contract for the Bureau of Indian Affairs (BIA) and the Bureau of Land Management (BLM) of the Department of the Interior as a natural resource management information system, but with a component which would provide for the input, manipulation, and retrieval of natural resource maps of various types. Although the Boeing Company (1972 a,b) of Seattle, Washington, had first completed a feasibility study for BIA/BLM, proposing a system essentially similar to PIOS or MAP/MODEL, Raytheon Corporation was awarded the developmental contract. Raytheon proposed a system which emphasized the geographic information system aspect and further proposed an approach which was to be based upon line segment digitizing. This approach was not then used in any other GIS input procedure but promised to eliminate many of the errors produced by other systems. This approach was adopted by the GIS group for use in the CARETS project and later by the USGS.

Although NRIS appeared to promise solutions for the CARETS project in the areas of data storage and retrieval, manipulation, and graphic production, close liaison with the Raytheon Corporation soon revealed that the NRIS contract was written to exclude any transfer of digitizer editing and correction programs to the public domain. If those programs were to remain proprietary to Raytheon, any digitizing and subsequent editing and correction of CARETS land use and other land cover maps would necessarily have to be performed solely by the Raytheon Corporation. Since a central objective of the CARETS project was to provide programs and subroutines for all aspects of a GIS in the public domain, it became evident that such programs would necessarily have to be developed in-house or obtained elsewhere.

The system proposed by Raytheon Corporation, however, promised to provide CARETS and GAP with most of the necessary geographic information systems capability. CARETS provided Raytheon Corporation with test data for digitizing processing, manipulation, and information retrieval tests with NRIS in October 1972. A demonstration of the system, scheduled for July 1973, was finally held in September 1973. The NRIS software was delivered to the Department of the Interior in November 1973, and the system was finally implemented on the USGS IBM-360 computer in December 1973 (figure 1, event E3). The USGS began testing the NRIS programs, and although testing continued into September 1974 (figure 1, event E4), by May it was apparent to USGS programmers that the NRIS compositing programs (the most critical of all manipulative programs in a geographic information system) were prohibitively expensive to run on USGS computers. In May 1975, the USGS agreed to share contract costs with the Bureau of Indian Affairs for an effort by Washington State University to revise the NRIS compositing programs to reduce their excessively high computer-run costs (figure 1, event E5). The attempt was unsuccessful, however, and the effort was abandoned in January 1976.

ADVISORY GROUP OF THE IGU COMMISSION ON GEOGRAPHICAL DATA SENSING AND PROCESSING

The problems in obtaining usable computer programs for CARETS led to the decision to contract for the services of an advisory group from the Commission on Geographical Data Sensing and Processing of the International Geographical Union. The IGU Commission had completed, in June 1972, a report on geographic information systems that resulted from a symposium held by the Commission in Ottawa, Canada, (Tomlinson, 1972). After some discussions, the Commission began to work cooperatively with the CARETS project and other projects in the USGS under USGS Grant No. 14008-001-G-67 and subsequent amendments (figure 1, events B1 through B7).

As a result of negotiations with the Commission, three main activities were planned:

1) The initial development of a geographical data-handling capability to meet then current program responsibilities of GAP;

2) The identification of the geographical data-handling requirements of the proposed USDI/USGS land use program and, to the extent possible, make the GAP system compatible with such requirements and use it as an experimental test bed to produce preliminary evaluation of the cost effectiveness of the techiques which it incorporated;

3) The determination of the cost effectiveness of the GAP system in comparison with alternative existing or experimental geographic information systems which deal with the same or similar data sets and query types.

Realizing that these three activities were necessarily interdependent and would have to be, to a large extent, carried out concurrently, the Commission identified four tasks:

1) Geographic information system specification;
2) Software/hardware recommendations;
3) Product review; and
4) System critique.

The responsibility of the Commission was to thoroughly examine the needs and requirements of the USGS geographic information system and, on the basis of its technical expertise, to provide sound recommendation and advice to the USGS on the development of the desired system. The Commission was also to critically examine the experimental system in such a way that the USGS could make viable decisions regarding future development of their geographic data handling capability. To accomplish this, the Commission established an Advisory Group (chaired by Dr. Duane F. Marble and composed of Drs. Waldo Tobler, Ray Boyle, and Dieter Steiner) to handle Tasks 1 and 2, and a Critique Group (headed by Dr. Hugh Calkins) to handle Task 4; and Task 3 was to be carried out by the two groups in cooperation (Tomlinson, Calkins, and Marble, 1976).

The responsibilities of the IGU Critique Group were defined in the work statement between the IGU and the USGS as follows:

1) To participate in the preparation of the preliminary system specifications so that the recommended system will meet the geographical data handling requirements of the USGS and be compatible with a proposed USDI/USGS Land Use Program;
2) To conduct a review of existing systems which have data handling characteristics similar to the needs of the USGS;

3) To compare and criticise the final recommendations to the USGS for a system that will meet its geographic data handling needs in light of the experience and results from the review of the other similar systems; and

4) To prepare general GIS recommendations that would address the major problems anticipated in the development of a proposed USDI/USGS Land Use Program.

Activities 1, 3, and 4 in the above list were to require substantial interaction between the Advisory Group and the Critique Group so that the final report of the Critique Group would incorporate recommendations of the entire IGU team. Activity 2 was essentially an independent task of the Critique Group and did not rely directly on extensive interaction with or feedback from the Advisory Group.

Between February and May 1973, the Critique Group participated in discussions with the USGS staff which resulted in the preparation of general system specifications [by D. F. Marble] (figure 1, event B-4). Further staff and IGU discussions resulted in a detailed work plan for the Critique Group (figure 1, event B5). This plan consisted of two major elements:

1) Five case studies of existing or experimental geographical data-handling systems to determine the various characteristics of the systems, including the major methods of geographical data manipulation. The significant factors that would be used in evaluating the system recommended for GAP would be extracted from these case studies.

2) Experiments in the area of data encoding (the conversion of graphical images to digital data records) in order to assess more accurately the resource requirements of this process, including manpower, hardware, and money.

Case studies were to be conducted on the Canada Information System, the Oak Ridge Regional Modeling Information System, the State of New York Land Use and Natural Resource Information System (LUNR), the Minnesota Land Management Information System (MLMIS), and the San Diego County Polygon Overlay Information System (PIOS). In addition, the data encoding (digitizing) tests were to be defined and applied to each system in the case study in an attempt to develop comparative statistics and other measures of the cost effectiveness of the various operations of each system (figure 1, event B5).

During the period from June 1973 to March 1974, the majority of the case-study and data-encoding experiment information was assembled. Documentation of each case study and the data encoding experiment was prepared. The documents describing each system and data encoding experiment were returned to the individual or agency responsible for each system for review and comment. As of July 1974, all comments had been received from the participating individuals and agencies (figure 1, event B6). It was decided that the Critique Group's report should summarize the major lessons learned from the case study materials and the data encoding experiment as a set of general recommendations which would be applicable to either the specific geographic data handling problems of the USGS or to a broader, more extensive land use program within the Department of the Interior.

The final report of the Critique Group was organized as follows:

1) Evaluation methodology;

2) Overview of case studies and data encoding experiment;

3) General recommendations resulting from the case studies and data encoding experiment;

4) Recommendations relevant to perceived USGS objectives and the general land use mapping program;

5) Conclusions emphasizing areas of uncertainty and requirements for continuing investigations;

6) Appendices containing detailed descriptions of each case study system and the data encoding experiment.

The draft report (figure 1, event B7), "Computer handling of geographic data," including the five case studies and the data encoding experiments description, was published by the Unesco Press in 1976 (Tomlinson, Calkins, and Marble, 1976).

CONCLUSIONS AND RECOMMENDATIONS RESULTING FROM THE IGU CASE STUDIES

The conclusions and recommendations of the Critique Group based on the data encoding experiment and comparative case studies are summarized in this section. The purpose of the summary is to highlight the recommendations which the IGU Commission believed were relevant to development of a geographic information system for the CARETS project and in the USGS and to identify areas of uncertainty in the developmental program as well as the requirements for continuing investigation. The conclusions and recommendations are placed in the context of the actual experience in developing a geographic information system in the USGS.

The five systems selected provided a range of encoding methods as well as a variety of encoding procedures (table 1). After the initiation of the study, the sponsor of LUNR system chose to include the system in the case studies but not in the data encoding experiment. The DIGIMAP system currently being developed by the CALSPAN Corporation was subsequently added to the data encoding experiment. The DIGIMAP system uses a fine-polygon and automatic-line-following scanning technique. DIGIMAP, however, is not represented in the comparative case studies.

Table 1.--Selected geographic information systems and related encoding methods

Geographic information system	Encoding method
The Canada Geographic Information System (CGIS)	Fine polygon,[1/] drum scanner
The Oak Ridge Regional Modelling Information System (ORMIS)	Small grid, flying-spot scanner
The Minnesota Land Management Information System (MLMIS)	Medium grid, manual grid overlay
The New York Land Use and Natural Resource Information System (LUNR)	Large grid, manual coding
The San Diego Comprehensive Planning Organization's Polygon Information Overlay System (PIOS)	Coarse polygon,[1/] manual digitizer

1/ Fine and coarse polygons are differentiated by the number of points used to define a polygon.

The Critique Group of the IGU Commission concluded that, for the five systems studied, there are probably more problems on the management side of the design and development of an information system than on the technical side. The Group found that the most critical problems in the management of a developing system were obtaining timely approval of the system, obtaining staffing slots and hiring personnel for them, acquiring the necessary equipment, obtaining fiscal support, and finally, implementing the system in the face of significant delays in accomplishing those tasks. The Group emphasized the need to estimate the probable delays in system design and implementation plans. The system design should be structured to minimize the adverse effects of delays. The Group pointed out that opponents of a system development project will capitalize on delays and that underestimating the time necessary in system development will necessarily adversely affect the implementation of its various parts.

In the five systems studied, there was a strong need to "sell" the system as far up the line of decisionmaking as possible. Major problems confronting this effort are the negative attitude that exists toward computer utilization and the misconceptions of people at the various levels through which approval must be obtained concerning the use of the system. Thus, in "selling" the system, it is necessary to involve personnel other than the higher level decisionmakers, that is, the data gatherers, the system designers, and the anticipated user groups.

In comparing the development of the five systems, the Critique Group noted several possible models of fiscal continuity for the development of a geographic information system, assuming that such development would take 5 to 10 years. One model orients the project so that deliverable products are scheduled at the end of each year or each funding period. The purpose of the products is to justify the continuation of the developmental program. The difficulty with this approach is that considerable emphasis and effort may be placed on the production of deliverable products with the consequent diminution of the effort on overall system development. Furthermore, it may later be found that the interim products may not be suitable for the final system. In general, the approach tends to extend the time required for system development.

An alternative approach to fiscal continuity is to plan system development over the shortest possible time without providing for interim deliverable products. System experience has indicated that the delivery of the first products might be possible in 5 years.

The Critique Group suggests that a variation of the last approach would be the inclusion of an extensive user education program within the project with a limited set of deliverable products in order to support the continuing development of the system. This approach anticipates a training program that keeps the user appraised of the capabilities and utilization of the system as it is developed. The training program, it is noted, should include the formulation of the user's problems in quantitative terms as well as instruction in utilizing the system. Those instructions would include operating characteristics, assumptions concerning the data and manipulative functions, reasonable expectations on system performance, and limitation of the system.

All of this assumes that the development program can be arranged so that less complex system tasks can be accomplished at an early date in the development, especially with users who lack extensive resources in manpower and money. Because this approach attempts to integrate user education and interim deliverable products without seriously delaying system development, the Critique Group concludes that this approach will require greater attention to the management of the system than has been evident in the systems studied.

Conclusions of the Critique Group relating to the data acquisition, input, storage, retrieval, and analysis subsystems of a geographic information system concerned the problems of data editing, the use of interactive system capabilities, file structure, and manipulative capabilities. They point out that a major technical constraint in the development of systems is digitization of data and especially the processing or converting an error-prone manuscript map into an acceptably error-free nongraphic file. Because in the present state of computer development, the computer cannot intelligently ignore nonlogical errors, the significance of an error-free digital file is particularly important. The process of creating such a file includes 1) pre-editing of graphic data, 2) digitizing, 3) detection and correction of errors, and 4) file structuring.

If a system is intended to handle multiformat (point, line, and area) data, the Group concludes, the file structure must be designed with this in mind at the inception of system development. The subsequent addition of one or more of these types to an existing file structure can be fairly expensive and can reduce efficiency considerably.

The use of interactive cartographic data manipulation capabilities can be of considerable benefit in the error correction process. The Critique Group concludes that manipulative functions remain a problem in geographic information systems because most users do not understand the assumptions or statistical bases upon which manipulations are made.

Solutions to the problems with the information output and use subsystems of the geographic information systems are included in the last set of recommendations and conclusions resulting from the comparative analysis of the five information systems by the Critique Group. Most of those problems relate to the utilization of manipulative capabilities or statistical analyses packages referred to above. Users generally believe that automated systems perform tasks much faster than manual systems and consequently judge a system's performance against that criterion. Thus, a user expects higher performance from the automated system even though he lacks an understanding of the manipulative or statistical techniques available. The user's lack of understanding of manipulative techniques, moreover, often leads to a failure to properly interpret the output of the system and further may lead to the acceptance of results without an appreciation of the underlying assumptions of the manipulation and the qualifications that should apply to data interpretation based upon those assumptions. This lack of understanding has also led to the underutilization of a system's manipulative capabilities.

A similar situation exists with regard to the use of the statistical packages currently available for computer-based analysis. The standard statistical packages available do not adequately meet the needs for spatial data analysis; on the other hand those statistical methods capable of handling spatial data are not generally used in the systems studied.

Education of users in the employment of spatial data manipulative techniques and statistical methods can be best approached, the Group believes, by a "go-between" process which essentially places persons knowledgable in spatial manipulative and statistical methods between the actual user and the system. Those "go-betweens" would have the task of interpreting or translating the user's problem into quantitative terms to which the system could respond. Higher level policy decisions, they note, are primarily based upon a mental model that is built upon information seen or heard by the decisionmaker. The degree to which the decisionmaker's mental model or "image" of the problem is modified by information from a geographic information system may critically depend upon how successfully the "go-between" or middle level technical personnel can transform the information (through manipulative or statistical techniques) into a form compatible with the policy maker's mental model or problem image. Unfortunately, the Group reports, "there are <u>no</u> mechanisms by which to determine the degree to which this image is subject to modification by the type of information which could come from an information system. None of the systems studied have a mechanism for either determining the effects of information or negotiating the use of information and consequently they are suffering as a result of lack of these mechanisms. Further, it must be recognized that even if these functions are supplied, there may be no perceivable change in decisionmaking, and that to base the continuity and survival of the information system on such notions could easily be fatal." Upon this rather pessimistic note, the Critique Group concludes that it is incumbent upon the system designer to include personnel capable of understanding the user's problem and to be able to translate that problem into terms suitable to the system's capabilities for spatial data manipulation and statistical analysis. "The designer must take the system to the user and on the basis that the user can understand it in terms of his problem." (Tomlinson, Calkins, and Marble, 1976).

CANADA GEOGRAPHIC INFORMATION SYSTEM

The Canada Geographic Information System (CGIS) had been used to provide digitizing services for the Government of Canada on an experimental basis since 1965, and was believed capable of handling the large number of CARETS map sheets. GCIS had an operational capability to produce composite maps from several overlays and had been used to accurately digitige high-density polygon maps. Because the CGIS was the single known system capable of providing accurate and economical digitization, the Geography Program awarded a $113,000 contract for use of the CGIS in digitizing the priority sheets of the CARETS project (figure 1, event C6).

Data Input Process

To be input to the CGIS, the CARETS maps first had to be converted from the Universal Transverse Mercator (UTM) coordinate system to the geographic coordinate system and were laid out in 1- by 2-degree map sheets corresponding to the standard USGS topographic map format. The maps were then reproduced on scribecoat and scribed. This scribed copy was later scanned by the drum scanner to enter the polygon boundaries into the system. Figure 2 is a schematic diagram of the input procedure.

The scanner consists of a black drum on which a map is mounted and a moveable carriage which moves the scanning head across the face of the map. The scanning head passes over the map and senses the amount of light reflected from the maps every 1/250 of an inch and records this information as a series of bits on magnetic tape. Where a line is present the drum scanner records a "1" on the scan tape. A normal full size map sheet (about 34 by 24 inches) takes approximately 10 to 15 minutes to scan. Figure 3 is an

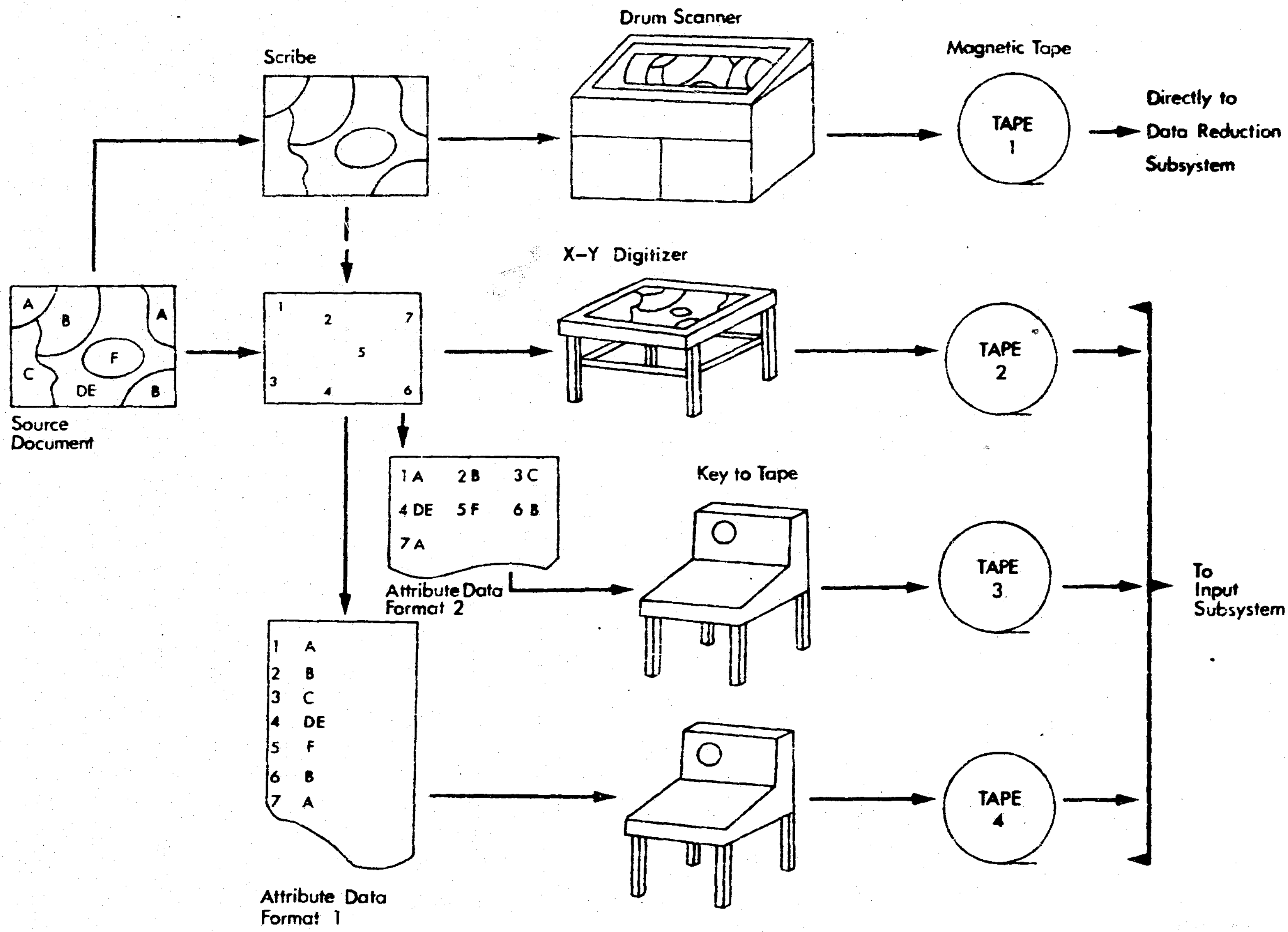


Figure 2--Manual procedures of data input process

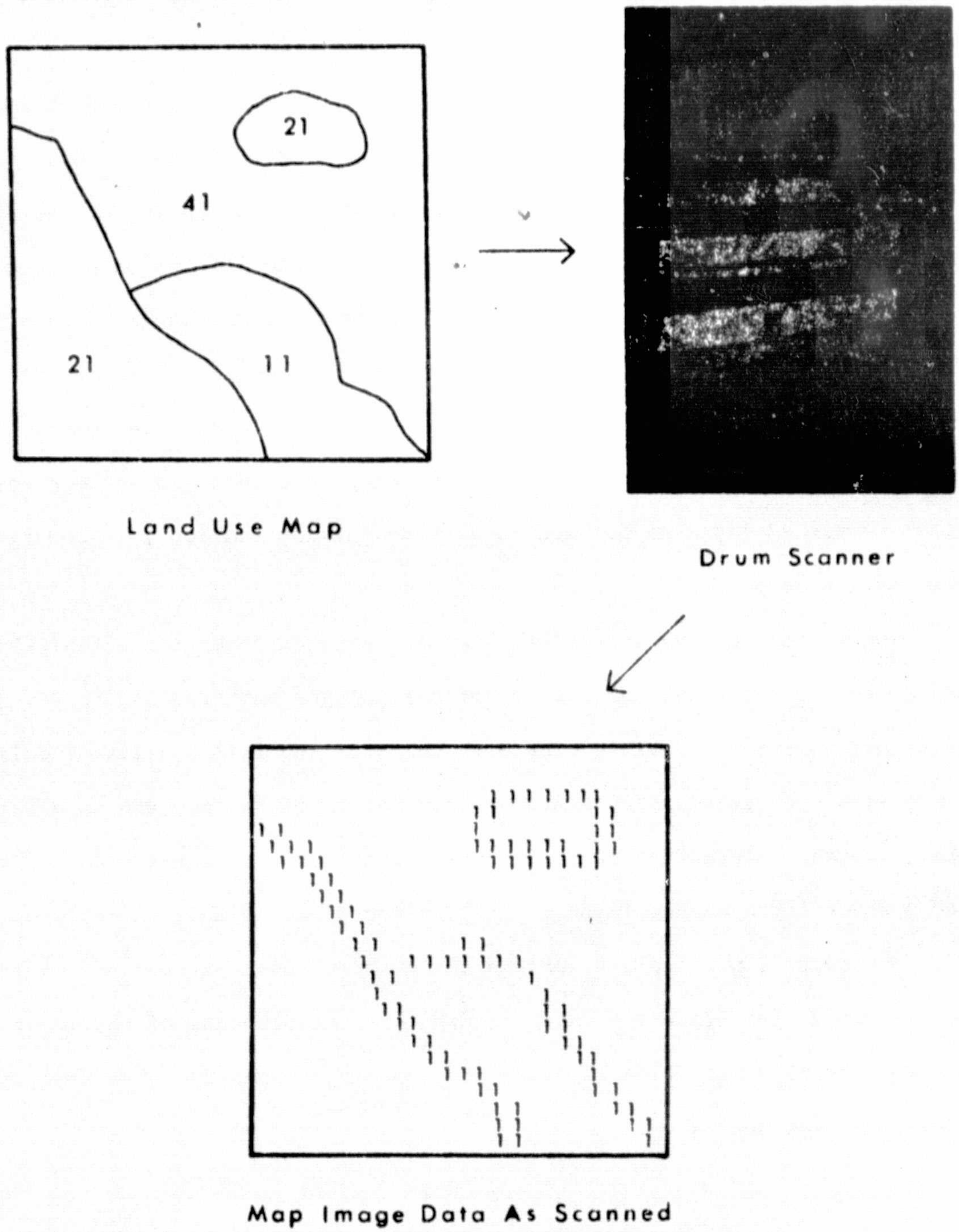


Figure 3--Map image data as processed by the drum scanner

example of the map image data as processed by the drum scanner; the zeros have been omitted for clarity. These map image tapes produced by the scanner go directly to the data reduction system where they become the image data set.

Once the map has been scanned, a numbered overlay to the map is prepared. On a drafting film overlay to the map sheet, a draftsman assigns a unique but sequential number to each polygon (from 1 through n, beginning at the upper left hand corner). Two different people then transcribe the data, each using a different format to list the polygon numbers and corresponding classification data. These two lists are keypunched directly to a magnetic tape, edited, and verified to eliminate transcription errors and other errors introduced in processing. In this manner, classification errors are detected and corrected. These completed lists of polygon numbers and classification data comprise the description data set.

The x and y coordinates of the map and polygon faces are established using the scribed copy, the overlay of the polygon numbers, and an x-y digitizer. After overlaying the scribed copy with the drafting film overlay of polygon numbers, an operator digitizes the corner points of the map and a point within each polygon, referencing it to its polygon number. This produces a digitizer tape, which links each polygon to an x, y coordinate within the polygon.

The data from the digitizer and the encoder tapes are merged and edited to produce a classification tape (figure 4). The products of the input subsystem--the classification tape and the map-image tape--are then sent to the data-reduction subsystem.

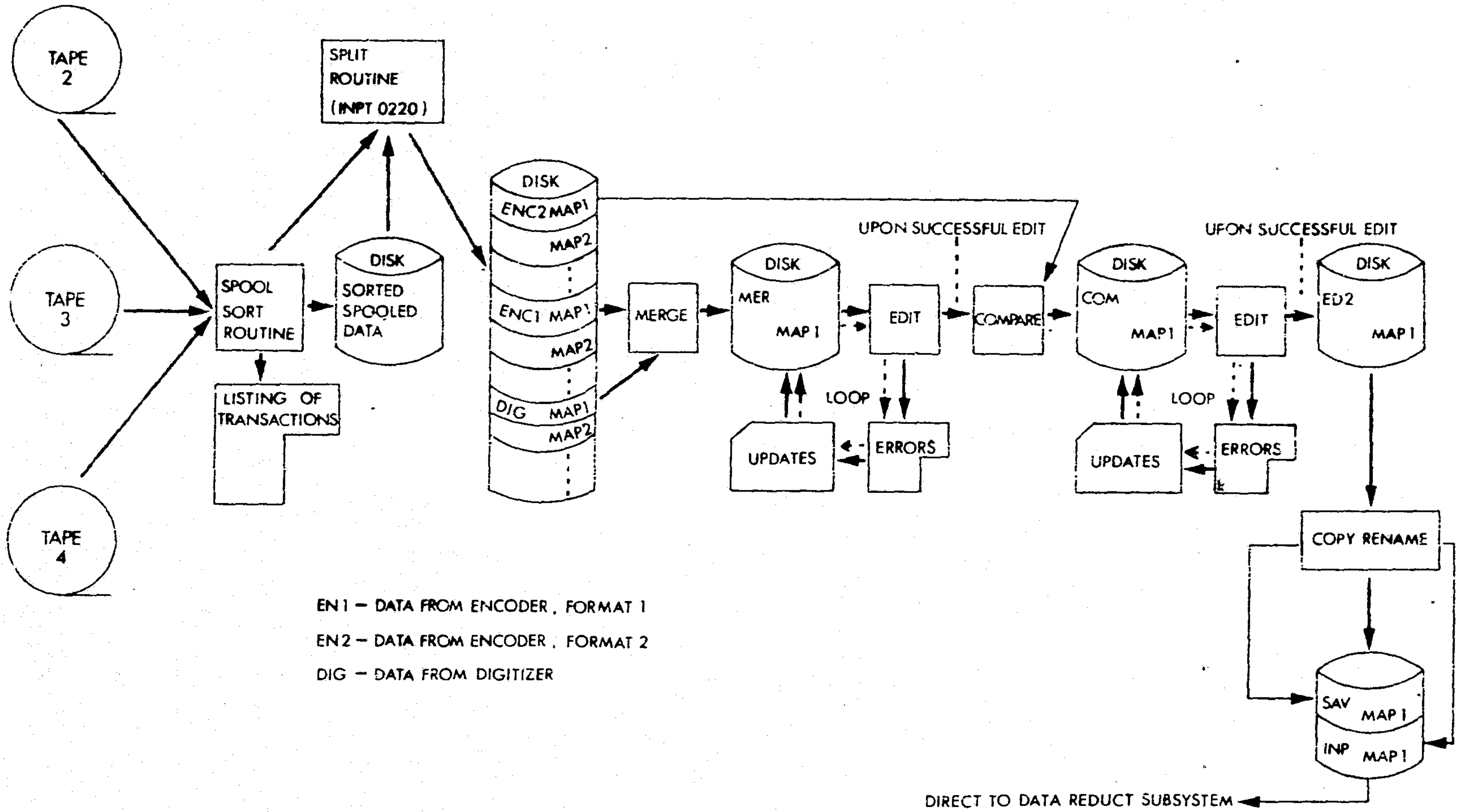


Figure 4--Computer operations of the data input process

Data Reduction

The data reduction subsystem is comprised of nine major groups of programs (phases 0-8). These programs transform the map-image tape and the classification tape into the image data set and the description data sets. The image data set is the coded image of the polygon boundaries. The descriptor data set is the classification data describing each of the polygons.

The data reduction subsystem modifies the map-image tape. In the data reduction subsystem, lines are closed and the several points identifying the width of a line are reduced to a single point so that each line is only one point wide. The polygons are checked for closure and then corrected. The arbitrary digitized coordinates of the map corner points and the points within each polygon are transformed to latitude and longitude coordinates. The map-image data are also converted from the arbitrary x, y coordinates of the scanner to latitude and longitude coordinates.

In the data-reduction subsystem, the image and classification data are edited for missing lines, gaps in lines, extraneous lines, and erroneous classification data. The list of errors produced is then corrected in the manual error correction subsystem (Tomlinson, Calkins, and Marble, 1976). The map edges are identified, and during the late phases of data reduction, polygons that may cross map borders are matched and assigned a new polygon number in both the image data set and the description data set. Areas for these polygons are recalculated, thereby eliminating the lines of the map edge.

The image data set and the description data set, stored on either magnetic tape or disk, comprise the data bank. The image data set consists of all point information necessary to reproduce a graphic display of the map.

Each polygon in the image data set is identified by a numerical descriptor. The image data set is stored in a series of frames. A frame is the basic unit of storage used by CGIS. The CARETS data were stored in frames 3.75 minutes on a side which resulted in 128 frames per map sheet at a scale of 1:100,000 and 512 frames per map sheet at a scale of 1:250,000. According to Environment Canada, Lands Directorate "the system is capable of maintaining the accuracy of this image to four one thousands of an inch (0.004 inches), regardless of scale." (Environment Canada, 1973).

The description data set is a list of the descriptive data for each polygon listed by the same number identifying it in the image data set. The descriptive data for each polygon would be the classification data, the area of the polygon, the latitude and longitude of the centroid, and a list of each frame in which the polygon appears.

Retrieval Subsystem

CGIS can retrieve the overlay map data for different types of coverages[1/] and display the data in either graphic or tabular form, or both. Once the map data are in the data bank the possibilities for retrieval format are almost infinite. When data for several types of map coverage are overlaid, the graphic display becomes difficult to visualize. CGIS can display in tabular form the descriptive data for each type of map coverage or for a combination of coverages. In the CARETS project, for example, Landsat-derived land use data were tabulated by both county and State.

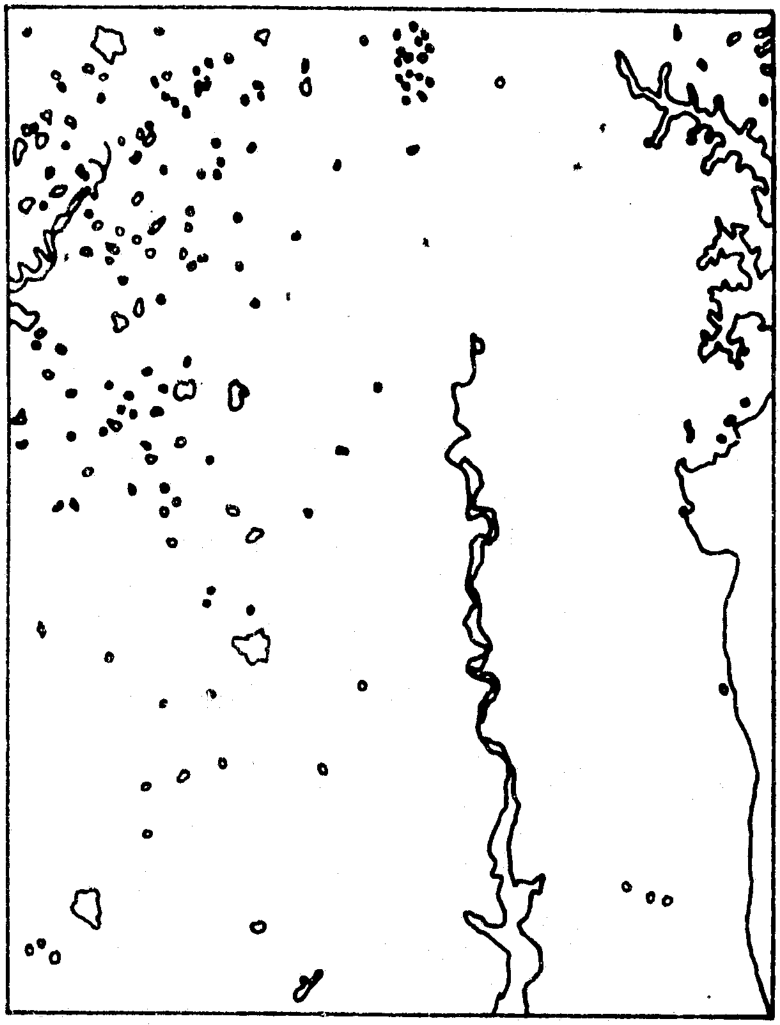
1/ Coverage is used to describe a map overlay of a single set of data, such as land use or political boundaries.

Maps at scales ranging from 1:370 to 1:1,000,000 may be put into the system. A separate program permits the retrieval of any data at any scale within that range. Plots or tabular data compositing two maps at different scales can then be obtained.

The CARETS project first received Landsat Level I land use data compiled by census tract and drainage basin from the CGIS. Following this, selected Level II land use maps, land use change overlays, census tract and county boundary overlays, geologic maps, and drainage basin maps at a scale of 1:100,000 were processed by the system; these maps cover the multicounty priority area of the Washington and the Norfolk-Portsmouth Standard Metropolitan Statistical Areas.

Once the maps are stored in the data bank, the information may be input to a graphics display system. Using a remote terminal such as a Tektronix 4012 or 4014 storage display system, the user may interrogate the system and manipulate the data. The Tektronix system can present a visual display on a viewing screen of either graphic or tabular data for a given type of map coverage or set of coverages and can produce the display as a hard copy. For example, a single category of land use for a given county on a map sheet or set of map sheets may be displayed as a plot, and the area of the category, in acres, may also be put out in tabular form. Figure 5 shows such a plot and listing for land use category 16 in the Patuxent River area between Chesapeake Bay and Washington, D.C.

The use of this system requires only that the information in the CGIS data base be converted to the graphics subsystem and that the user have access to a remote terminal. The system is somewhat limited because it lacks the storage capacity to handle detailed graphic data for a large area. By reducing the polygon size, the total area that can be viewed increases.



SUMMARY OF AREAS SELECTED

SELECTION 7 08/05/75

PLUS	- MAJOR		
	USE	AREA	PER CENT
	1	7,298.0	100.0
PLUS	-SUBGROUPING		
	USE	AREA	PER CENT
	6	7,298.0	100.0
TOTAL AREA SELECTED IS		7,298 ACRES	

Figure 5--Example of land use plot and tabular summaries

Both area measurements in acres and hectares can be obtained. On the other hand, tabular data can be displayed for even complex combinations of coverages. For example, a listing of the land use types and areas of the land use types for a given county could be tabulated.

The CARETS researchers will be able to assess the land use and other coverage data from a terminal in the USGS National Center. If the CGIS data base were converted to a format useable in-house, the USGS would then be able to access and manipulate the data for any user request.

Deliveries of data summaries derived from Landsat (land use by county at 1:250,000) and tape data on land use by county and drainage basin at 1:250,000 were made by CGIS in the first half of 1976 (figure 1, event C7).

Comparison of Area Values Determined by the CGIS and by Manual Methods

The CGIS personnel performed three types of area measurements using digitized maps in order to compare the accuracy of the areas calculated by their system to areas calculated by manual methods. In April 1974, CGIS personnel compared their calculated area for Prince Edward Island, to the value given by Statistics Canada. The two values differed by 11,522 acres or 0.82 percent of the total area. CGIS investigators then compared the CGIS area calculations with areas calculated by four manual methods for the total areas of two Canadian map sheets. The manual methods used were dot planimeter count, polar planimeter, weight measurement, and quadrilateral tables.

Researchers compared the results of each of the above methods to the areas determined by the CGIS. Table 2 shows the deviations for the first map sheet (11L03). Of all the manual methods, only the polar planimeter determinations varied more than 1 percent from the CGIS determined area for this map.

Table 2.-- Area determinations for CGIS map sheet 11L03

METHOD	AREA	DEVIATION FROM AVERAGE OF ALL METHODS EXCEPT CGIS		DEVIATION FROM CGIS	
	(ACRES)	(ACRES)	%	(ACRES)	%
CGIS	265,363.6	+ 3419.6	+ 1.31	-	-
DOT COUNT	266,532.6	+ 4588.6	+ 1.75	+ 1169.0	+ 0.44
WEIGHT CALCULATION	263,127.9	+ 1133.9	+ 0.45	- 2235.7	- 0.84
POLAR PLANIMETER	254,556.3	- 7387.7	- 2.82	-10,807.3	- 4.07
QUADRILATERAL TABLES	263,559.2	+ 1615.2	+ 0.62	- 1804.4	- 0.68
AVERAGE OF ALL METHODS EXCEPT CGIS	261,944	-	-	- 3219.6	- 1.29

The CGIS determined value varied only 1.31 percent from the average of the manual methods, a deviation less than either dot counted or polar planimeter calculated deviations from the average area.

Table 3 shows the same type statistics for the second Canadian map sheet (82J). It also shows each of the three dot counted values used to derive the average dot counted areas and each of three polar planimeter calculated values used to find the average polar planimeter area. These figures show the variation possible when using either the dot planimeter or polar planimeter for area determination. For the second sheet, none of the manually determined areas vary by as much as 1 percent from the CGIS determined area. As a result, the CGIS value varies only 0.06 percent from the average of all the manually determined values.

The CARETS researchers consider the difference between area values determined by the CGIS method, and manual methods, as described above, insignificant.

Once computer processing of CARETS map began, the CGIS staff conducted a series of area checks for a sample from the CARETS data base. The area of a map sheet computed by CGIS was not only checked against areas from other methods, but against itself. Table 4 displays the inconsistency of the dot count method for a full CARETS sheet (0.15 percent between counts 1 and 2), the deviations of the average dot count value from the area of the map sheet (0.40 percent), and the difference between the average CGIS overlay area and the average dot counted area (0.46 percent).

The information in table 4 concerns the 12 maps listed as "Composite Maps" and "Data Reduction Process Maps." The data reduction process maps were combined to make the composite overlay maps. The largest difference in the 12 separate determinations is less than 1-1/2 acres for a map of nearly 600,000 acres. This implies that the system is consistent in its determination of large areas.

Table 3.--Area determinations for CGIS map sheet 82J

METHOD	AREA	DEVIATION FROM AVERAGE OF ALL EXCEPT CGIS AREA		DEVIATION FROM CGIS AREA	
	(ACRES)	(ACRES)	%	(ACRES)	%
CGIS	3,900,508	2,639	0.06	-	-
DOT COUNT 1	3,894,144	9,003	0.23	6,364	0.16
DOT COUNT 2	3,937,643	34,496	0.88	37,135	0.95
DOT COUNT 3	3,909,789	6,642	0.17	9,281	0.24
AVERAGE	3,913,858	10,711	0.27	13,350	0.34
WEIGHT CALCULATION	3,900,431	2,716	0.06	77	0.002
POLAR PLANIMETER 1	3,875,500	27,647	0.70	25,008	0.64
POLAR PLANIMETER 2	3,907,000	3,853	0.09	6,492	0.17
POLAR PLANIMETER 3	3,912,100	8,953	0.22	11,592	0.30
AVERAGE	3,898,200	4,947	0.12	2,308	0.06
QUADRILATERIAL TABLE	3,900,100	3,047	0.07	408	0.01
AVERAGE OF ALL METHODS EXCEPT CGIS	3,903,147	-	-	2,639	0.07

Next, researchers determined the area of a single polygon for comparison. They conducted two types of dot counts and a polar planimeter determination of area. Dot Count 1 included all dots falling on the boundary of the polygon. Dot Count 2 included only those dots within and mostly within the polygon. The researchers repeated each determination 12 times. The results, in table 5, show CGIS area calculations to be comparable to area calculations made by manual methods.

In further tests, researchers determined and compared the area of a selected polygon from a map sheet by dot count and polar planimeter. The results are recorded in table 6 with appropriate deviations calculated. The CGIS determined area for the polygon shows little variation from either the dot counted area (0.59 percent) or from the polar planimeter area (-1.57 percent).

In one final test, investigators determined areas of selected polygons from each of three CARETS map sheets by both dot and polar planimeter methods and compared them to CGIS determined areas. The results are presented in tables 7-9. Summing the data from table 7-9, the dot counted area differed from the CGIS determined area by 2.15 percent; the polar planimeter determined area differed from the CGIS determined area by -0.87 percent; and the CGIS determined area differed from the average of the two areas determined by 0.51 percent. Again the variations show CGIS determined figures to be as acceptable as manually determined figures.

Table 4.--Area validation for CARETS map sheet JRG3W

Manual	Acres	Absolute Difference	Percent Difference
Dot Count 1	593,839.50		
		940.58	0.15
Dot Count 2	592,898.92		
Average	593,369.21		
		2404.67	0.40
True Area of Map Sheet	595,773.88		
		2784.97	0.46
Composite Overlay* maps			
U124	596,154.37		
UABG	596,153.81		
		380.30	0.06
UDEH	596,154.36		
Average	596,154.18		
Data Reduction* process maps @			
US01	596,153.13		
US02	596,153.47		
US04	596,153.30		
USOA	596,153.27		
USOB	596,152.92		
USOG	596,153.32		
USOD	596,153.21		
USOE	596,153.40		
USOH	596,153.30		

* The area of the sheet was determined three times by CGIS. Map U124 is a composite overlay composed of maps US01, US02, and US04. Map UABG is a composite overlay of maps USOA, USOB, and USOG. Map UDEH is a composite overlay of maps USOD, USOE, and USOH. The area was then determined for each of the three more complex composite overlay maps (U124, UABG, UDEH).

@ The data reduction process is described in the section titled, "Canada Geographic Information Systems."

Table 5.--Area determinations for a single polygon (No. 209), in acres

Dot Count 1	Dot Count 2	Planimeter	CGIS
648.06 1/	613.53 to 625.48	606.51 to 654.39	618.19
1/ Mean			

1. Each count was done 12 times (36 counts in total).

2. Dot Count 1

   All dots on the line were counted as within the polygon.

   Dot Count 2

   All dots on the line that were mostly within the polygon were counted.

Table 6.--Area determinations for a single polygon (No. 319), in acres

			Average	CGIS	DEVIATION Absolute	DEVIATION %
Dot count	1	12,712.94				
	2	12,872.30				
	3	12,489.84				
	4	12,948.00				
	5	12,908.16	12,848.84			
	6	12,900.19				
	7	12,916.13			+ 76.08	+ 0.59
	8	12,896.21				
	9	12,995.81				
				12,924.92		
					-206.34	-1.57
Planimeter	1	13,083.45				
	2	13,147.20				
	3	13,051.58				
	4	13,147.20				
	5	13,131.26	13,131.26			
	6	13,147.20				
	7	13,210.94				

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 7 --Area Validation of Randomly Selected Polygons Values in Acres

FACE NUMBER US01	DOT COUNT					PLANIMETER					CGIS CALCULATE VALUE
	1	2	3	AVERAGE	BEST 2 OF 3	1	2	3	AVERAGE	BEST 2 OF 3	
122	382.46'	370.51	382.46	378.48	382.46	350.59	414.98	350.59	372.05	350.59	355.56
133	2772.86	3123.45	3294.76	3063.69	3209.10*	3235.00	3016.61	3171.26	3140.89	3203.13*	3201.81
189	18573.40	18589.34	18776.59	18646.44	18581.37	19011.64	18770.01	17959.87	18580.51	18890.82	18915.69
209	589.63	641.42	613.53	614.86	601.58	669.31	637.44	557.76	621.50*	653.37*	615.19
211	4455.09	4434.19	4529.80	4474.03	4446.14	4525.82	4414.27	4541.76	4487.28*	4533.79*	4435.93
212	1243.00	1398.38	1386.43	1309.27	1392.40*	1497.98	1402.36	1147.39	1349.58	1450.17*	1330.06
323	294.81	338.64	342.62	322.02	340.64	350.59	334.65	366.52	350.59*	358.55*	346.59
496	405.36	454.17	490.03	450.19	472.10	430.27	478.03	446.20	451.85	438.23	437.74
656	709.15	733.05	780.86	741.02	721.10	717.12	812.73	764.92	764.92	741.02	778.63
705	342.62	346.60	354.57	347.93	344.61	350.59	350.59	366.52	355.90	350.59	353.65
879	876.48	920.30	984.04	926.94	898.39	1083.64	940.22	1083.64	1035.83*	1053.64*	955.15
1135	466.12	521.90	529.87	505.96	525.88	462.14	494.01	621.50	525.88	478.07	532.32
1312	760.94	505.96	768.91	678.60	764.92*	764.92	717.12	844.60	775.55*	741.02	741.39
1774	1665.31	1713.12	1705.15	1694.53	1709.13*	1768.89	1721.08	2645.37	2045.11*	1744.53*	1705.53
1789	645.40	665.32	641.42	650.71*	643.41*	685.24	621.50	669.31	658.68*	677.27*	641.01
1791	525.88	517.92	525.88	523.23	525.88*	589.63	509.95	525.88	541.82*	517.91	523.44
1933	585.64	613.53	637.44	612.20	625.48*	685.24	653.37	653.37	663.99*	653.37*	638.35
				35960.10	36184.59				36722.43	36866.52	36727.77

224.59
0.62%

144.09
0.39%

5.34
0.015%

+762.33
2.12%

767.61
2.14%

* Value higher than calculated by CGIS.

Using averages, 1 dot count higher, 9 planimeter counts higher }
Using best 2 of 3, 6 dot counts higher, 9 planimeter counts higher } of 17 counts.

TABLE 8 —Area Validation of Randomly Selected Polygons Values in Acres

FACE NUMBER US02	DOT COUNT 1	DOT COUNT 2	DOT COUNT 3	DOT COUNT AVERAGE	PLANIMETER 1	PLANIMETER 2	PLANIMETER 3	PLANIMETER AVERAGE	CGIS CALCULATED VALUE
120	167.32	167.32	183.26	172.63	207.16	175.29	175.29	185.91*	172.96
129	155.37	139.44	167.32	154.04	175.29	159.36	159.36	164.67*	155.33
135	179.28	179.28	183.26	180.60	207.16	191.23	191.23	196.54*	183.14
145	31.87	23.90	23.90	26.55*	63.74	31.87	31.87	42.49*	25.59
157	227.08	215.13	211.15	217.78*	254.97	223.10	207.16	228.41*	217.48
171	239.04	231.07	247.00	239.03	223.10	254.97	270.91	249.66*	242.70
177	15.93	11.95	15.93	14.60*	15.93	15.93	15.93	15.93*	12.97
188	286.84	274.89	290.83	284.18	334.65	302.78	302.78	313.40*	289.64
203	67.72	83.66	71.71	74.36*	127.48	79.68	79.68	95.61*	70.35
214	23.90	31.07	31.87	28.94	79.68	47.80	31.87	62.76*	37.77
245	235.05	207.16	227.08	223.16	207.16	223.10	223.10	217.78	233.35
249	107.56	95.61	111.55	104.90*	175.29	95.61	95.61	122.17*	100.86
				1720.77				1895.33	1747.64

175.56

10.14%

-147.69

-7.79%

26.87

+1.56%

* Value higher than calculated by CGIS

Using averages: 5 dot counts higher
Using averages: 11 planimeter counts higher
of 12 counts total.

TABLE 9 --Area Validation of Randomly Selected Polygons Values in Acres

FACE NUMBER US04	DOT COUNT 1	DOT COUNT 2	DOT COUNT 3	DOT COUNT AVERAGE	PLANIMETER 1	PLANIMETER 2	PLANIMETER 3	PLANIMETER AVERAGE	CGIS CALCULATED VALUE
101	601.58	665.32	314.73	527.21	701.18	654.39	638.43	664.66*	655.51
111	1207.15	1258.94	1266.91	1244.33*	1306.75	1244.95	1388.59	1313.43*	1240.87
117	1988.01	1984.03	2035.82	2002.62	2042.99	2011.07	2027.03	2027.03*	2002.69
125	1262.92	1243.00	1243.00	1249.64	1340.71	1308.79	1292.83	1314.11*	1258.94
155	187.24	199.20	203.18	196.54	287.29	191.53	223.45	234.09*	199.85
165	354.57	362.54	362.54	359.88	367.10	367.10	287.29	340.49	360.67
192	294.81	294.81	314.73	301.45	287.29	335.17	271.33	297.93	314.79
204	258.96	250.99	250.99	253.64	239.41	271.33	287.29	266.01*	257.96
223	239.04	247.00	274.89	253.64*	255.37	255.37	239.41	250.05*	249.29
240	928.27	860.54	928.27	905.69	1101.30	925.73	1037.45	1021.49*	939.46
272	298.80	310.75	310.75	306.76*	351.13	319.21	335.17	335.17*	300.45
311	820.70	804.76	828.67	818.04	877.84	861.88	734.20	824.64	834.30
				8419.44				8889.10	8614.78

469.66
5.58%

-274.32
-3.08%

+195.34
+2.32%

* Value higher than CGIS

Using averages, 3 dot counts higher
Using averages, 9 planimeter counts higher } of 12 counts.

U.S. GEOLOGICAL SURVEY GEOGRAPHIC INFORMATION SYSTEM

The Geographic Information System of the USGS is designed to input, store, manipulate, and retrieve digital spatial data developed from land use and land cover maps at a compilation scale of approximately 1:125,000 plus overlays showing (a) Federal land ownership, (b) river basin and subbasin, (c) counties, and (d) county census subdivisions. In addition, the system will accommodate maps at other scales. Computer-generated products from the system include a digital data base tape, graphic and statistical data, and a specialized statistical and spatial data analysis. The basic data units of the system data base are (1) boundaries that are identified by the categories of land use and land cover, such as a river basin, that lie on either side of the boundary, (2) polygons identified by categories inside the boundaries, and (3) the boundary lengths plus the areas of the polygons. Boundaries are stored as strings of points defined by geographic (latitude and longitude) coordinates and are organized by 1:250,000-scale quadrangles and 7-1/2 minute quadrangles.

The system operations flow of the GIS is shown in figure 6. Following receipt of land use and land cover or other overlays from the Compilation and Interpretation Branch of the Geography Program, overlays are either manually digitized on a standard commercial digitizer table by a human operator or they are automatically digitized on a laser scanner by I/O Metrics Corporation (figure 1, event F7). Some digitizing is accomplished with an interactively operated manual digitizer table associated with the INTERMAP interactive cartographic system in the GIS Branch, but this system is not used in a production mode and its digitizing capabilities will not be discussed.

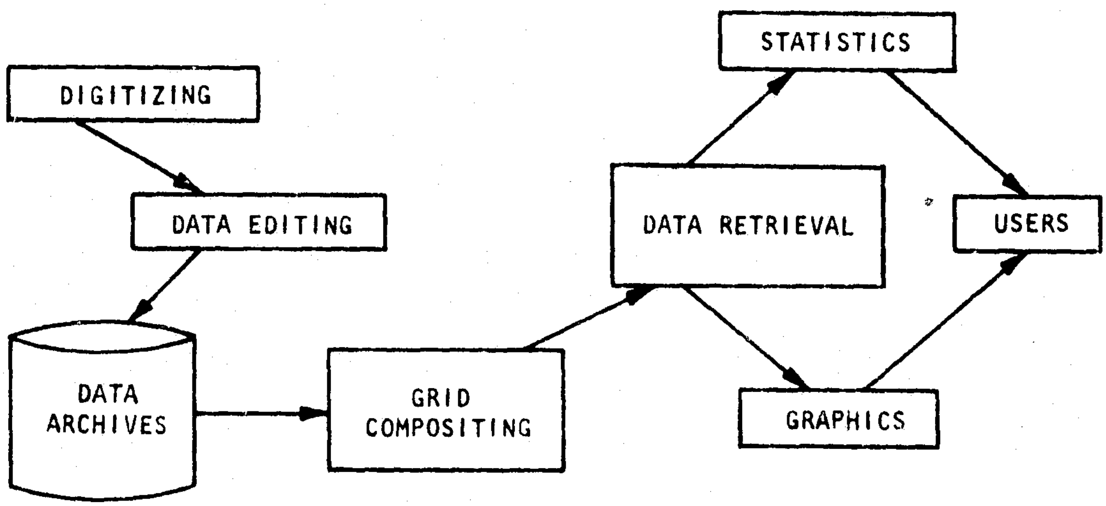


Figure 6--Geographic information system operations flow diagram

Whether overlays or maps have been digitized by manual or automatic methods, the digital record of boundaries and polygons must be edited, and errors found therein corrected by a "graphic input procedure" (GIP) developed by the USGS. Briefly, the procedure involves the use of computer plots and programs to detect and flag arc (boundary) errors committed in the digitizing process or those that remain from the compilation process before the overlay was digitized. Because these errors will compound the problems in processing or manipulating spatial data in later phases of the system, it is imperative that they be detected and corrected at this stage in processing.

Lists of corrections detected are dealt with in several ways. The method chosen depends upon the type of error and correction needed. Compilation errors often require an inquiry to the compilation staff; some corrections are performed by punch card processing through the IBM 370/155 system in the USGS and others are accomplished with the INTERMAP interactive cartographic system. Following corrections, the arcs or boundaries are chained into polygons and combined with polygon identification numbers to form arc/polygon files. Any additional errors are detected and corrected.

At this point spatial data files in the system are comprised of arc/polygon files of digitized single-subject overlays for a single 1:250,000-scale quadrangle map. The next step in the system is the computer-manipulation of those spatial data files. "Compositing" is one form of data manipulation that has been singled out in figure 6 for special attention because of its critical importance and because of the special difficulties it presents in a geographic information system. Compositing denotes performing in a computer the same process traditionally used by an analyst when he stacks (or "composites") several layers of transparent overlays of different phenomena to facilitate the study of the interrelationships of these phenomena. Computer compositing is presently accomplished through software which transform polygon maps into variable-sized grids which can be used to composite two or more single-subject arc/polygon file tapes. Development of polygon compositing software for the GIS has not been completed.

The boxes following the "grid compositing" operation in figure 6 refer to additional types of data manipulation, such as retrieval, analysis, and display that are presently performed or are under development in the GIS. Types of data manipulation that are possible are scale and/or projections change, contiguity or proximity analysis, generalization, extraction of data from composited files, spatial searches, and point radius and corridor searches.

Present capabilities of the GIS include the following:

1) Coordinates may be converted to UTM, State plane coordinates, or any other rectangular or spherical frame of reference.
2) Polygons may be converted to grid cells of any size and in any frame of reference.
3) Data may be plotted back on any projection, at any scale, and with a wide range of selection criteria.
4) Data may be retrieved on the basis of individual polygons or boundaries, type of boundaries or polygons (for example, land use category or county) or combinations, such as land use by census tract or by drainage, through the grid compositing operation.

5) Data may be retrieved by geographic location (that is, by 7-1/2 minute or 1:250,000-scale quadrangle, by any arbitrary area, by proximity to any point, or by proximity to any route).

6) Data may be retrieved by any combination of the above.

Graphic outputs of the system include editing plots, custom plots of single or composited data sets for users, and scribe and peel-coat stable base plastic film for publication purposes. A continuing aspect of the GIS will be the support of State cooperative users and others within and outside USGS in the utilization of the GIS as well as the provision of spatial data, plots and statistical analysis, and exportable computer programs of the GIS.

Digitization of Polygon Map Data

The graphic input procedure includes the initial conversion of maps into digital data, the editing process by which logically correct or "clean" data files are produced, and generating polygon and line files for further manipulation and data retrieval.

A number of alternatives are available both in hardware and method to accomplish the initial conversion to digital data. Two prime considerations in selecting a method are: 1) hardware independence and 2) minimization to digitizer-operator induced errors. The graphic input procedure, under development at USGS since early 1973, attempts to accomplish this by: 1) limiting the amount of information required from the digitizer and 2) using software to perform automatic editing, error checking, and file creation.

The digital data from any given digitizer must be identified by map title, data, and whatever other textual description may be necessary. The digitizer coordinate frame of reference must be defined by map projection and the digitization of selected control points defined by geographic (latitude-longitude) coordinates. Beyond these initial identification data, the graphic input procedure requires two files: 1) unlabeled lines defined by series of (x,y) coordinate pairs and 2) polygon labels, each tied to an arbitrary point within the polygon. The following is an excerpt from instructions given to digitizer operators by the USGS. Note that operators (or scanners) are not required to tag or label lines in any way. All that is required while digitizing lines is the ability to recognize line intersections (nodes): "All maps consist of lines. All points which are common to three or more lines are called nodes. Lines joining nodes will be called arcs and will consist of one or more line segments joining the digitized points to each other. The boundary of the map will intersect some of these arcs. For the purpose of digitizing a particular map, arcs intersecting the boundary of the map will be terminated by the intersection, which will then become a node. The straight line segments of the boundary joining adjacent nodes will also be arcs. An arc will be completely defined by a string of x,y coordinates beginning with the node at either end and proceeding segment by segment to and including the node at the other end. The lines of the map will be completely defined when each arc, including arcs which are the boundaries of the map, has been digitized at least once. The only points which need be digitized more than once are the nodes.

The arcs of the maps can be chained together at their common nodes around each completely closed area. These areas, which are assumed to be homogeneous in the characteristic being mapped, are called polygons. Polygons which cross the boundary of the map are assumed to end at the boundary. A single polygon surrounded by one larger polygon is a simple island. A group of adjacent polygons totally surrounded by one larger polygon is called a compound island and contains identifiable nodes at arc junctions. For purposes of digitization it will be necessary to choose one point on the boundaries of simple islands to be the node. This node will be both the first and last point in the coordinate string defining the arc which is the boundary of the simple island. The direction in which an arc is digitized is immaterial.

Every polygon of each map will be identified by a not-necessarily-unique integer of from one to nine digits. A polygon will be completely defined when its identifier has been recorded at least once followed by the coordinates of some arbitrary interior point within the polygon which is neither on an arc nor within an included island."

The types and sophistication levels of equipment which has been used to supply initial data to the graphic input procedure span a wide range of graphic input devices, from a simple manual encoding and card-punch operation to an automatic line-following laser scanner. Any device which can translate lines into series of (x,y) coordinate pairs may be used. Most often this has been some variety of digitizing table with a manually operated cursor.

Experimental projects digitizing land use, land ownership, census tracts and subdivisions, drainage basins, geology, soils, and State and county boundaries of regional test sites in the CARETS area were conducted in 1973 using the Bendix Datagrid table. The success of the project demonstrated the feasibility of using manually operated digitizing tables to convert land data compiled as polygon maps to digital form.

As a followup from the preliminary digitizing of CARETS maps, production digitizing of the Land Use Data and Analysis Program (LUDA) data began in the fall of 1974 using a Computer Equipment Corporation digitizing station operated by the Johns Hopkins Applied Physics Laboratory. Land use and land cover, land ownership, political boundaries, census subdivision, and hydrologic units covering the entire State of Louisiana at scales from 1:125,000 to 1:250,000 were digitized, edited, and delivered to the Office of State Planning in Baton Rouge by the end of fiscal year 1975. That office has used the data and a copy of the system software to manipulate the data themselves, generating land use and land cover area summary statistics by parish and various special plots and retrievals. A special application involved the assessment of April 1975 flood damage, given within a week after the flood. Statistics showing the total acreage of each land use and land cover type flooded within each parish were given.

Many problems have been encountered and much has been learned in the past 2 1/2 years about processing data digitized on tables with manually operated cursors. The number of problems and therefore the amount of editing necessary has varied with the machine or digitizing station used and the skill of the operator.

A wide range of capabilities exists from machine to machine. Some, designed for automatic drafting work, have proved unusable due to the high inertia of the cursor or arm while attempting to depart from a straight line in the x or y direction. Most tables create problems merely because they do not permit the operator to view his work until after he has completed the digitizing operation, if at all. Problems created using such "blind" tables include: 1) missing data, either lines or polygon lables; 2) lines digitized more than once; 3) incorrect labels; and 4) loss of origin, resulting in different frames of reference for various parts of the same map. Digitizing systems which allow viewing of the data while digitizing and some degree of backtracking capability produce much better results.

The experience and skill of the digitizer operator plays an even more important role in the quality of the manually operated digitizer table output. Map topology must be understood and care must be taken so that control points are digitized correctly and lines followed exactly. If the digitizer mode[1/] used calls for the operator to select the points by which a line is described, even more care must be taken. Constant attention must be paid so that line intersections (nodes) are recognized and two arcs are not digitized as one. Labels must be correct and good records kept so that all labels and lines are digitized. It must be said, however, that in spite of the problems inherent in using manually operated digitizing tables, most of the editing of USGS land use and land cover data has been necessary not because of digitizing errors, but because of compilation errors.

Once extensive experience with manual digitizing table output was attained within the USGS, a decision was made in 1974 to contract digitizing with I/O Metrics Corporation of Sunnyvale, California. I/O Metrics uses an extremely accurate laser scanning device designed originally to digitize bubble and streamer chamber photographs. The scanner is used in a "line-following" mode rather than raster scanning so that no raster-to-vector conversion is necessary and label data may be present on the film as well as lines. Production digitizing by I/O Metrics began in the spring of 1975 (figure 1, events F7).

The use of the laser scanner solved many of the problems inherent in manual digitizing, but brought up new problems to replace the old. Most of these problems have been dealt with by the Compilation Branch and I/O Metrics.

1/ Three modes may be used: 1) points generated as a function of time; 2) points generated as a function of distance traveled; and, 3) points selected by the operator.

The first problem has involved the preparation of the maps to be digitized. Since the scanner just "sees" opaque and transparent areas, the lines must be distinct, solid, and opaque while everything else within the laser beam's search radius must be transparent. While a manual digitizing operation can be done with a manuscript compilation copy, the scanner demands two film positives be made with just lines and labels (no labels may be within a given tolerance of a line). Two film copies must be provided as the scanner can only digitize a 3- by 4-inch (approximately) area at a time and the film positives must be cut into overlapping strips. The Topographic Division photo labs superimpose a grid upon the overlays to delineate these 3- by 4-inch rectangles. Problems have occurred where this grid falls directly upon an internal map boundary. The most serious problems, however, have come from the poor quality of the line data after reduction and photo processing of the original inked compilation maps. A negative must be made before the film positive. As ink lines naturally vary in thickness and density, weak lines on the original drop out and strong lines "bleed" into neighboring lines and labels. This has meant that either the digitizer operator at I/O Metrics must constantly reposition the probe manually while digitizing or the original map must be scribed. Experience has shown that if the original maps are scribed in compilation with digitizing by automatic scanner in mind, considerable savings may be made. Since many inked or printed maps are compiled without consideration of the requirements of the digitizing operation, however, the cost of the preparation of the material for scanning may equal or surpass the cost of manual digitizing and editing.

If the input requirements of the automatic scanner are met, whether in compilation or extra preparation, the digitized data returned are far more accurate and error-free than any manual digitizing operation. In fact, this accuracy at first created a problem in processing the data as returned. I/O Metrics returns to the Geography Program data describing the lines of a map by a series of points <u>8 mils apart or less</u>. The 125 to 200 points per inch means the lines are described to a far more accurate degree than that to which they were compiled. The weeding algorithm used successfully with manually digitized data could not operate cost effectively on the I/O Metric data. A new algorithm had to be developed using a look-ahead tolerance corridor search technique, the last point within the corridor being kept and used for the start of the next look-ahead corridor. Using this algorithm, we have routinely reduced the number of points to 10 percent or less of that returned by the scanner while retaining the accuracy of the original compiled lines. The data have proved to be very free from errors when compared to manual digitizing results.

Editing Process

After the polygon map data have been captured, regardless of the device used, the raw digitizer output consists of three parts: 1) geographic control data, 2) unlabeled series of (x,y) points describing the lines or arcs of the map, and 3) labeled polygon interior (position arbitrary) points. To produce final files the data are put through the following editing steps (see figure 7):

1. Conversion to standard format;
2. Data compaction (weeding out points unnecessary to define lines within a given spatial tolerance);
3. Conversion to latitude-longitude coordinates, if desired;

4. Split into either 7-1/2 minute sections or other sections, if desired or necessary;
5. Limited automatic editing and error detection of line (arc) data;
6. Manual batch or interactive editing of line data with a return to previous step until line data are "clean";
7. Merging of labeled polygon points with the line (arc) data, providing either further error detection or clean files;
8. Manual batch or interactive editing of polygon label data and a return to step 6 or 7, if necessary;
9. Edge match of each map section with neighboring map sections either in this same map or others already on file, if necessary.

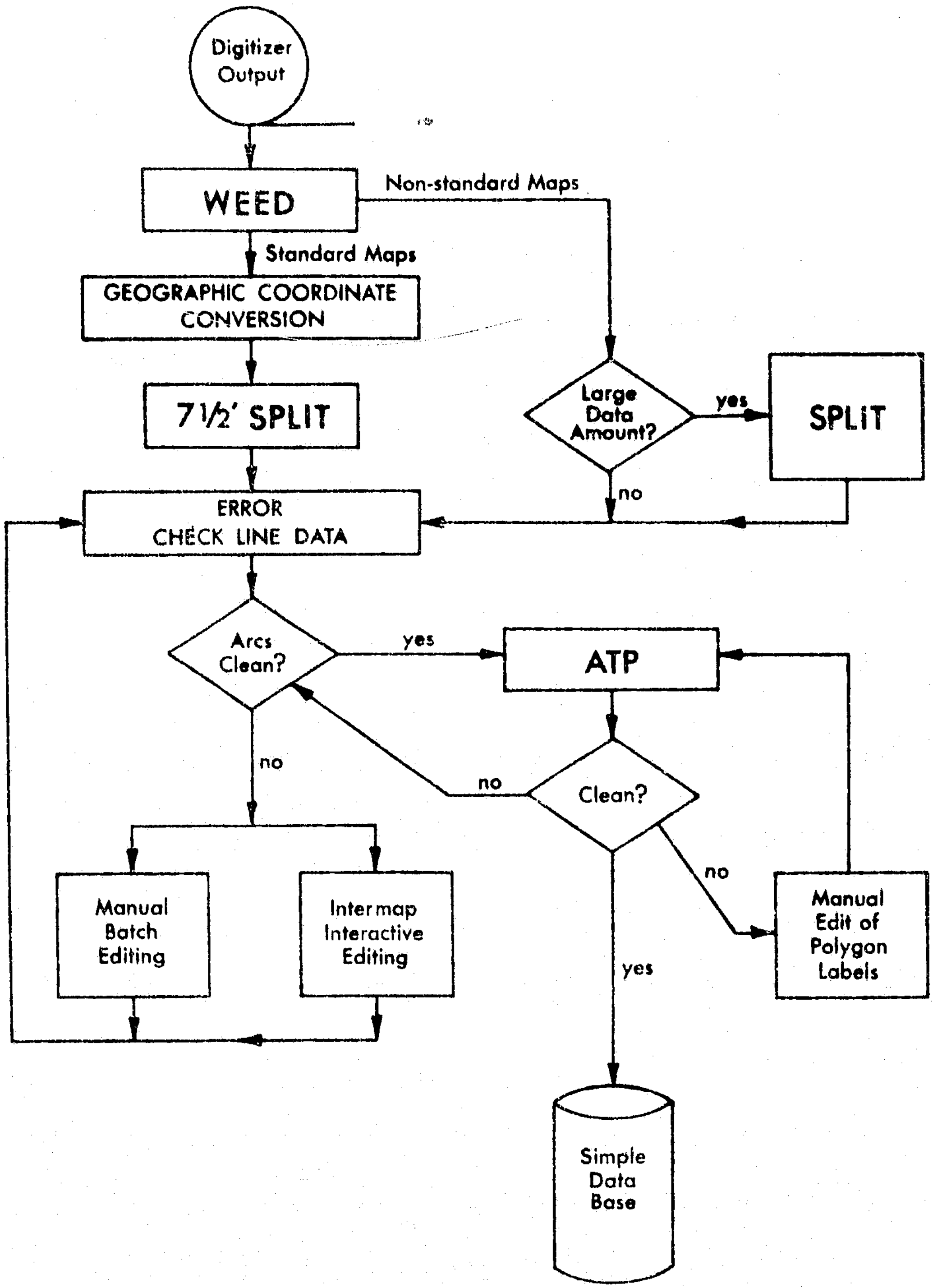


Figure 7—Graphic input procedure flow diagram

All raw digitizer data is first converted to a standard format. The program used must, of course, be different for almost every different digitizer. The USGS has processed data from six different sources, each requiring a special format conversion program. For I/O Metrics data the first two steps (format conversion and data compaction) are being done simultaneously to reduce input/output charges.

The data compaction step or weeding has already been mentioned in connection with processing I/O Metrics data. This is a very important step since all further editing, manipulation, and retrieval costs are dependent upon the amount of data which must be processed. The basic principle is that in order to spatially describe a given straight line segment only the two end points of that segment need be stored. An efficient cost-effective and logically correct algorithm must be used to ensure optimum data compaction while retaining positional accuracy. If a large amount of data is to be weeded, the algorithm should require that each input point be "looked at" (calculations performed) only once. The USGS uses the following algorithm: 1) use the point kept last (first point in string is always kept) and the next point in the series (at a greater distance than the given tolerance from the first point) to define the center line of a tolerance corridor; 2) calculate the distance of each following point to the center line of the corridor until a point falls out of the corridor or the end of the line is reached; 3) save only the last point in the string within the tolerance corridor; and 4) return to step 1 if the end of the line has not been reached.

If the maps being processed are standard USGS land use and land cover maps or related overlays, the weeded data are then transformed into a standard Mercator-like projection where a given x or y coordinate value may be equated to a latitude or longitude value, but the longitude values vary with the cosine of the latitude. This allows easy splitting of the data into 7 1/2-minute sections while retaining scale consistency and spatial relationships from section to section. The algorithm used to transform the data into this projection assumes the data was digitized from bases in an approximate Universal Transverse Mercator projection. Control points at the corners (must be 7 1/2-minute intersections) of the geographic "rectangle" bounding the map plus control points at the midpoints of the latitude lines of the rectangle are used to accomplish the conversion by single precision Aitken interpolation. If no projection transformation is desired, the data may still be split into sections along UTM or arbitrary lines. Experience has shown that the cost of editing may be reduced significantly by splitting map data into sections. In any case, if the number of coordinates needed to describe the lines of the map exceeds 32,767, the map must be split (only 16 bits are used to store arc pointers). This has meant that before maps were split into 7 1/2-minute sections, most land use maps covering 1:250,000 quadrangle had to be split into six to nine sections. As part of the split routine arcs are added along split lines to form a complete continuous outside boundary to each section.

As the last step of a single automatic production procedure begun with the raw digitized data, each map section line data are edited and checked for node errors.

Manual editors use the error listings generated by the arc edit program plus a plot of the data or the INTERMAP system along with a copy of the original compilation manuscript to visually check the data and decide what edit commands are needed whether in batch or interactive node. The edited arc data are again checked for node errors until "clean".

The heart of the graphic input procedure is the arc to polygon program (ATP) the step which ties the labeled polygon interior points to the arc data, creating the cross-referencing polygon-arc files. ATP serves as the final stage of topological error check and verification. Most of the polygon label errors and compilation errors are caught in ATP as well as false arc crossings (if a given arc does not recross another arc) and arcs duplicating two or more arcs. Errors listed may indicate a need for more arc editing and/or editing of the labeled polygon interior points. As duplicate labels and misplaced labels are automatically deleted, the editing of labeled polygon points consists of either changing a label or adding a labeled point.

The final step in building the files is edge-matching the outside boundary arcs of each map section with neighboring map sections either from the same map or other maps previously digitized and filed. Often this means a split of one boundary arc into two to ensure unique identification of attributes left or right of the arc. Note that this edge match process not only means the completion of the map presently processed but the update of edge information of maps already stored.

Much has been done to perfect the system and to streamline the operation. The large amount of data processed has helped debug the software and has illustrated many improvements which either have been or will be made. The number of man hours needed to edit a typical 1:250,000-scale quadrangle land use and land cover map has been decreased from something in excess of 200 to less than 40. More and more production time is being spent in file maintenance and bookkeeping rather than actual editing.

INTERACTIVE DIGITIZING AND CARTOGRAPHIC EDITING SYSTEM

After its review of USGS and CARETS digital mapping requirements, the IGU advisory group recommended that USGS acquire the hardware necessary to support programs developed by Ray Boyle of the Department of Electrical Engineering at the University of Saskatchewan (figure 1, event G2). Definition of the hardware specifications was completed and the decision to purchase the equipment was made in August 1973.

INTERMAP is a cartographic digitizing and editing system based on the Digital Equipment Corporation (DEC) PDP8/e Digital Minicomputer and various interfacing hardware. In addition to IOS/8 (a modification of DEC's OS/8 software support system), a large number of programs provide capabilities for digitizing map sheets into map files that can be stored on any IOS/8 file structured device, transforming the data from digitizer table coordinates to actual map coordinates, editing the data outlined using Tektronix storage display terminals, editing the raw digital data, and converting the data to IBM compatible 9-track magnetic tape format from IOS/8 file structure format (and vice versa). With the above capabilities the system can operate in "stand alone" or as a digitizing and editing system providing data on IBM compatible tape for larger computer systems. The USGS acquired INTERMAP[1/] to develop an in-house geographic information systems laboratory and to allow in-house digitization and interactive editing of the digitized polygon files.

The PDP8/e minicomputer and an extended arithmetic unit for high-speed calculations are the main components of the hardware system. The system is based on disk operation, which provides the speed necessary to handle large

1/ When initially acquired in 1975, the INTERMAP System was called CART/8.

amounts of data. The Tektronix storage display units with an interactive pointer provide the graphic interface to the user. The magnetic tape unit provides for bulk storage and input/output to other computer systems via a 9-track IBM compatible magnetic tape.

The hardware originally acquired by the USGS included the following:

1) Two Tektronix 611 CRT storage display units;

2) A PDP8/e minicomputer with 16,000 octal words of memory, each word being 12 bits in length; power-fail auto-restart hardware:

3) A magnetic tape unit;

4) A standard RCA teletype terminal for standard user keyboard input;

5) One disk drive utilizing removable disk cartridges, each cartridge having a capacity of 1.6 million words;

6) A Gradicon digitizer;

7) An extended arithmetic element;

8) Interfaces for the Tektronix display units and the Gradicon digitizer to the minicomputer.

INTERMAP systems can be used to digitize, display and edit point, line, symbol, alphanumeric, and polygon data. Editing can be done at three times: (1) as data are being digitized, (2) after data have been digitized, and (3) finally when the data have been extracted for a special use, such as preparing overlays for automatic map drafting. All displays of the data may be symbolized. The data can be used either within the INTERMAP system or put in a standarized format, stored on magnetic tape in IBM compatible 9-track format, and transferred to other systems. In this manner the INTERMAP system can be regarded as being compatible with existing customer programs and other computer systems.

For digitizing, the INTERMAP system utilizes a Gradicon table with a handheld digitizing tool or cursor controlled by the PDP8/e and a Tektronix 611. The INTERMAP system is considered an "intelligent" digitizer. "Intelligent" digitizer software refers to the ability of DIGIT Programs to warn the operator when he moves the cursor too rapidly for accurate line-following, tell him when he has closed an island (polygon), and inform him when he has closed a line during a backtrack procedure.

The first operation is to digitize a number of control points on the base map. These control points are later used to translate, rotate, and scale the digitized table coordinates into a rectangular coordinate system (INTERMAP uses data with lat-long, UTM, digitizer or any decimal rectangular coordinate system).

Once the control points are digitized, the digitizer operator has any number of methods of digitizing (with preliminary editing features available). He may digitize in a point mode. He may also digitize lines in a stream mode. Thus, he can digitize contours, hydrography, or any other line information on the base map.

At the same time, he is able to specify feature codes for any of the map attributes that he is digitizing. This feature selection is done using a series of thumb wheels or keyboard characters supplied with the INTERMAP system.

The digitizer software includes several preliminary edit capabilities. An automatic island closure permits the digitizing of closed arcs. There is also a backtrack capability, which allows the operator to stop digitizing at a point and move back any distance along a line, and recommence digitizing.

All of the information before the point where he recommences digitizing is saved and all the previous information after that point is discarded. The user is aided by an audio device attached to the digitizer table that emits one tone as long as he is digitizing in a proper mode. If his hand moves abruptly, the tone will change and indicate the need to backtrack. The audio device aids in backtracking by changing tone when the previously digitized line is crossed.

A second mode of input is from outside sources. For example, land use and land cover data digitized by a commercial contractor may be edited using the INTERMAP system. In this case, two INTERMAP utility programs are used to get the information from the outside source into a format compatible with the INTERMAP system. Once this has been done, the full manipulative and extractive capabilities of the INTERMAP system can be utilized. The prerequisite for running INTERMAP is the DEC IOS/8 Operating System for the PDP8/e.

Once the data are available in the INTERMAP format, the user has the interactive display and editing (DSPEDT) capabilities at his disposal. Line-mode editing capabilities can be used for line or polygon data, permitting the removal of lines, or parts of lines, addition of lines or parts of lines, removal of spikes, filling in of gaps, and cleaning up of junction closings. For symbol data, symbols may be added or deleted, or moved about within the data base. Alphanumeric data, such as the name of a lake, can also be changed in the data base. For example, the user can change the spelling of a name, delete the name, or change the position of individual characters in that name. He can thus take the name of a river and manipulate the individual characters to follow the shape of the river much as is done in lettering on a topographic map.

All manipulative capabilities are closely associated with the data extraction capabilities of the INTERMAP system. More precisely, each time the user wishes to edit a portion of his data base, he need display on the cathode ray tube only a small portion of the data base. The data can be magnified as much as 32 times and special line symbols, such as dashing, line thinning and thickening, or overlay of name and symbol information can be done interactively.

Delivery of the hardware required for the INTERMAP system was gradual, with the minicomputer being delivered in March 1974, the display equipment in June 1974, and the digitizer table and interface equipment in December 1974 (figure 1, events G4 through G6). The system was finally implemented with the INTERMAP software in December 1974 (figure 1, event G7).

As a result of the extended time needed to acquire and implement the interactive cartographic system, it was not available to the CARETS project for the production of digitized maps within the time limitations of the project. The system is being used intensively in the editing and correcting of USGS land use and land cover maps, however, and use of the system is anticipated for future processing and analysis of the CARETS data base resulting from other CGIS work. In addition, the interactive cartographic system--although not procured from CARETS funds--is one of the accomplishments in the development of a geographic information system in which the CARETS project play a part.

CONCLUSIONS

This report has described the problems, decision points, activities, and research involved in designing, implementing, and operating a geographic information system to input, store, manipulate and retrieve digital spatial data developed from land use and other natural resource data sets. The system began with the CARETS project, but was expanded and extended over the 4 years covered by this report. The attempt to provide the CARETS project with a geographic information system, while not altogether successful during the life of the project, provided the incentive and support for a research and development activity which is continuing to cope with managerial, technical, and scientific problems characteristic of such systems.

One of the more serious problems is the significant time delays that occur in obtaining project approval, acquiring equipment and facilities, hiring and training staff, and moving from the design and development stages of a GIS to its implementation. The sequence of activities shown in figure 1 provides multiple examples of the magnitude of that problem. GIS development could not keep pace with the CARETS schedule. This scheduling problem was further complicated by the fact that the interpretation and compilation of CARETS maps involved conventional techniques which did not depend upon a new technology, acquisition of specially trained personnel or special equipment, or the development of complex software. CARETS data compilation activities, therefore, proceeded at the start of the project, while GIS development required an extended start-up period and soon lagged behind the overall CARETS timetable.

Reliance on the assumption that operational geographic information system software, techniques, and procedures would be generally available for importation and utilization in the CARETS project was unjustified; even now, the transferability of systems which are generally considered operational (see the five systems described by the IGU Commission and systems considered under "Survey and Evaluation of Existing GIS Approaches") cannot be considered feasible because of inadequate documentation, machine dependency, or proprietary restrictions. It should be noted that even the utilization of the Canada Geographic Information System--a truly operational system--was extensively delayed over the life of the CARETS project due to the time necessary to negotiate with the Canadian Government, conduct evaluation tests, and consumate an agreement. As a result, the digitizing and editing of CARETS maps within the extended time frame of the project actually could have been accomplished by the USGS GIS, although the production of statistical summaries would have taken longer (see figure 1, events C7 and F7).

It might be argued that the selection of a grid-cell approach from the alternative methods of spatial data encoding that were available at the start of the CARETS project might have avoided the developmental problems attendant on the arc-polygon approach used by the GIS. At the time, the grid-cell method was in use by several organizations and its adaptation might have reduced the start-up and lead times experienced in the CARETS project. A preliminary analysis of the characteristics, advantages, and disadvantages of the various approaches, however, strongly indicated that future requirements in handling spatial data should dictate that approach which would most closely accommodate all types of spatial data--point, line, and area data--and provide the highest degree of flexibility possible in storing, manipulating, retrieving, and outputting that data. In addition, the early promise of the development of a geographic information system by the Raytheon Corporation which would afford highly useful manipulative and output facilities made a strong argument for the arc-polygon approach to encoding. Subsequently, the IGU advisory group's analysis of five spatial data encoding approaches actually did confirm the suitability of the decision from the standpoint of flexibility and accuracy.

It is on the management side of designing, developing, and implementing a GIS, however, in which the experiences detailed in this report vary from the procedures and criteria that are generally considered good practice. The comparative analysis of five systems by the IGU suggested that recommended practice in system development includes: a) conceptualizing the system and entrepreneurial activities, b) research management, c) system debugging, and d) system operation and maintenance. The managerial experience relating to these phases suggests the following:

1. A long-term staff plan which recognizes substantial staff changes through the phases of design, development, and implementation.

2. A fiscal plan which relates to the total resources necessary to develop the system and which provides for fiscal continuity.

3. A program for "selling" the system while at the same time providing interim products to potential users.

4. An extensive education program for users relating to applications problems and techniques of using the system to attack those problems.

5. A feedback system which allows the user to communicate his experiences in using the system.

System development in the CARETS project and the USGS contrasted with accepted procedure in that, from the beginning, system operation, conceptualization and development, debugging, and research management were carried out in parallel. Shortly after the initiation of the CARETS project, the immediate need to encode and edit CARETS spatial data sets imposed a clear and urgent requirement for the development of a graphic input procedure. Successively, the production of data sets in the USGS and the requirements of an expanding number of State cooperative programs with the USGS intensified the pressure for parallel system development with an equal or even greater production effort and to place developmental emphasis upon the spatial data processing phases of the system rather than upon the system as a whole.

The arbitrary emphasis dictated by large-scale processing requirements for spatial data coupled with the strong and insistent interest of users in the preliminary products of spatial data processing still complicates the rational development and implementation of the GIS and exacerbates many of the system management problems because of the competiton between production and development for fiscal support, personnel, time, and facilities. Overemphasis on the development of the input and output facilities of a system exists because these two phases of system development are most understandable to potential supporters and users of the system, who are likely to be interested only in what goes into it or what comes out of it. Senior management and applications users generally are not concerned with the other more complex aspects of system development such as manipulative techniques, data base management, and interactive retrieval. The result of that emphasis has been to place stress upon the effort needed to produce deliverable products at the expense of overall system development and to refocus staffing and fiscal plans from long-term development to short- and near-term production requirements. Meanwhile, total system development time is continuously deferred or extended and the possbility increases that later system development may render interim products useless.

Nonetheless, the considerable advantages of paralleling the development of a GIS with a reasonable and balanced system production effort cannot be understated. Despite the management problems inherent in the approach, and recognizing the difficulties in maintaining a balance of resources and emphasis between production and system development, a continuation of the present GIS objective is indicated: to integrate user education and interim deliverable products within the continuing development of the system without excessively delaying it.

REFERENCES

Arms, Samuel, 1968, MAP/MODEL system--technical concepts and program descriptions: Columbia Regional Assoc. of Govts., Portland, Oreg.

_________ 1969, MAP/MODEL system--data entry and retrieval: Bur. of Governmental Research and Service, Univ. of Oregon, Eugene, Oreg.

_________ 1970, MAP/MODEL system: system description and user's guide: Bur. of Governmental Research and Service, Univ. of Oregon, Eugene, Oreg.

Boeing Computer Service, Inc., 1972a, Natural resource information system--volume 1, overall description: Seattle, Wash.

_________ 1972b, Natural resource information system--volume 2, design and analysis: Seattle, Wash.

Dangermond, Jack 1971a, Polygon information overlay system: Comprehensive Planning Org., San Diego County, San Diego, Calif.

_________ 1971b, Proposed information system for planning: Comprehensive Planning Org., San Diego County, San Diego, Calif.

Environment Canada, 1973, Canada geographic information system--overview: Lands Directorate, Ottawa, Ontario, Can.

Raytheon Corporation, 1973, The development of a natural resource information system v. 1-5: Prepared by Raytheon corporation under U.S. Dept. of Interior Contract No. K5 1C 14200693, Wayland, Mass.

Tomlinson, R.F., ed., 1972, Geographic data handling, UNESCO/IGU second symposium on geographical information systems v. 1-2: IGU Comm. on Geographical Data Sensing and Processing, Ottawa, Ontario, Can.

Tomlinson, R.F., Calkins, H.W., and Marble, D.F., 1976, Computer handling of geographical data: Paris, UNESCO.