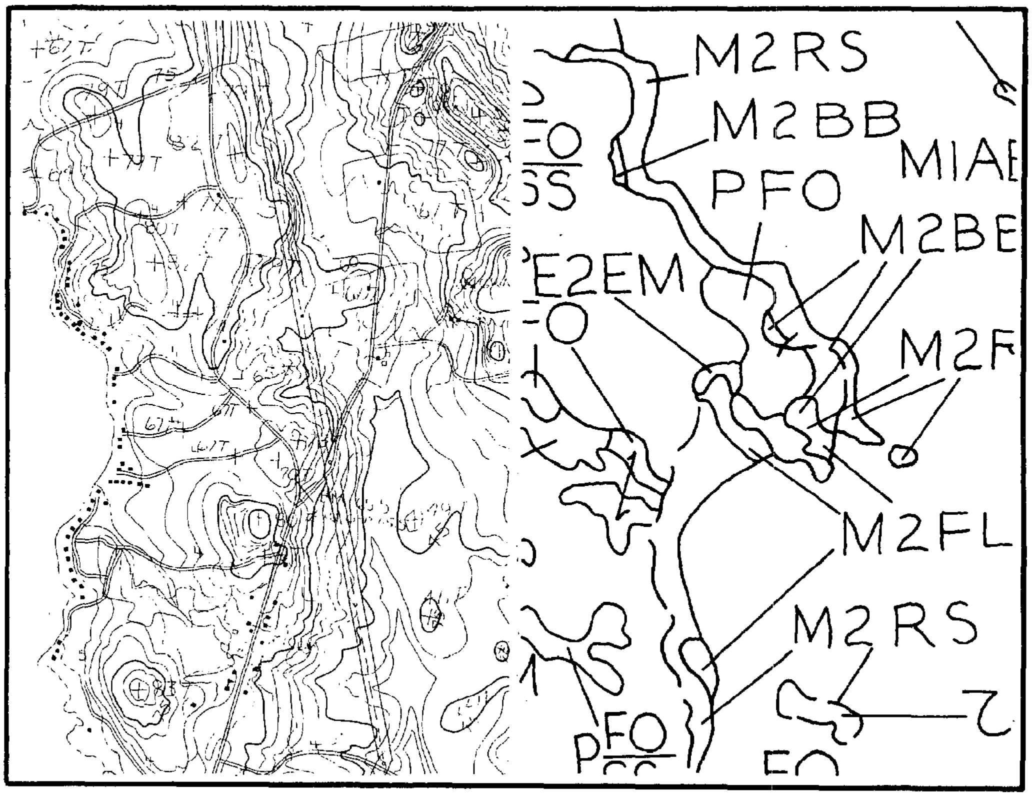
18656

INVESTIGATION INTO THE FEASIBILITY OF ESTABLISHING A STATEWIDE GEOGRAPHIC INFORMATION SYSTEM

November 1979



...urces Program
...ource Planning Division
...e Planning Office

Maine State Planning Office. Water Resources Programs.

INVESTIGATION INTO THE FEASIBILITY OF ESTABLISHING A STATEWIDE, NATURAL RESOURCES, GEOGRAPHIC INFORMATION SYSTEM FOR MAINE (FINAL REPORT)

OCTOBER, 1979

Maine State Planning Office and

Maine Land and Water Resources Council

Acknowledgements:

The principal author was CHARLES YUILL, a Massachusetts Audubon Society environmental intern working with the State Planning Office. MAL CAREY of the Social Sciences Research Institute, University of Maine, prepared Section V; JOSEPH CHAISSON of the State Planning Office directed this study, prepared portions of the report, and provided final editing. Many thanks are given as well to other state and regional agency staff members who patiently assisted this project through a series of interviews and by providing review comments.

Financial assistance for the preparation of this report was provided by the Department of Housing and Urban Development under their Comprehensive Planning Assistance (Section 701) Program; the U. S. Water Resources Council, and the Massachusetts Audubon Society Environmental Intern Program.

TABLE OF CONTENTS

APPENDICES:

RECOMMENDATIONS AND EXECUTIVE SUMMARY

A. RECOMMENDATIONS:

1. Because of the current level of costs associated with establishing a statewide GIS, the anticipated steady decline in computer hardware costs, and the federal government commitment to establish a number of GIS's with instate applications, MAINE SHOULD NOT PROCEED AT THE PRESENT TIME TO ESTABLISH A STATEWIDE GIS. HOWEVER, DEVELOPMENT AND EVENTUAL ESTABLISHMENT OF AN EFFECTIVE STATEWIDE GIS THAT WOULD INCLUDE CAPABILITIES ADEQUATE TO MEET REGIONAL PLANNING AND RESOURCE MANAGEMENT NEEDS AND THAT WOULD BE COMPATIBLE WITH EMERGING CADASTRAL LAND RECORD SYSTEMS SHOULD BE CONSIDERED A LONG-TERM GOAL OF THE STATE.
2. A DATA COORDINATOR SHOULD BE DESIGNATED WITHIN THE STATE PLANNING OFFICE. INCLUDED WITHIN THIS COORDINATOR'S RESPONSIBILITIES WOULD BE:
   A. PROVISION OF STAFF SUPPORT AS NECESSARY TO THE DATA MANAGEMENT SUBCOMMITTEE OF THE LAND AND WATER RESOURCES COUNCIL;
   B. PURSUING IMPLEMENTATION, EITHER DIRECTLY OR AS A PROJECT SUPERVISOR, OF THE GIS RECOMMENDATIONS THAT FOLLOW.
3. USER NEEDS ASSESSMENTS FOR DATA STORAGE, RETRIEVAL, AND ANALYSIS SHOULD BE UNDERTAKEN ON A REGULAR (BI-ANNUAL) BASIS. THESE ASSESSMENTS SHOULD INCLUDE USERS AT BOTH STATE AND REGIONAL LEVELS AND SHOULD EMPHASIZE DATA NEEDED TO ADDRESS PRIORITY RESOURCE MANAGEMENT PROBLEMS.

4. THE STATE NATURAL RESOURCE DATA COLLECTION PLAN SHOULD BE COMPLETED AND USED AS THE BASIS FOR AN ACTION PLAN TO PROVIDE THE COMPLETE DATA BASE NECESSARY FOR AN EFFECTIVE STATEWIDE GIS.

5. A PROCESS SHOULD BE INITIATED TO CONTINUE THE DESIGN OF A GIS, SPECIFICALLY INCLUDING THE DEVELOPMENT OF METHODOLOGIES FOR THE ANALYSIS OF MAINE NATURAL RESOURCES DATA THAT WOULD BE CONTAINED WITHIN A STATEWIDE GEOGRAPHIC SYSTEM.

6. ADVANCES IN SOFTWARE FOR GIS APPLICATIONS AND THE EVOLUTION OF CAPABILITIES AND COSTS OF COMPUTER HARDWARE SUITABLE FOR GIS APPLICATIONS SHOULD BE REGULARLY MONITORED.

7. PROGRESS MADE BY FEDERAL AGENCIES TOWARDS DEVELOPMENT OF AN AUTOMATED DATA BASE FOR MAINE AND INFORMATION SYSTEMS THAT USE SUCH DATA SHOULD BE REVIEWED ON AN ANNUAL BASIS.

8. THAT ALL CONTINUATION OF WORK TOWARDS THE DEVELOPMENT AND EVENTUAL ESTABLISHMENT OF A STATEWIDE GIS BE CONDUCTED UNDER POLICY GUIDANCE OF THE LAND AND WATER RESOURCES COUNCIL.

9. IF A STATE OR REGIONAL PLANNING AGENCY INTENDS TO ESTABLISH AN AUTOMATED GIS FOR EITHER GENERAL USE OR FOR SPECIFIC PROJECT WORK, THE AGENCY SHOULD SUBMIT A DETAILED PROJECT DESCRIPTION TO THE DATA MANAGEMENT SUBCOMMITTEE OF THE LAND AND WATER RESOURCES COUNCIL FOR REVIEW. THE SUBCOMMITTEE SHALL REVIEW SUCH PLANS AND MAKE COMMENT

ON THE RELATIONSHIP OF THE SPECIFIC PROPOSAL TO ONGOING PLANS FOR ESTABLISHING A STATEWIDE GIS. THE OBJECTIVE OF THIS REVIEW PROCESS IS TO ACHIEVE THE MAXIMUM CONSISTENCY WITH AN EVENTUAL STATE SYSTEM THAT IS POSSIBLE WITHIN SPECIFIC STATE OR REGIONAL AGENCY PROGRAM REQUIREMENTS AND TO AVOID UNNECESSARY DUPLICATION OF EFFORT.

B. SUMMARY

The specific purpose of this study has been to investigate the feasibility of establishing a statewide geographic information system for the storage, retrieval, and analysis of spatially distributed natural resources data. The work outlined in this report is a preliminary investigation of Geographic Information Systems (GIS) and their relationship to land planning and management in Maine; it contains a series of findings and recommendations adopted by the L&WRC concerning state GIS development.

As envisioned in this report, the goal of a "Maine Natural Resource Information System" is to provide the best possible vehicle for agencies to fulfill their statutory requirements and satisfy needs in terms of resource planning, development, management, and the conservation of Maine's natural resources. Natural resource related information is collected by various state agencies in accord with their separate legislative mandates. However, agency specialization alone does not identify completely the type of information that is being accumulated or what is available to other agencies. At the present time specialized data systems are functioning at the regional and state level in Maine, but the mechanism does not exist to link, in common network, the sum of sources and users of natural resource data/information. It is important that the data/information needs of the agencies involved be met in the most cost-effective and efficient way possible. A geographic information system can accomplish this task by functioning as an integrating force, not by interfering with primary data gathering activities, but by properly tapping existing and potential sources, services, funding and systems in a way that will effectively unite the user with needed data.

In Maine the value of such a system by necessity would have to be judged in terms of addressing:

- Consistent format, quality, standards and scales of data
- Improved indexing and storage of existing information
- Improved coordination of ongoing planning programs
- Less duplication in data storage
- Elimination of conflicting data
- Eventual hardware and software compatibility

Such a system need not be totally a system in the strict engineering sense. It should merely be a service mechanism for 1) assembling data in both machine readible and non-machine formats; 2) processing raw data into meaningful information; 3) adjusting and organizing data into formats and forms suitable for modern storage, retrieval, and manipulation procedures; 4) storage of data in a systematic information base; and 5) displaying this data in graphics, models, maps, studies, and simulations appropriate to specific audiences and problems.

Basic to the implementation of a successful system are three critical elements: 1) knowledgeable users of the system; 2) the requisite data base; and 3) operational system capabilities determined by the needs of the users. These elements are the focus of this study.

C. STUDY FINDINGS

1. The establishment of a statewide GIS would benefit Maine by

    a) Improving the availability of information to support state resource management and decision making activities;
    b) Reducing the costs of analysis and application of natural resource information compared with current, decentralized manual analysis techniques; and
    c) Providing information useful in regional (substate) planning and resource management activities.

2. Natural resource information is currently collected and used by many state agencies in carrying out legislative mandates. A wide range of "data systems" exist within state agencies that use natural resources data to support resource planning, management, and decision making activities. These "systems" range from manual use of secondary data to agencies that rely upon in-house, primary data collection with associated automated data analysis.

3. Important data series that would be used by a statewide GIS (soils, surficial geology, land cover, etc.), are in many cases incomplete, inaccurate, or out of date. Data that exists is scattered through files, libraries, and information systems of federal, state, and private organizations. Therefore, the full potential of a statewide GIS would not be achieved until a complete, accurate, and timely natural resource data base exists.

4. The federal government plans to establish an extensive series of digital format, natural resource data files for Maine, and also has developed or is developing a number of GIS's for the storage, manipulation, and retrieval of such data. Many of these systems could be adapted for use at the state level. It is likely that a general federal GIS for natural resources data will evolve, combining major federally established natural resource digital data bases that would be available for use by the state.

5. The Department of Conservation, Bureau of Forestry, is in the process of establishing in-state use of GIS's operated by the U. S. Forest Service, U. S. Department of Agriculture. This involvement is funded under the Renewable Resources Planning Act (RPA), and has included hiring a full-time professional staff person to undertake the necessary programming to support such involvement.

6. Computer hardware suitable for GIS applications is rapidly evolving. The initial and operating costs of such hardware are steadily decreasing at the present time. The technology of performing input and output of geographic data is changing very rapidly with higher quality choices and possibly lower costs expected. A proliferation of excellent computer software applicable to geographic information systems currently exists in stages of application, design, and development. Available software usable for establishing a statewide GIS ranges from that available for free within the public sector to expensive private sector products. Software for GIS applications is evolving at a very rapid rate.

7. A statewide GIS could be established at the present time in the following ways:

   a) In-house system using existing Central Computer Services hardware;

   b) In-house system using a new minicomputer that would be "designed" for and dedicated to GIS use.

   c) Contract with University of Maine

   d) Private contract.

8. Staffing costs associated with any of the four alternatives would be approximately a minimum of $60,000/year.

9. The optimum approach to establishing a state system at the present time would be through an in-house, dedicated minicomputer system. The hardware configuration of such a system would consist of:

   a) Dedicated Central Processing Unit (CPU).
   A stand-alone mini or micro computer system with large address space and substantial computational capability will be required to meet the demands created by GIS development and production activities. A stand-alone system will offer many advantages over shared use of a larger computer facility: radically cheaper cycles, better thruput, control of hours of operation and operating priorities, control of tape and disk mounting and printout dispatching, and a much simpler systems environment.

b) Tape Subsystem.
An industry-standard 1600 bp. tape subsystem will be required for disk back-up and for large-file communication and archiving.

c) Disk Subsystem.
A disk system (possibly using fixed, Winchester drives) is required to provide on-line and working storage of data and programs.

d) Line Printer.
A medium-speed line printer is required to provide reasonable printing turn-around. It is strongly suggested that the unit also have dot matrix plotting capability for displaying pixel-oriented data.

e) Digitizer.
A large-bed (42" x 60") digitizer is recommended to allow convenient digitizing of even, large format maps without requiring segmentation. Interface to the system should be serial for simplicity and flexibility.

f) Graphic Display Tube.
A high-resolution graphic terminal (either storage tube or refresh) is required for quick and inexpensive inspection of geographical data. The interface should be serial for most choices of CPU and graphic terminal to allow future system reconfiguration and enhancement with minimal hassle.

g) Plotter.
A 30-36 inch drum plotter will offer the best performance/price ratio in the early years of a state GIS activity.

10. Data processing hardware that would be used to establish the recommended state system configuration is improving so rapidly that cost estimates are difficult to prepare and would have little meaning. Annual hardware and software costs would probably be less than the minimum annual staff costs.

INTRODUCTION

Interagency cooperation in natural resource planning and management in Maine is encouraged through the efforts of the Maine Land and Water Resources Council (L&WRC). The Council consists of representatives of state agencies with natural resource management and planning-related responsibilities, the University of Maine, Regional Planning Commissions, and the Legislature.

In June of 1976, William Adams, then Commissioner of the Department of Environmental Protection and Council Chairman, requested that a study be done for the Council to determine what Maine's approach should be towards establishing a statewide, automated, geographic information system for natural resources data. Commissioner Adams made this request, in part, because of concerns about the rather large investments some Regional Planning Commissions were making in such systems as part of their 208 programs. Responsibility for advising the State Planning Office in conducting the study was assigned to the to the Council Data Management Subcommittee. Funding and staff resources to conduct the study were not obtained until June 1977. A draft of the study report was widely circulated for review during the spring and summer of 1978. Based upon review comments received, revisions were made to the study report and a set of findings and recommendations for state action were prepared for the Data Management Subcommittee to review.

The Subcommittee met in January 1979 to discuss the draft findings and recommendations. Numerous changes to the recommendations were proposed, agreed upon, and forwarded to the Council for adoption. The Council met on October 25, 1979 and voted to adopt recommendations concerning establishing a statewide GIS as presented in this report.

The purposes of this study are to:

1. Determine to what extent a need exists for a state-level geographic information system (GIS) for natural resources data;
2. Inventory and describe federal activities directly related to the establishment of a state level GIS;
3. Review alternative methods of establishing a state-level GIS; and
4. To recommend what specific actions should be taken by the Land and Water Resources Council concerning the establishment of a state-level GIS.

To accomplish these purposes, the study examines the history, structure, design, and current utility of GIS's and also describes a framework for examining information requirements for land and water resources planning and management in Maine. The examination of information requirements addresses operations associated with data collection, handling, storage, retrieval, analysis, and display.

This study does not recommend either the selection of a specific GIS or the setting of specifications for a specific GIS. Rather, a process is proposed that would work systematically over time to establish an appropriate, well thought out state level GIS.

It is clearly recognized that numerous existing needs of resource planners and managers are currently being met with available statistically based data management systems and various nonautomated spatial data management systems; and we expect this situation to continue. Consequently, it is recognized that, as they are developed, geographic information systems that contain such spatial natural resource data as land cover types or surficial geologic materials, statistical systems containing summaries of such data types as deer kill, and systems concerned with land parcel ownership and boundaries must act together with flexible boundaries established between them.

The study is organized into six sections and a number of appendices.

Section I is an executive summary of the entire study, including its findings and recommendations.

Section II is this introduction.

Section III, "Basics of Geographic Information Systems" provides a brief review of the components of spatial data handling systems, the frameworks within which they are expected to operate, typical products of such systems, and a detailed discussion of the considerations that are important in designing spatial data handling systems.

Section IV, "Elements of Spatial Data System Design," focuses upon laying the groundwork for a continuing long-term data and system availability monitoring program. It is divided into two subsections. The first, "Determination of System Specifications" presents a framework for a long-term work effort whose objective would be to assess the information needs, data handling requirements, and data manipulation capabilities of potential system users. This subsection also contains a preliminary survey of the system needs of various state resource planning and management agencies. The second subsection, "Data Availability," reviews the current and potential availability of system data from federal sources.

Section V, "Hardware and Software Availability," reviews the current availability of software and computer hardware for establishing a state level GIS; and it also proposes an ongoing process for monitoring future developments in these areas.

Section VI, "Determining Specifications for a Geographic Information System," presents a hierarchy of potential system specifications based upon the information concerning data needs, data handling, data availability, and the state of the cut in hardware and software development that was presented in earlier sections.

Appendix A is a glossary of technical terms used in this report.

Appendix B presents a review and evaluation of current and developing GISs in other states.

Appendix C presents a proposed overview description of an integrated natural resources information planning program for Maine; and

Appendix D is a list of references cited in the report.

BASICS OF GEOGRAPHIC INFORMATION SYSTEMS

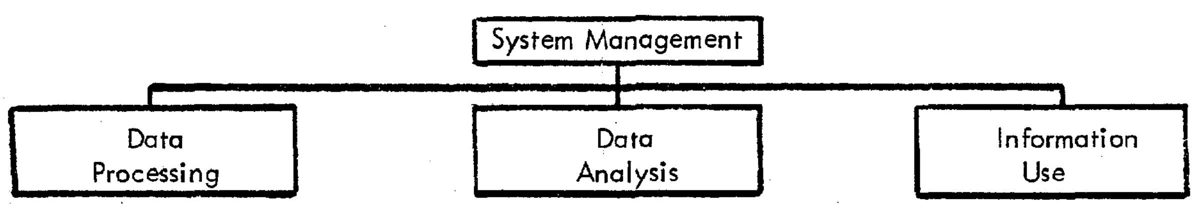
Basics of Geographic Information Systems

3.1 Introduction

3.1.1 Definitions:

A geographic information system is a class of information systems which are defined by specifying the characteristics which qualify information as being geographic. (Tomlinson, 1972). An information system is an ordered combination of data bases, resources (staff, equipment, etc.) and procedures designed to produce information useful to a decision making process. An information system therefor, is user demand driven, the user being defined as anyone responsible for responding to problems faced during the management of a resource. The relationship among the decision making process, systems users and various elements of the information system are outlined in Figure 1.1. The activities central to any information system can be described as being one of four categories or subsystems: 1) management of data base; 2) data processing; 3) data analysis and; 4) use of information (Fig. 1.2).

*Figure 1.2 Basic Information Systems Functions



*Source: Tomlinson, 1972

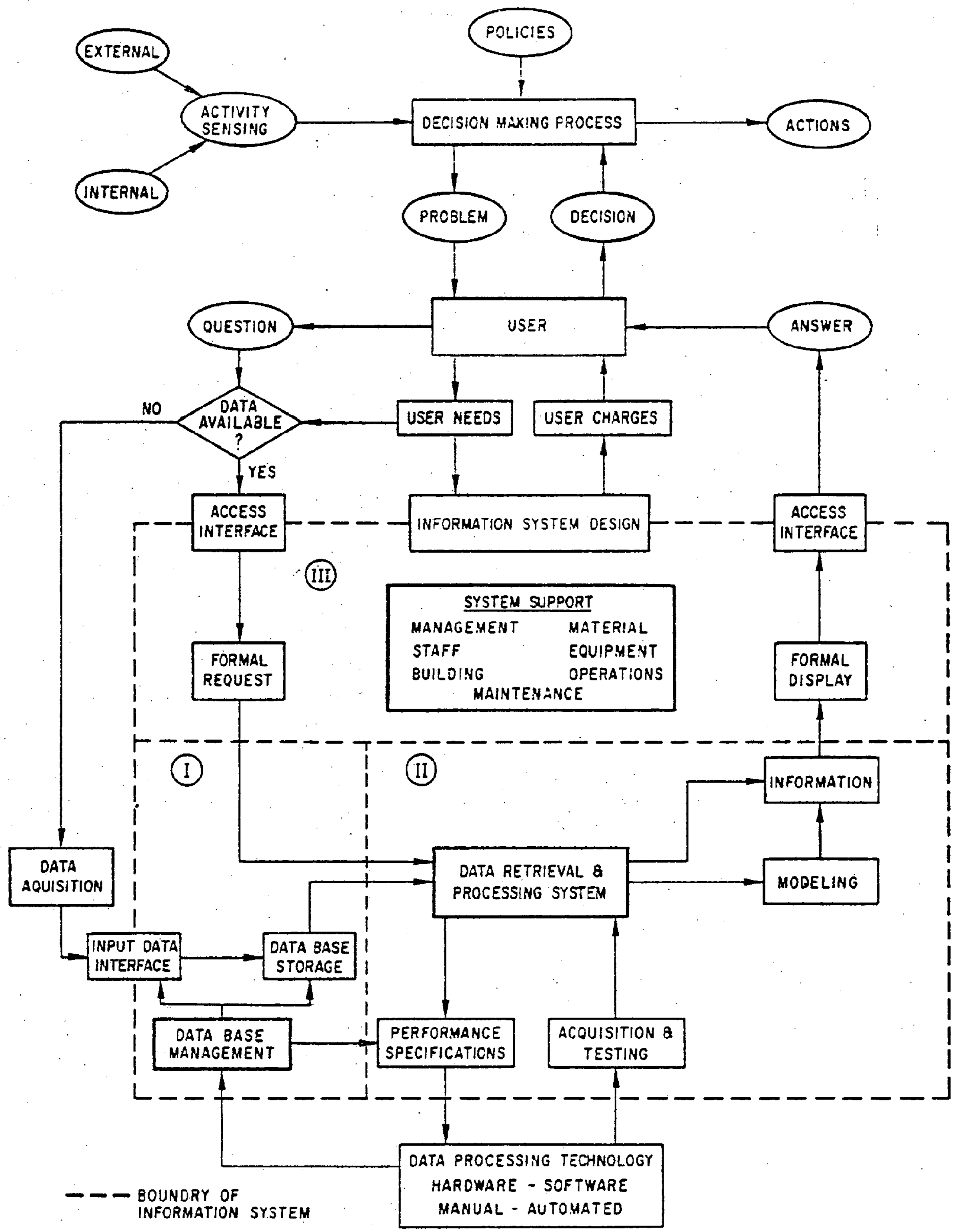


Fig. 1.1 Basic Elements of a Geographic Information System

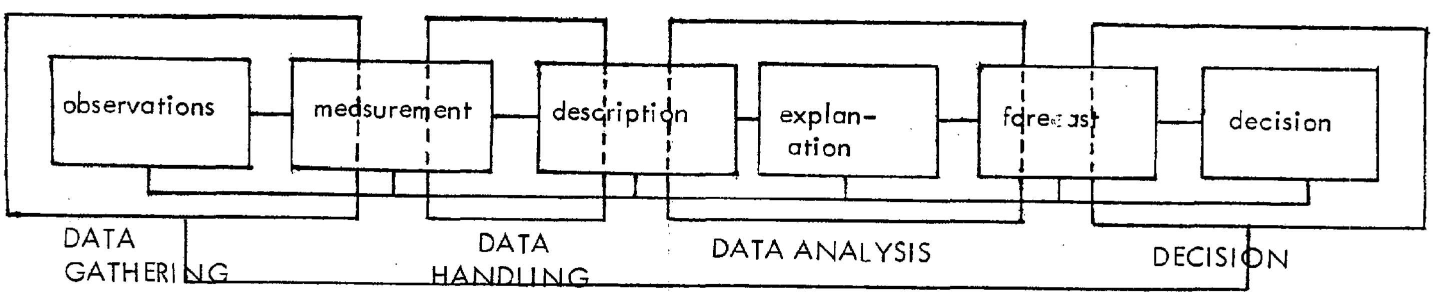
source: Tschanz and Kennedy, 1975.

Implicit in Figure 1.2 is the notion that data becomes information only after it has been retrieved and processed for particular purposes. The boundary condition which separates geographic information systems from other types is the requirement that the data be referenced in a manner which will allow retrieval, analysis, and display based on spatial criteria (i.e., map display of soils information).

Data description of various natural resource and socio-cultural aspects of the landscape generally have a describable spatial component. That is, data describing objects, entities or conditions incorporate identifiers that prescribe where they are. These data types can be directly related to places. They are linked to the places by locational identifiers such as coordinates of latitude and longitude, arbitrary grid cells, administrative areas, or merely by the position of one factor relative to all its adjacencies. When a set of techniques for handling data includes procedures for handling locational identifiers in concert with descriptive information, then such techniques are referred to as "spatial data handling systems." The term "spatial data" is commonly used as a synonym for "geographical data" and such techniques are referred to as "geographic data handling techniques."

As previously reviewed, the concept of geographic or spatial systems is much broader than the concept of geographic data handling. The spatial information system must consist of the functions which include the observation and collection of data through data analysis to theur use in some decision making process. The range of spatial information system functions are shown below.

Figure 1.3 Spatial Information Function Boundaries



To be of any utility, a geographic information system must be part of an institution. Its function is to gather data and process them for some use of the institution. Data handling is a limited, though essential, part of the process. If the data being handled by the system are spatial data as defined before, and the techniques being utilized are spatial data handling techniques, then the system is referred to as a "geographic information system."

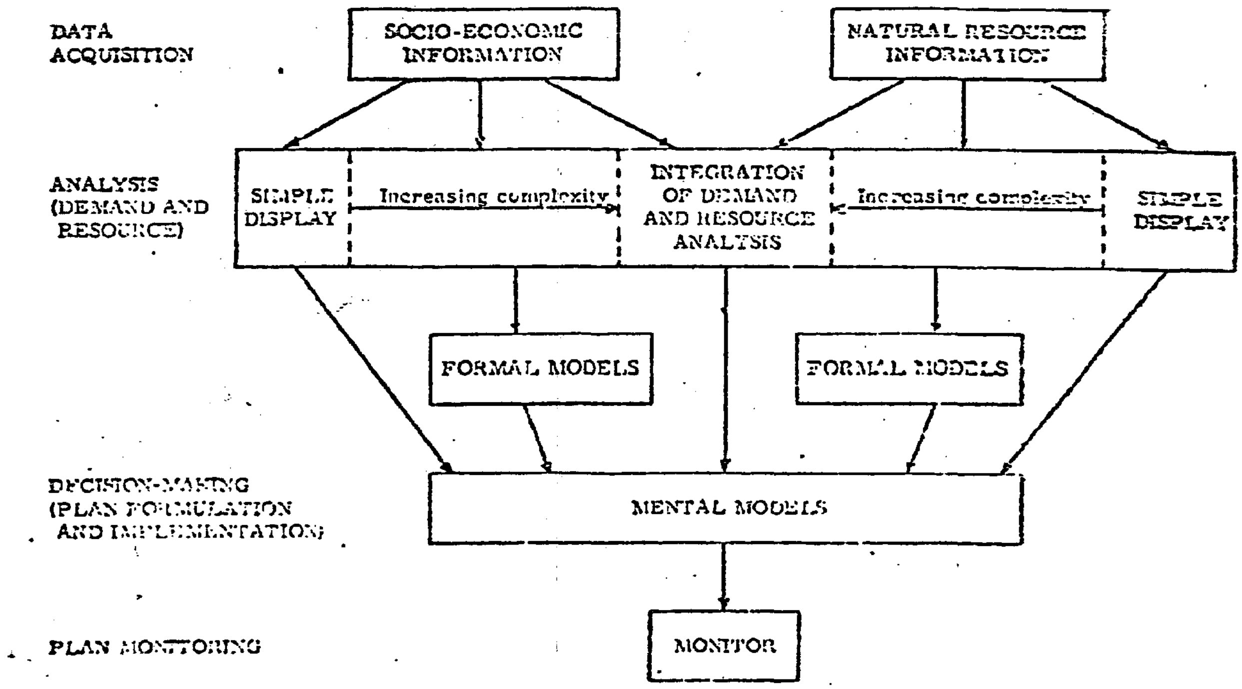
3.1.2. Some functions of geographic information systems in planning

Generally, natural resource planning responds to both demand projections and the relative abundance or scarcity of natural resources. This enables an assessment of the relationship between demand satisfaction and actual resource utilization. This definition assumes that all physical characteristics associated with a given landscape (land, water, and plant and animal ecosystems) are potentially within the domain of natural resource planning.

New concepts aimed at defining and evaluating landscape characteristics and land use activities more rigorously than in the past are now required in resource planning. Historically, land classification systems have combined concepts of land use activity, intensity, and several other factors including zoning, ownership and natural features. In the future, it appears that land use and landscape unit classification schemes defined specifically by activity, intensity and any number of hydrologic and ecologic parameters related to descriptions of impacts and not only appearance will be required. It will be critical to refine classification schemes to depict factors clearly and unambiguously as the focus of planning shifts from growth to a focus which better balances growth and environmental conservation.

Generally, natural resource planning is best designed for implementation at the large regional (i.e., river basin), state or multi-state regional scale. Implicit in such natural resource planning are six basic activities: 1) inventory; 2) estimation of resource demand; 3) resource analysis; 4) plan formulation; 5) plan implementation; and 6) plan monitoring. (Figure 1.4).

Fig. 1.4 Resource Planning and Managment



Source: U.S.G.S. 1975

1) Inventories cover the data required for estimating the demands and the data on natural resources which are identified to be of primary importance in the description of a particular region.

2) Inventoried demographics and economic conditions permits the assessment of internal and external demands forces which can be analyzed. This step also involves projecting various demographic and economic parameters for the given region. Alternatives range from no and slow growth to accellerated growth can be examined and detailed.

3) Resource assessment is an estimation of capability, suitability, and carrying capacity of the landscape and its various resources. The carrying capacity of a region is reflective of the maximum resource demand that can be accomodated within acceptable limits. Such limits cannot be objectively defined but are derived from the goals and objectives of the population of the region.

4) Plan formulation involves balancing resource demand with resource availability within broad constraints set by the broad regional goals. The process of plan formulation is ideally designed to produce sets of alternatives or scenarios which are reviewed by appropriate decision makers and are expressed as either physical plans or policy plans against which specific proposals are evaluated.

5) Evaluation of alternatives proceeds from alternative formulation and the selection of an alternative based on predetermined criteria is the required output.

6) The final stage is plan monitoring. This process often involves new data gathering, measurement and comparison of implementation achievements with previously stated goals and objectives

A complete planning process deals with the whole of a region and total demands on resources. As well, specific functional planning is often required for dealing with issues such as facilities siting or resources which are of unique significance for the region. Such activities fall into two main categories: activities motivated by conservation or preservation (i.e., critical areas); or ones motivated by the development or provision of special services (i.e., power plant siting). An additional data load results from these specific planning activities because of the necessity for greater precision and resolution in data being utilized to accurately depict selected geographic areas (Fig. 1.5).

This general resource planning process is outlined only to serve as a reminder of the problems to which a geographic information and data handling system will be expected to respond. The amount of data implied by this process is so extensive that systematic techniques (manual and computerized) must be explored and implemented for data acquisition, storage, retrieval and utility. Generally, all alternatives to purely manual techniques involve the use of computers.

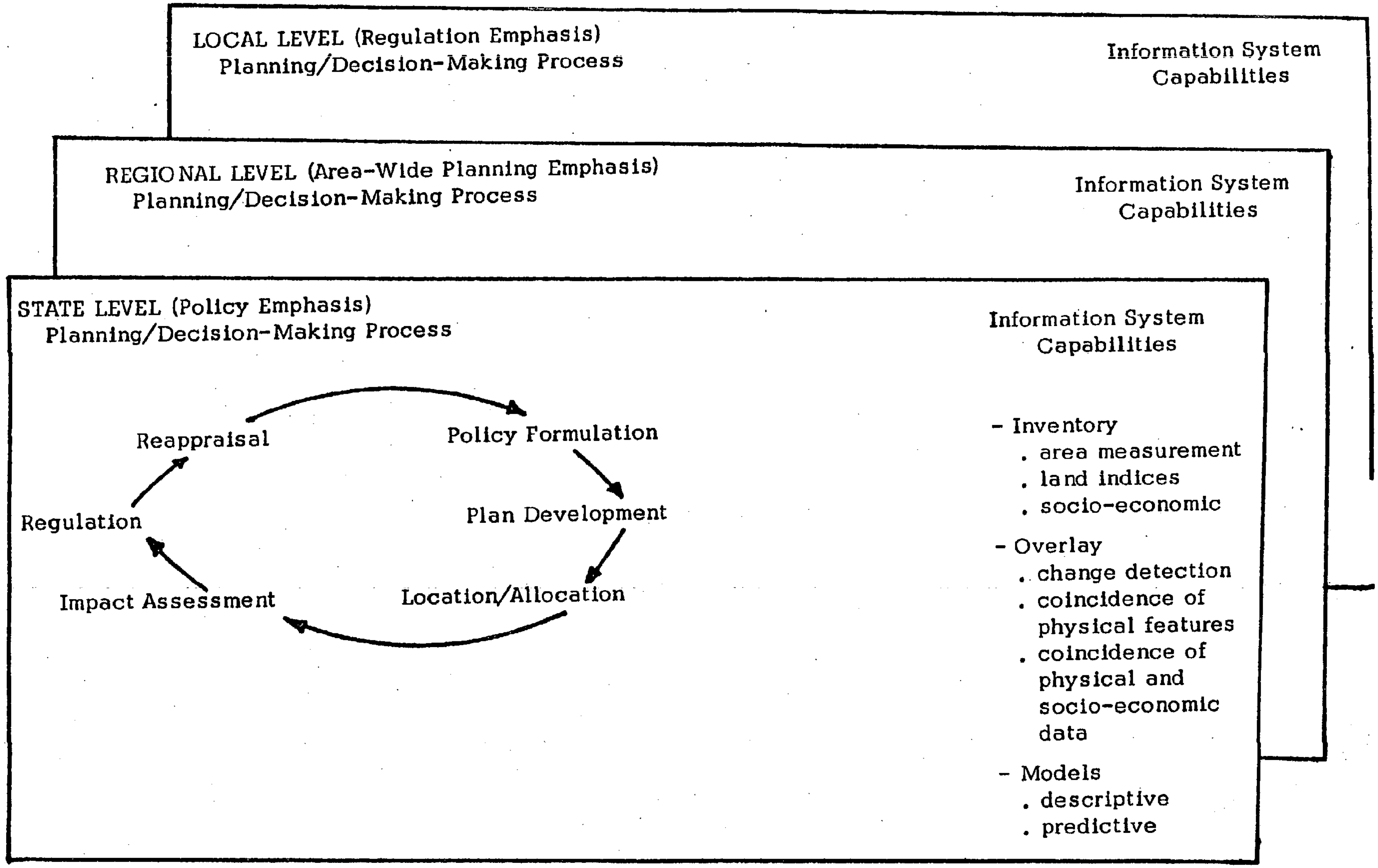


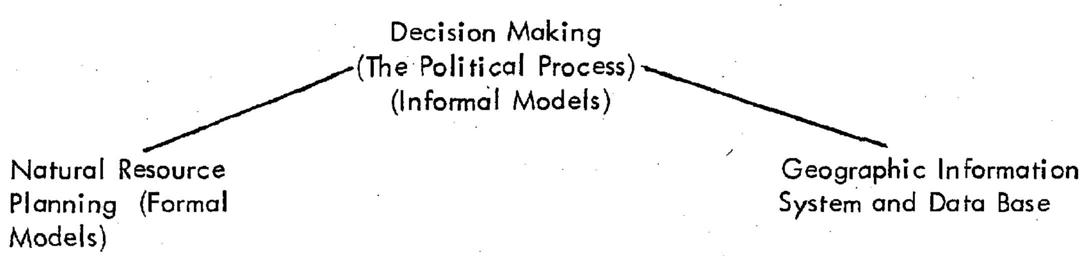
Fig. 1.5 Hierarchy of Resource Planning Processes Relative to Information Systems

source: Duecker 1975

As well, geographic information systems present powerful tools for dealing with issues associated with environmental impact assessment. State wide systems have been used to generate data about facility sites, service areas and zones of impact. (See Table 1.1) The use of an information system for impact analysis can help to determine the achievements of objectives of efficiency and equity for the assessment by providing easily accessible data to measure external effects, benefits and costs, both for general regions and for specific areas and population groups.

The expected use and utility of data types determines how data must be collected and handled (Figure 1.6). Formal data analysis models (i.e., suitability for residential development mapping) are generally the required end use of geographic resource information. However, even when formal models are the analysis method, the data still must serve as imputs into what are called mental or informal models (the politically based decision making process). Data handling techniques must respond to the requirements of both formal models (accurate, reliable, timely, etc.) and informal models (understandable, clear, concise data outputs).

Figure 1.7 Information Use



In this context, information systems must be multi-faceted, particularly if faced with responding to different appropriate levels of spatial and temporal scale. The information system must respond to the relationships between levels of government, and the nature of functions such as area-wide planning, impact assessment and environmental regulations.

IMPACT TYPE	LAND USE CHARACTERISTICS	SOCIO-ECONOMIC DATA
Spatial Incidence		
- Facility Site	Amount of land taken, by type, for facility	Number of persons, homes, businesses, etc., taken by facility
- Service Area	Amount of land, by type, in service area	Number of persons, households, in service area
- Impact Zone(s)	Amount of land, by type, in zones impacted by facility	Number of persons, households, in zones impacted by facility
Achievement of Objectives		
- Allocational Efficiency	Measurement of externalities	Aggregate benefits and costs
- Distributional Equity	Measurement of external effects to specific properties	Distribution of benefits and costs to population groups

Table 1.1

Use of Information System for Impact Assessment

(source:Duecker 1975)

Information System Elements	Reappraisal	Policy Formulation	Plan Development	Location/ Allocation (Site-Selection)	Impact Assessment	Regulation
		PLANNING/DECISION-MAKING PROCESS				
Capabilities						
- Inventory	Change detection	Indicators of achieving objectives	Indicators of implementation	Population, public facilities, travel time	Development applications	Building permits
- Overlay	Change from $landuse_i$ to $landuse_j$			Coincidence of land features	Coincidence of land features	
- Models	Land indices	Allocation efficiency, distribution equity	Allocation, efficiency, distribution equity	Location/allocation algorithms, efficiency, equity	Impact diffusion models, efficiency equity	
Spatial Scale						
- Analysis	Municipality County, Region	State, region, county, municipality	State, region	Distance between points of supply & demand	Service area, impact zone	Ownership panel
- Basic Unit . Land Use	Natural unit (cell, polygon)	County	Natural units	Eligible sites	Natural unit	Ownership panel
. Socio-Econ - rural	Enumeration District (ED)	ED or County	ED	ED, MCD, or County	ED	
- urban	Census Block	MCD	ED or Tract	Block, ED, Tract	ED Block	

Figure 1.6
Capability and Spatial Scale Requirements for Information System to Support a Statewide Planning/Decision Making Process (source:Duecker 1974)

3 .1.3. Examples of Geographic Information System Outputs

Output products are usually defined as documents (either data listings or maps) which can be used directly by the planning profession and/or the decision maker. This infers that no additional manipulation or analysis may be required before the information is utilized. For example, resource planning at the regional or state level is often concerned with the amount and distribution of various land uses. This type of information can be obtained from an information system in a variety of ways. A map is usually prepared from aerial photographs which outlines the boundaries of each use for which identification was required. The map contains both the boundaries of each use and a code which identifies the use. From mapped parcels, areas occupied before each use can be estimated. If more accurate information is required, areas can be measured by dot grid or planimeter. Such manual techniques often require several interations to ensure an acceptable level of accuracy. Another alternative is to code the mapped data for computer processing and direct the computer to measure each area (see figures 1.8. 1.9, and 1.10).

It is also often desireable to note changes in land use over time. The same basic techniques are available to note changes in land use as to record present land use. One can obtain the photographs, hand draw and then measure changes or such changes can be measured with the computer. Often the most useful products from the latter are tables showing changes which have occurred and specific maps depicting various kinds of land use change.

Maps such as agricultural productivity and land surface cover, if overlaid, can produce comparisons of immediate interest to resource planners. Questions concerning land conversion, amounts of remaining agricultural land, the relationship of productivity to actual use can be easily answered with graphic and tabular analysis procedures. As well, other different map types can be combined manually or by computer to derive composite map results. For example, maps of salt water and topography can be overlaid to produce a map of non-fresh water coastal wetlands. Maps of physical capability, topoclimate and scenic resources can be overlaid to delineate areas most suitable

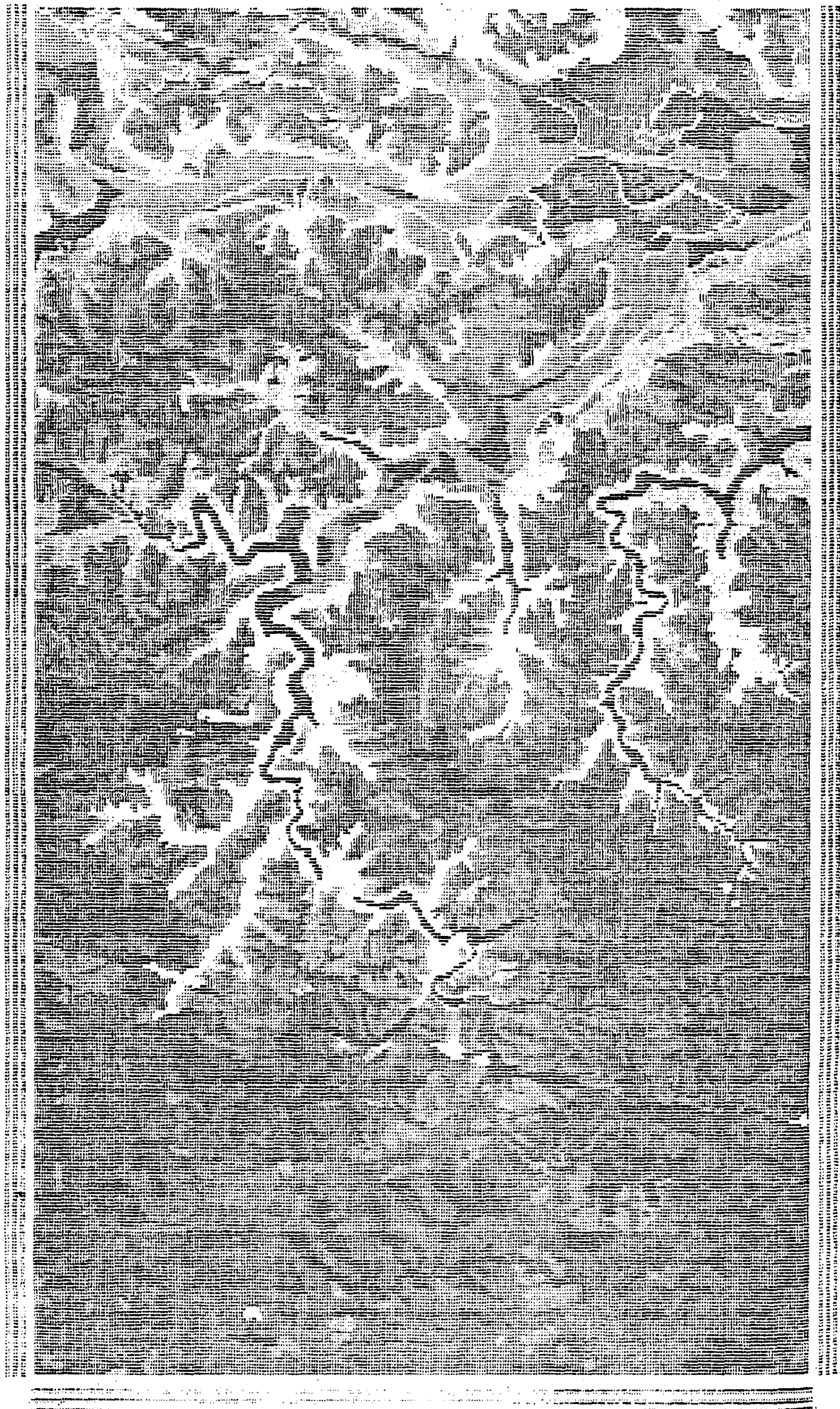


Figure I.8 Example Geographic Information System Mapping
(soil suitability for agriculture in Iowa)
source: Duecker 1975

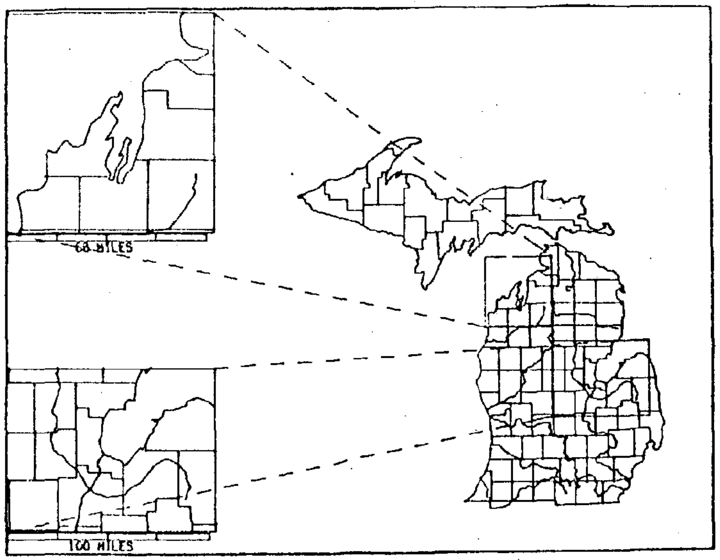


Figure 1.--Base Map

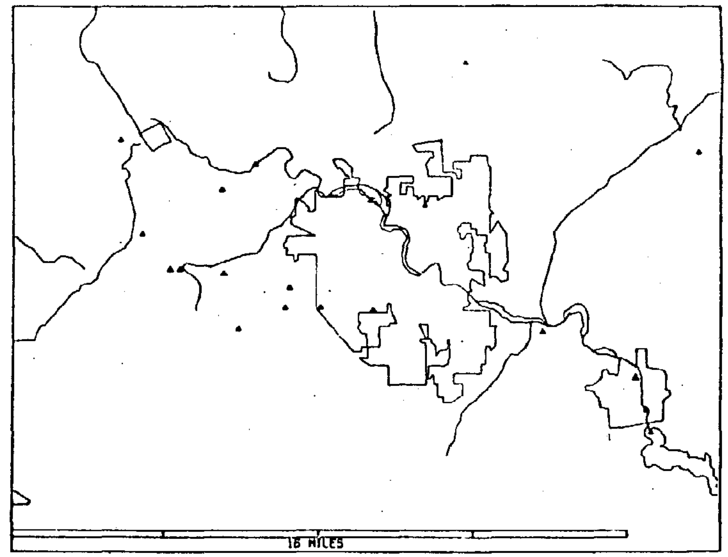


Figure 2.--Magnified portion of Huron River Basin

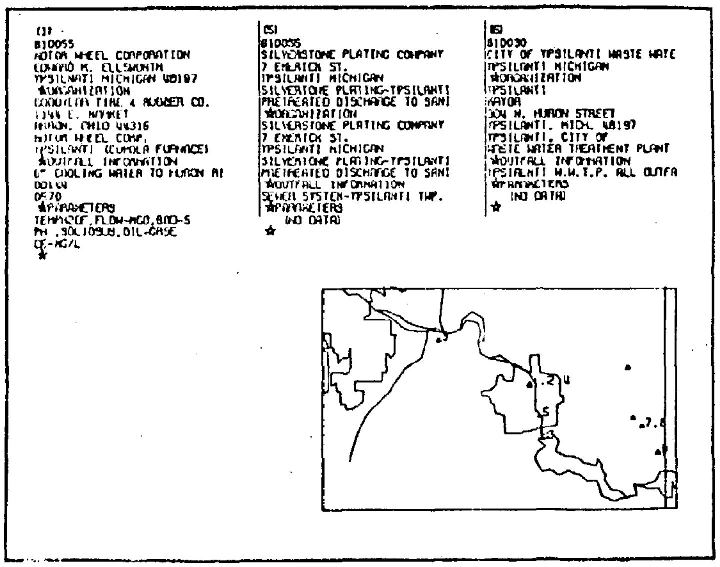


Figure 3.--Typical Station Data

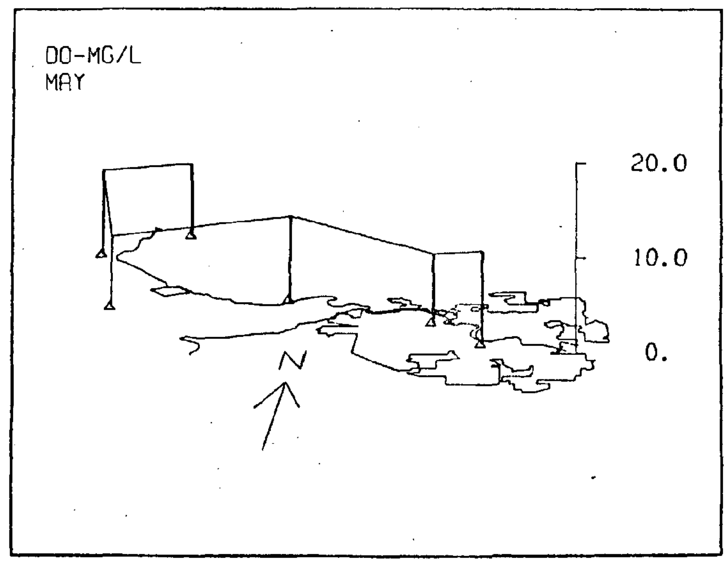


Figure 4.--Water Quality Degradation

WATER RESOURCE DATA

Figure I.9 Example of Geographic Information System Output

(source IGU 1972)

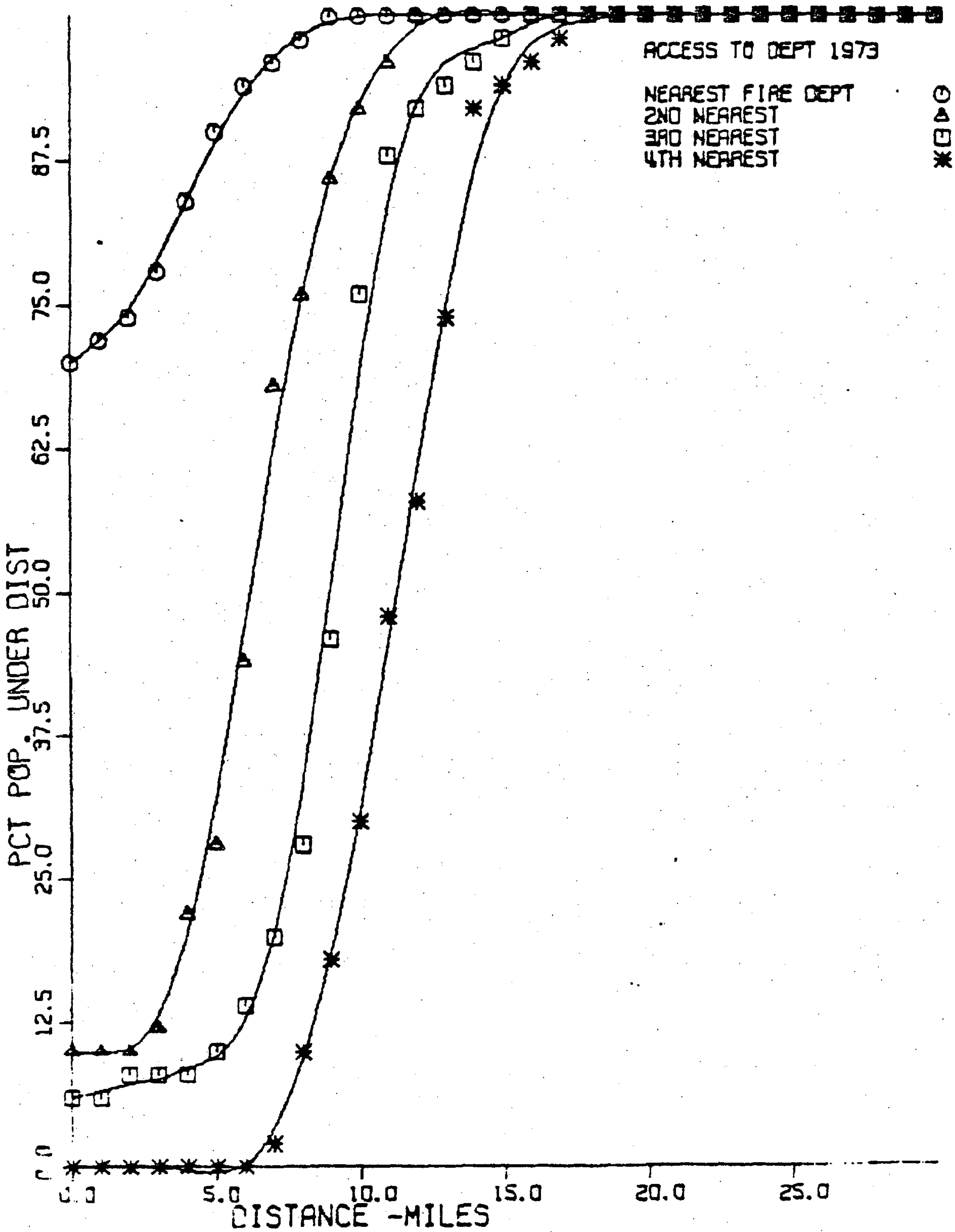


Figure 1.10 Example of a Geographic Information System Output (distance to police stations)

(source: IGU 1972)

for new housing development.

Essentially, the typical kinds of questions faced by resource and land planners are concerned with the distribution, quantity, quality, and change in the characteristics of natural resources. It is of geographically referenced natural resource information which deals best with understanding these resource concerns with the provision of tabular and mapped information required for professional planners and decision makers alike.

3'.1.4. Spatial Information System Options

An agency faced with the task of building a data and information base for planning has the problem of creating inventories which describe the state of all relevant data elements. There are three basic approaches: (1) the required data elements can be identified and programs developed to collect these data in a set of inventories; or (2) data collected from other sources, usually for specific other purposes, may be acquired and combined in various ways to be useful in resource planning; and (3) a combination of (1) and (2).

Each of these approaches have significant advantages and disadvantages. In the first approach, the design and process of data collection operations are time consuming and very costly. The major advantage is the ability to control the definition of relevant data elements, and the manner in which they are collected. In the second approach, the time and costs for data collection are not serious problems; however, the responsible party must design data handling procedures which can accomodate varying record formats, recording media, and data columes represented in the totality of the data identified as relevant. It is possible that the savings in cost at the acquisition stage may be offset by subsequent costs of data manipulation for varying formats and record volumes.

It must be noted, unfortunately, that data quality is not at the center of most agencies concerns as they approach planning. Data seem to be assessed only on the basis of :How precisely does the data item tell me what I need to know? Can the data be used directly or must it be manipulated first?

Tomlinson (1971) notes the significance of data development through a comparison of information system component development in terms of the current state of the art for each component (see Figure 1.11).

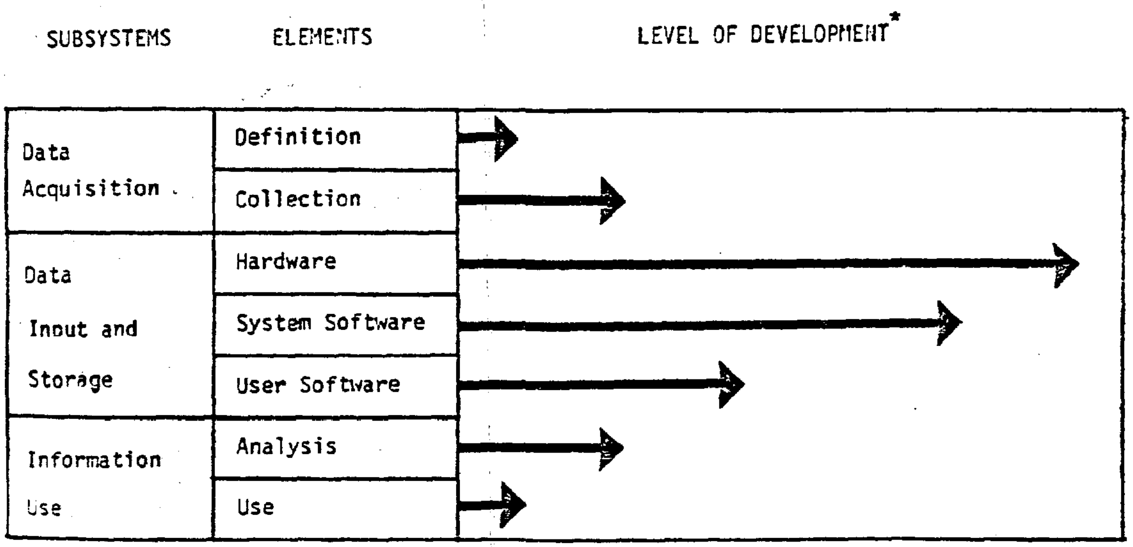
Hardware is by far the most advanced, followed by system software. Data documentation and collection are the least advanced and most difficult part of any information system effort.

There are four basic options available for performing the range of technical operations associated with geographic information systems: manual; computer assisted: interactive computer assisted; and automated. No known systems are currently completely automated, so it is recognized that complete automation is questionable as an alternative.

Data handling systems must maintain a high degree of flexibility in order to allow easy adaptions to changes in preceived needs by system users. The methods used to maintain flexibility in the various technical operations are the cornerstone of information system development. Flexibility is best mainatined by planning for continued system openess (i.e. avoidance of needless automation). For example, if information system functions (i.e., storage of data) can be accomplished efficiently using only manual techniques then a manual system should evolve. In other situations, manual data handling methods can be combined with computer oriented techniques relegating the computer only to those tasks best accomplished by the machine (i.e., repetitive statistical analysis or map overlaying). The technical options have implications for methods of information transfer, storage, retrieval, analysis and display.

For manual systems, operations include interpreting photos, drawing maps and preparing output reports, analysis and graphic presentations. Computer associated procedures include encoding the data base into a computer readible form which often requires the addition of specific equipment such as digitizers and interactive editing devices and storage devices such as cards, tapes, or discs.

Figure 1.11
Information System Development



*Relative order of magnitude (expressed as proportion development)

FIGURE 1.11 is presented to illustrate the relative developmental stages of the important aspects of geographic information systems.

For manual system options, data retrieval consists of extracting a required map, photo or statistical table from a larger data file based on some criteria for extracting the specific observations. With computer assisted options, retrieval operations are accomplished by software which are utilized for the selection of desired data from the files located in various computer storage devices. The interactive computer aided option includes built in feedback mechanisms that allow retrieved items to be visually displayed for examination to determine if they are to be included in a special data file being created for the intended analysis. Manipulations are carried out by additional software.

Outputs, such as maps and tabular listings can be also produced manually or with computer tied peripheral equipment (line printer, plotter, television screen). Outputs from computer aided options sometimes require manual modifications for use by decision makers but should not require modification for use by professional planners.

3.2 The Structure of Geographic Information Systems

Geographic information systems are not only composed of data and data handling techniques, but are a composite of these and a number of other related primary components. The components recognized as being central to geographic information systems are: (1) data acquisition; (2) software and hardware; (3) the technical operating procedures of the system; (4) the system's relation to its potential users; (5) the management of the system; and (6) system component procurement requirements.

3.2.1. Data Acquisition

The data acquisition component of geographic information systems must deal with all the system's data, their characteristics and the institutions and systems which are vital in data supply or in assisting in their acquisition. The data acquisition component of geographic information systems must address the issues of:

A. Data Characteristics:

- data formatting, record formats, and data volumes
- location and ownership of data
- documentation for data
- classification of data
- data collection plan

B. Data Framework Parameters

- procedures for acquisition
- continuity in data flow
- informal data provision agreements
- data supplier – data user program coordinator
- changes as required in data collection

C. Evaluation of Data

- applicibility to the problems of users
- assessments of accuracy and reliability

3.2.1.1. Data Characteristics

Within systems a number of data characteristics affect their subsequent use and ease with which they are processed. Within the framework of an information system, some important characteristics are:

accuracy, precision, reliability;

coverage, scale, relevance;

format, volume, density.

A. Accuracy can be assessed by determining the uniformity of data over areas, by measuring uniformity of recording medium (maps); and by estimating spatial precision and timeliness.

B. Precision of spatial data is affected by the choice of medium and data format, by the geographic referencing system, by the nature of assigned descriptors. Every change in format or scale introduces possible errors and generally a loss of information. Also, precision may be reduced if information is interpolated from data obtained from scattered points, such as with ambient water quality data.

C. Reliability is a measure of the quality of data as it relates to a problem and to the assumptions and requirements of specific data analyses. The ultimate test for reliability is simply how the data set corresponds to reality - both spatially and descriptively.

D. Coverage relates to the geographical - spatial extent and completeness of the data and the content of the coverage (does it have the desired classification scheme).

E. Scale refers to the size of a recording medium for spatial data in relation to real world dimensions. The importance of scale is recognized as significant when data may be recorded at one scale and archived or analyzed at another, often resulting in the loss or distortion of information.

F. Relevance refers to the appropriateness of data to performing a task or answering a question. To be relevant, data must be able to meet specific analytical requirements and meet standards of accuracy, scale, and timeliness.

Data characteristics (A) through (F) are characteristics which affect data use. Do they tell the user what he wants to know, make inferences he needs to make to perform essential manipulations and analyses?

The following data characteristics are those which affect data handling. That is, these are characteristics with implications for data processing (manual or automated) and they relate to the formats, volumes, and density of data. These characteristics have significant implications on information system design and must be recognized in the initial stages of system design.

G. Format refers to both the format of data records and individual data elements. Record format relates to the arrangement of data for capture and storage (maps, cards, magnetic tape, tabular, etc.), and graphic symbols and other delineations used (i.e., legends for maps). Written records, graphic records and machine readable records are the three general types. All three are generally a part of information systems and are useful for various specific data types. Record formats are the observable results of a data collection effort. They define input for geographic information systems and are therefor a prime concern in all such systems.

Data format refers to the mode in which real world objects are represented as data. The format is both the representation of a real world element and a locational identifier marking its position, area, and extent. Data format is important for software development as it affects both data manipulation and program structure. In terms of location data may be point, line or area data. Standard maps, such as USGS quads, often contain all three data types, but information systems generally require separate inputting of lines, areas, and point information. Where both locational and descriptive data are to be utilized, the two are separated for computer inputting and then merged internally via special software.

H. Volume again refers to both records and individual data elements. Record volume is the number of separate physical records on which data is recorded, such as the number of maps in a specific mapping series. Record volume provides an important delineator of data handling requirements as it is usually a specific number where as the amount of data per record usually varies greatly. Record volume is important in both hardware and software development because record volume can overwhelm a data handling system just by the sheer number of cards to be punched or maps to be digitized. Records volume is important in frequency determinations for handling and in retrieval. A system must know the peak load of requests to access records that it can handle.

Data volume is an extremely difficult parameter to estimate. Counts may be conducted on sample areas, but actual volumes are generally not known until actual data inputting. Data volume may be only estimated, but it is important to estimate its likely growth rate for continued system data handling success.

I. Data density is an important contributor to data handling operations. It serves to convey the degree of difficulty to be expected in processing data from map sources. Data density is often expressed as the number of data elements per square mile at any map scale on any map. Generally, technical difficulty and the costs of data input decrease significantly as data density decreases, so it is uneconomic for a system to accept data more dense than required by system users. Data density is perhaps the most significant parameter in determining the magnitude of imputs, storage, and other data handling tasks of a geographic system.

3.2.1.2. Data Framework

The data framework refers to the physical existence of data and the ability to obtain and use those data. Procedures for acquisition relates to data ownership. For example, many data are subject to provisions of confidentiality which may impact their use or collection. Also, their are tremendous differences in both the physical location and accessibility of what are often similar types of information. Continuity of data is significant if a geographic system relies on specific temporal or geographic coverages or extensions of coverages.

The important bottom line parameter for data which is available relates to the process of extracting a relevant, reliable data base from the available sources. For example, data series such as soils and surficial geology often require additional interpretation at the time of input into the system.

When available, most parameters relevant to the framework for data acquisition relate to cost. The full costs of making data available and usable must be weighed against the costs of creating new data. When budget constraints permit,actual data collection permits the use of data whose relevance and reliability can be ensured.

3.2.1.3. Data Evaluation

Evaluating data relates to the applicibility of the data by examining documentation for accuracy and reliability. This step is important because any attempt to use other than error free data files can be disasterous in terms of data analysis, outputs and end use, and will jeopardize confidence in the entire system.

> "The quality of data is important. One bad piece of data can ruin a good data set. The ability of data to stand as evidence and be acceptable to both sides is the essential criteria for the construction of an information system." (Bossleman, 1975)

The level of effort which can be devoted to data inspection usually reflects staffing, budget and time constraints. The advantage of secondary data is that often they have already been inspected and interpreted for some use, but, as previously discussed, since this previous use usually will not correspond with the new use, additional efforts aimed at data documentation for evaluation are generally required.

3 .2.1.4. Data Most Often Included in Geographic Information Systems

The emerging problem solving capabilities of state initiated and statewide resource planning and management associated information systems calls for data typical of more localized sources concerned with land use and natural resources. This data has generally been concerned with planning processes related to aesthetic, ecological, geological, hydrological, physical and socioeconomic values, conditions and trends, along with any number of factors that determine the suitability of land for various uses.

Data most often included are: soils, geology, vegetation, land use activity, hydrology, flood hazards, and natural features and to a lesser extent, cadastral parcels, streets, roads, and highways, utilities, zoning, census tracts, plans, redevelopment districts, and specialized factors such as moise levels, seismic hazards, and ambient air and water quality data.

3 .2.2. Software and Hardware

Software and hardware are computer oriented terms which refer to procedures for manipulating data and the equipment that performs tha actual manipulations. An addition to standard software and hardware, geographic systems require specialized procedures and equipment to handle spatially referenced data.

3 .2.2.1. Software (System Procedures)

With most systems, there are generally manual procedures (instructions for individuals to follow, such as in interpreting aerial photographs) and computer aided procedures (software for data base management and information analysis and display). Software will not be reviewed here but specific discussions can be found in Appendix A and Chapter 5 of this report in an IGU (1976) report to the Geologic Survey on computer[1] system software related to geographic information.

3.2.2.2. Hardware[1]

Systems generally require both manual devices (i.e., optical devices) and computer associated devices. The hardware for a system must be capable of handling both cartographic and attribute data, allowing for its entry, retrieval, examination, and often modifications. In this brief section, the concern is with introducing special hardware requirements as an introductory exposure by the reader to automated central processing units, computer memory alternatives and standard input-output devices is assumed. It is recognized that many geographic information systems in partial use or development, do not have any special hardware in that they merely use terminal displays on large time-sharing computers. But many new systems are cost effectively using certain types of special display and edit, drafting, and digitizing equipment and these are the focus of this discussion.

The Digitizing Sub-system. Digitizing, the process of recording the areal dimensions of spatial data in computer readable form, can be done manually and automatically. As well, both grid cell and polygon systems can be digitized manually or using automated techniques.

Grid cell digitizing can merely incolve the placing of a gridded overlay over a map and reading along each of the squares, allocate values for the parameters being considered to each square. However, a similar output can be gained automatically if the map is scanned by optical or vaster scanning devices.

In digitizational of polygon data line digitization is the most common method employed. For map work, free pencil digitizers are most effectively used, in which the scribing device

---

[1] This discussion is based heavily on IGU Appendix C and Geographic Data Software, IGU, 1976 800+ pages.

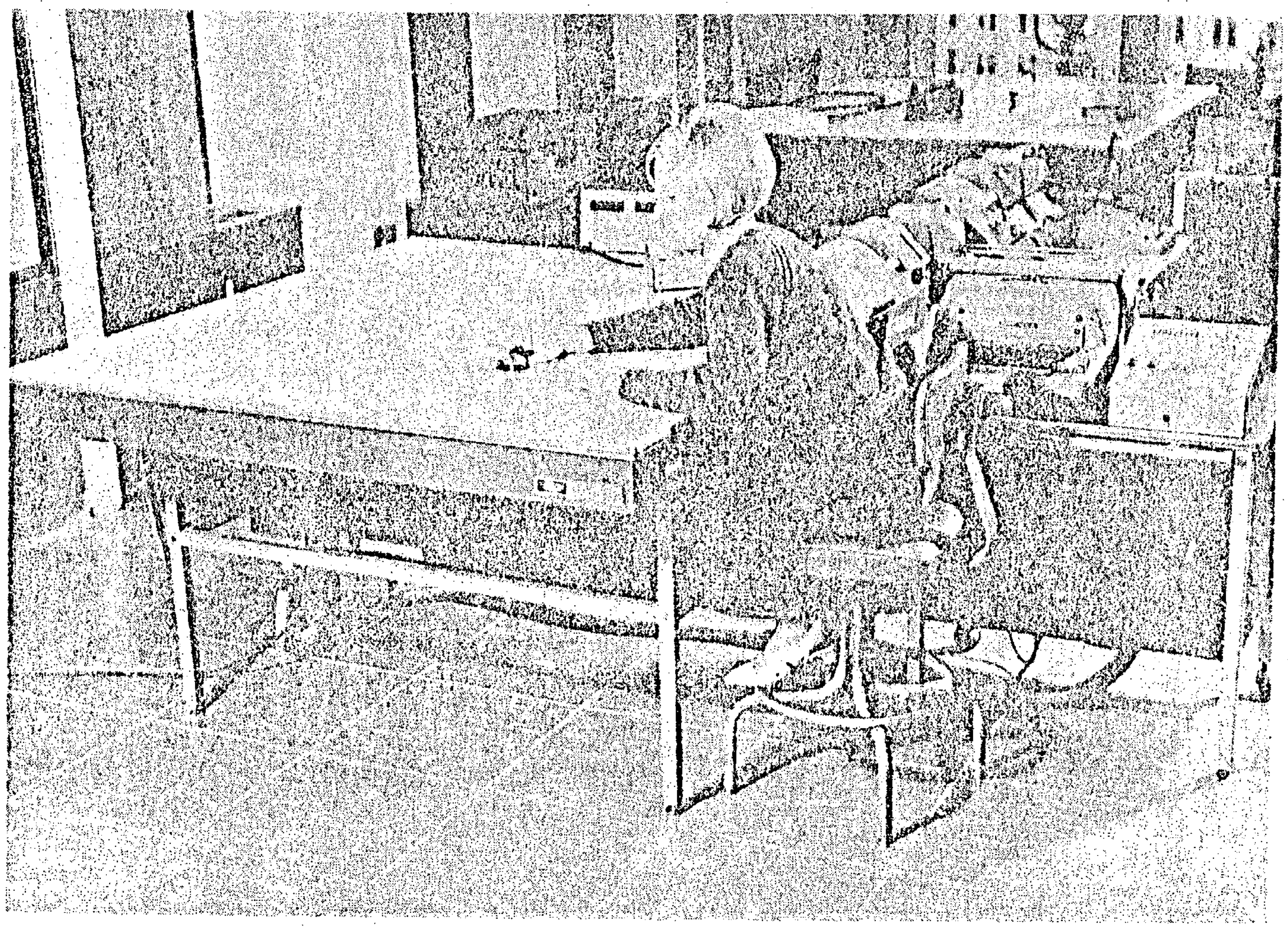


Figure I.12 Example of Digitizer Operation (electric pen which records point coordinates along lines) is not mounted on an arm but can be moved and placed anywhere, the only restraint being the electrical connection. The alternative is to mount the pen on a trolley on the arm of a drafting table (see Figure 1.12).

One of the problems in off-line or non-interactive digitizing is there is no way to check the data as it is being recorded and there is no way to efficiently correct mistakes. The addition of a minicomputer controller to carry out the checking, and the reporting and erasure of unwanted points adds greatly to confidence in the data and the reliability of the entire operation. This cost is only slightly greater than systems with no minicomputer control.

Automatic line followers and drum scanner digitizers are also available but are not widely utilized. Automatic line followers contain a viewing photocell which is carried by an x, y incremental motor; between the photocell and the field of view is a segmented disc. The relationship of the disc to black lines on a document serves as the basis for x, y, motor instructions for line following. A number of hardware (problems in automatic accuracy checking) and software (difficulty in locating islands) have prevented the widespread use of line followers in cartography. Raster scanners are more common and these involve a line by line sweep across a display surface to generate the map record. One basic problem with scan digitization is the difficulty in converting from raster to vector formats within real time limits. Also there are few user programs which allow this to be done economically and efficiently. The difference lies in the fact that raster formats involve holding the data as horizontal lines and vector formats involve actual line by line encoding and storage. So most scanning devices output directly onto magnetic tape and the conversion is carried out off line. Because of the great bulk of data often encountered, this conversion is a major data handling issue to be addressed. However, the Canadian Geographic Information System has been operating efficiently in this manner for a number of years.

The edit subsystem often consists of one or more displays associated by a minicomputer and controlling software. Such self standing edit units currently cost between $25,000 and $100,000. The purpose of the edit subsystem is generally verification of data which has been either digitizing or scanned in terms of completeness, accuracy and cleanliness. To properly accomplish this a first requirement for hardware should be a self-standing interactive display system. A second requirement is a manual digitizing capability to carry out modifications of cartographic data.

The choices of display equipment for the system are numerous. The most widely used display is the raster refresh scanner which produces images by drawing horizontal lines on a television screen. Various vector type plotting systems are absolutized. Other types of storage displays include laser screens which are large television screens with very clear picture images. Such systems best operate when they allow interaction with the display for correcting, updating and modifying the displayed information so storage of some type is generally an integral part of the edit subsystem.

The drafting subsystem is required to produce output display. Most display systems involve drafting units and plotters operated for off-line cartography and many such units contain mini-computers for control. Drafting units produce cartographic lines drawing in a continuous line mode and lower quality drafting units are referred to as plotters. Flat bed plotters and drafting units look like ordinary drafting boards and drum type machine is a revolving drum with an axial motion for drawing. Generally, drum units are preferred because of ease in manufacture and maintenance and better speed response in operation.

The quality of the map which is obtained from either unit type depends on the quality of the mechanism that determines the size of steps in the x and y increment motors. This simply refers to the drawing pens which are driven vertically and horizontally by separate motors relying on these motors acting simultaneously to produce angular or curvilinear line patterns.

Microfilm plotters, electrostatic pointers, electron beam recorders and light heads are also available but few systems have currently incorporated either because of hardware expense and complexity.

This discussion is both brief and incomplete but it was only intended to point to the range and functions of important peripheral equipment types. The next section will deal with digitizing, editing and outputting functions in much greater detail.

In summary, most of the minicomputer systems currently being utilized in digitizing and editing functions have cost on the order of $100,000 for complete equipment and software. Over a five year period, 1,000 hours per year, this would mean an hourly rate of $20/hour. This can be compared to an hourly rate of $250 + for large time sharing computer systems. When it is appreciated that data handling functions and search operations are as fast and efficient on minicomputers, the economics of the smaller units is appreciated. As well, a review of existing systems seem to show that systems with smaller in-house systems have much faster turn around, which must be considered if visual interaction with large data bases is desired.

3.2.3. Technical Operations of the System

3.2.3.1. Introduction

Technical operations relate to the total set of operations which must be performed on data to achieve requisite products from the system.

Data input functions cover all operations needed to develop an error free data base ready for analysis. These operations are important because documents used to create data files are error prone and in the past, human interpretation has been the primary data check. As well, data inputting covers processes which relate to the analysis and transformation of data to bring them to a standard base. So data manipulation at this stage only provides for checking the original information and placing it into a computer ready format (see Figure 1.13).

The next group of techniques relate to the manipulation of data to change it before some analysis. This is important because the needs and requirements of many analysis techniques require that the data be transformed to a format suitable for that analysis technique. All major geographic information systems contain the capability for data manipulation prior to analysis.

The third group of procedures relate to data retrieval and analysis. These procedures encompass management of locational and attribute data, retrieval from the data base, and any procedures for data update, further correcting, reformulating or deleting.

The final group of procedures are information output procedures. These relate to methods and formats of tabular and graphic output and to validity checks and verification procedures.

3.2.3.2. Data Input Functions

The first step in data inputting is an evaluation of the quality of the source material. Data documentation should be available to make the following judgements:

1) Is the source acceptably error free?

2) Are there extraneous data?

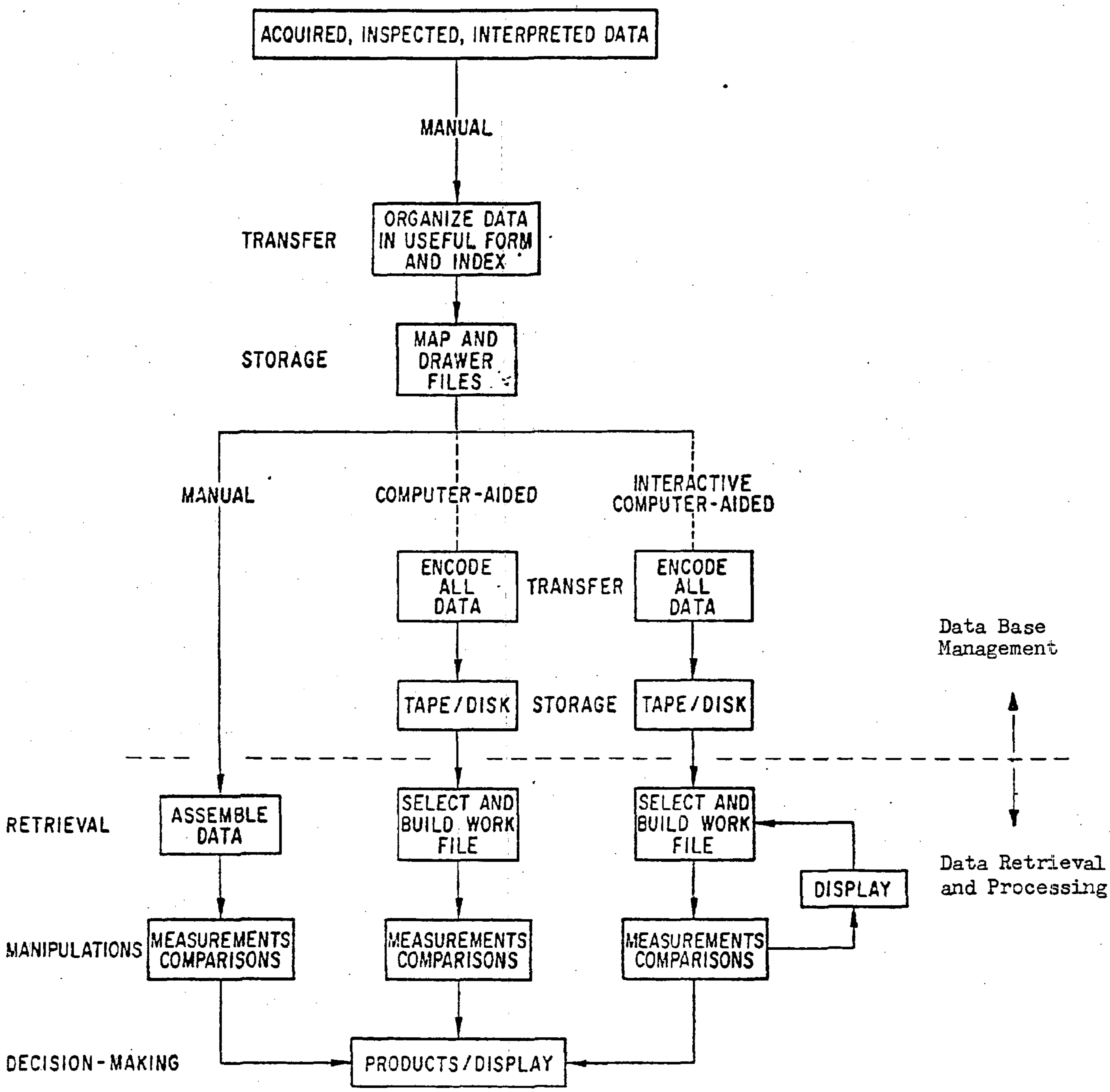


Figure 1.13 Geographic Information System Options

(source: Tschanz and Kennedy 1975)

3) Are there missing data?

If required,corrections are usually performed before actual data entry. Correction is a manual process and is likely to introduce additional errors and failure to detect them can lead to a very costly editing process after graphic to digital conversion (see Figure 1.14).

Many source documents require reformatting before conversion from graphic to digital data. The nature of the conversion process, particularly the level of automation, determines whether or not reformatting is required. An example of this is preparation of line maps representing polygons on stable base materials to be used on an optical drum scanner. If the graphic to digital conversion is likely to be error prone, the source material may be reformatted to avoid new errors. Reformatting may also be required to simply separate geographic data from attribute data.

Once checked and properly formatted, it is important to estimate data volumes and relate them to formats to be utilized. The desired relationship between real world elements and their data representation must be studied to ensure acceptable levels of information loss.

Before data entry,initial classification schemes must be developed using inputs from existing standardized schemes (i.e., standard statewide land use coding), user needs assessments, and actual characteristics of the data. Where flexible, a choice must be made concerning coding scheme, and the advantages of different schemes must be weighed as the complexity of data manipulation procedures are strongly affected by the nature of these codes.

The final step before actual data entry (digitization) is partitioning the graphic data base. Source documents must generally be partitioned or combined into optimal working units. This step involves locating boundaries between records, determining appropriate coding, and developing procedures for linking records if spatial aggregation is required.

The next step, digitizing, is the major data transfer function in setting up a geographic

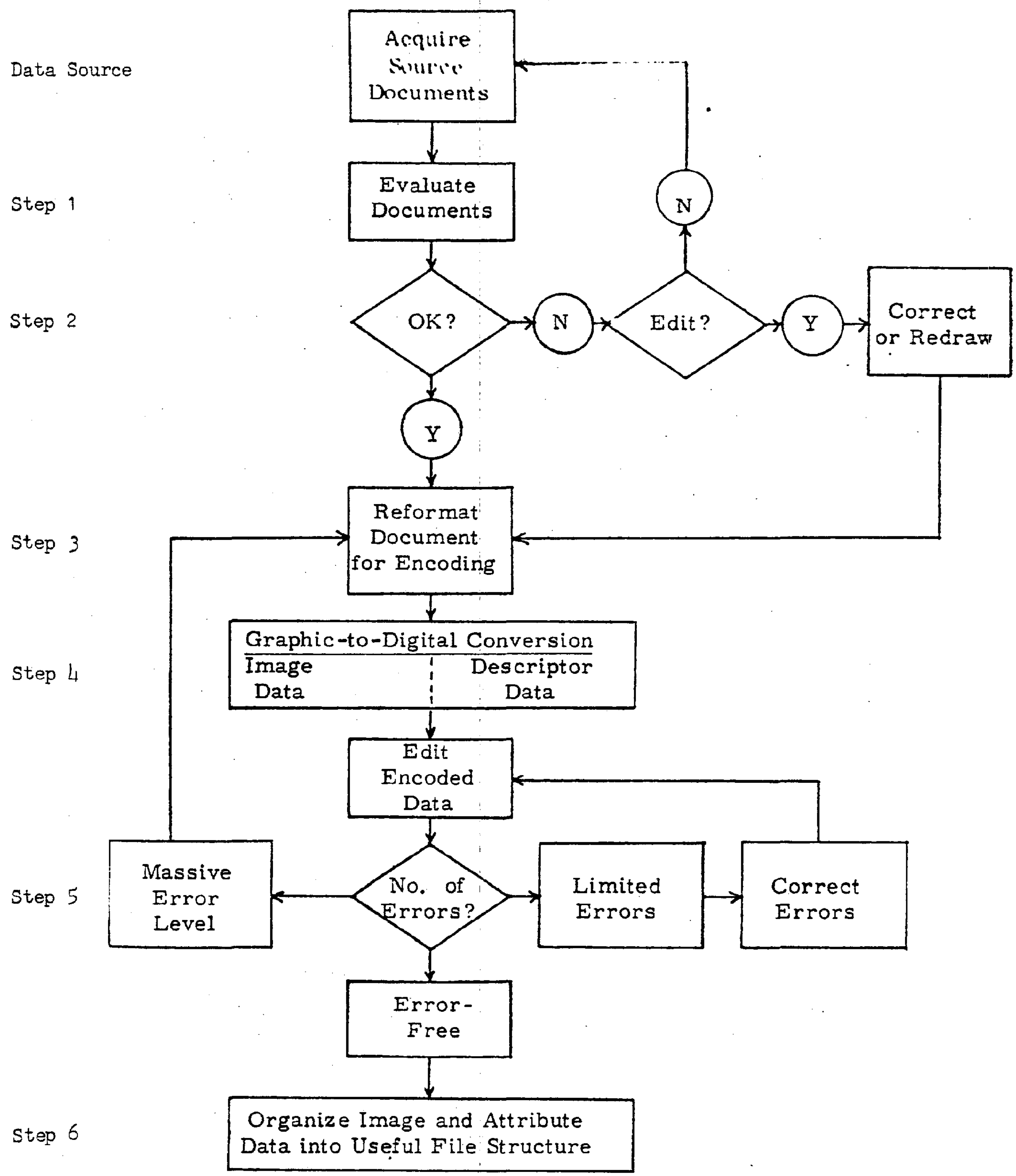


Figure 1.14 The Digitizing Process

(source: IGU 1976)

information system. Digitizing is simply a sequence of tasks for encoding spatially oriented data. In its narrowest form it can be viewed as determining x, y coordinate values to describe the location of points, lines, and areas as they are depicted on maps. In a broader sense, digitizing can be viewed as the task of creating and acceptable, machine readable, data file from a variety of sources. Figure 1.15 illustrates a generalized process whereby a useful data file is created with digitization the necessary element in conversion. The figure illustrates the paths that may be followed to computer format geographic data.

Actual data encoding involves separate tasks for encoding geographic and attribute data. For image or area data there are:

A) identification and selection of the entity to be encoded from the data source;

B) determining the location of entities by measurement with either manual or automated techniques; and

C) recording the locations on some medium such as cards, tape or disc.

For attribute data the tasks are:

D) for each entity, recording on machine readable base, the codes for the various of the attributes; and

E) recording corresponding identification codes for linking geographic boundaries to descriptions for those boundaries.

The options for identifying geographic entities are relatively straight forward. Point data usually describe separate entities while grid or polygon data rarely describe completely homogeneous areas. So explicit decisions are required on assignment of values and boundaries. The amount of descriptive information retained must be weighed against the problems inherent in the handling of various classification codes.

SPATIAL DATA ENCODING

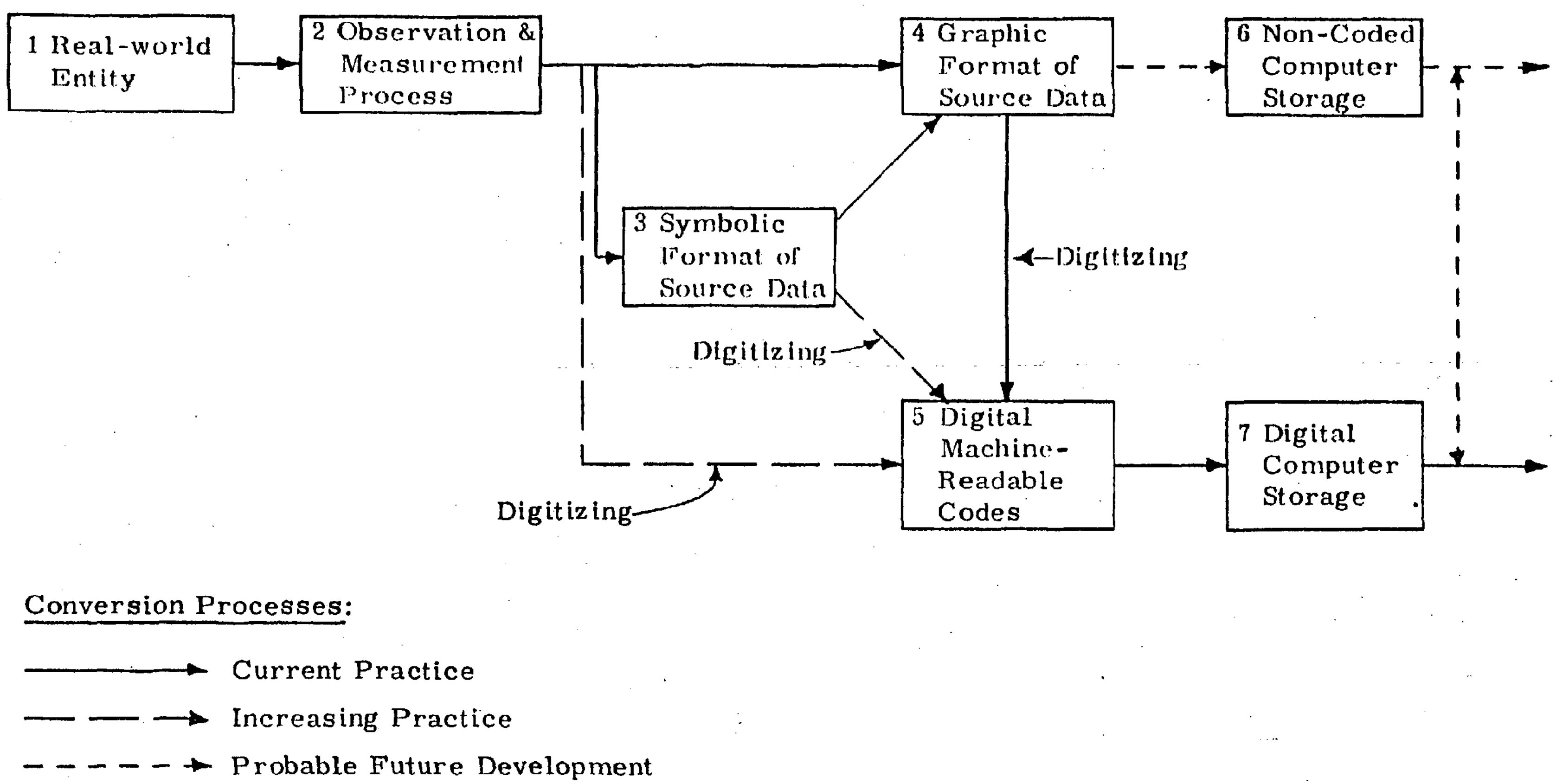


Figure 1.15

(source: IGU 1976)

Compared to encoding descriptive data, the encoding of graphic data is a much more complex process, as when complex spatial entities must be described, the location identifier component of data can become cumbersome because of large data columes. There are various methods to lessen these problems, however, it is easy and logical to represent point data by points in computer storage but it has been found convenient to represent irregular area data types by grouping information into arbitrary grids. This process involves a loss of information in order to achieve an easier data management function. Most systems in operation are based on grid encoding but some newer systems are now acquiring the ability to handle multiple data formats including line, point, and area data.

The points codes for point data are generally recorded as x, y coordinates and coding can be accomplished by either completely manual means or with digitizing. For large computer assisted systems it is typical to use a digitizer that is capable of recording x and y values for points used by an operator who moves the digitizer point over each point on the map and depresses a key which automatically records relative x and y locations with respect to an arbitrary origin.

The optimal line encoding method is to encode each line segment by identifying end and curvature points with x, y coordinates. Data in this format lends itself to specialized analyses such as network and flow analysis.

Encoding data to grid cells has been extensively done in the past. Typically, an arbitrary grid is defined, then overlaid on the map area to be encoded and then manually, semi-automatically, or automatically record the appropriate classification codes. Although grid cell encoding always results in a loss of information, the loss is not uniform and may not affect utility of the data. Loss of information relates to both grid cell size and the manner in which classification codes relate to the data and how they are assigned. Not to be confused with grid cell, encoding a number of systems convert point, line, and polygon data to gridded formats for analytical use even though they were recorded in other formats. This is because the computations associated with the overlay of grid data are relatively simple when compared to other data formats.

For areal data represented as polygons, the optimal encoding method is to describe polygon boundaries with strings of x, y coordinates to define approximate polygon boundaries. For most applications, the polygon is utilized because it leads to the least amount of information loss. Ideally, of course, systems should have multi-format analytical capabilities because such systems can measure and compare data encoded in different formats.

The final aspect of data inputting is data structuring and storage. The important parameters associated with data storage are the medium (cards, tape, disc), the organization of data (file structure), and the advantages of various types of organization, as there are four basic modes of digitizing. The distinguishing characteristics are the amount of capital investment requires, speed of digitizing, degree of error proneness, volume capacities and the amount of labor required. Table 1.2 presents four basic methods for accomplishing actual graphic to digital data conversion.

The digital data files that are created will always require editing. The kind and rate of errors are a function of the digitizing method. Different digitizing methods create different general error types so editing requirements are generally very system specific as there is no list of typical errors which applies to all systems.

In contrast to geographic boundary data, attribute data encoding is a straight forward process. Three important aspects of attribute encoding are:

1) accuracy of the coded data;

2) standardization of classification codes;

3) structure of the classification code.

These three significant parameters relate to the fact that the advantages of various coding schemes must be weighed as the complexity of data manipulation procedures and storage file structure is strongly affected by the nature of classification codes.

GRAPHIC-TO-DIGITAL CONVERSION

TYPE*	METHOD	CHARACTERISTICS
PURELY MANUAL	Manual conversion to digital format, or Key punching of symbolic records, or both	Minimum capital investment, slow, prone to error, not suitable for medium or large volumes of data, labor intensive, and highly selective
MANUALLY AIDED MACHINE	Point & Line Tracing Digitizers - Batch (for digitizing graphic records with direct human assistance) - Interactive	Low to medium capital investment ($10K-80K), high selection capability, prone to error but can have high editing capability, not suitable for large volumes of data, high labor requirements, and moderate speed
SEMI-AUTOMATIC	Automatic Line Following	Medium to high capital investment ($300K+), low selection but good editing capability, no human errors, suitable for large volumes of data, low labor requirements, and relatively fast
AUTOMATIC	Graphic Scanning Devices OCR (Optical Character Recognition) Digitizing at Source (direct from sending devices)	High initial capital investment, low error rate, low capability for intelligent data selection (1976), suitable for large volumes of data, low labor requirements, and fast

↓ Increasing Automation

* **A continuum of methods can be recognized, which extends from those that are basically manual** techniques to those that are highly automated. Four arbitrary subdivisions of this continuum are given above.

(source: IGU 1976)

The first step in data entry is partitioning the graphic data base. Source documents must be divided into logical workable units. This involves locating all boundaries between records, determining appropriate coding and developing procedures for linking records if spatial aggregation is required.

The next step, the encoding (digitizing) of graphic and descriptive data is the major data transfer function in setting up a computer oriented data handling system. The process involves recording geographic boundaries and descriptors. For each type of spatial unit (lines, points, grids, or polygons) the options for coding descriptive data (i.e., a soil or land cover type are as follows: a single value for the predominate characteristics; recording the distribution of elements across the unit thus showing proportions; recording the presence or absence of entities; or recording the proximity of elements to other elements.

The final data inputting step is to organize the data into useful data files for storage and easy retrieval. For large systems, there is a need for a number of techniques created just to create, maintain and retrieve data from storage. Such programs are referred to as data base management programs and will be discussed in the next two sections.

Without becoming to specific, this area of discussion is important because data organization in computer memory greatly affects both the efficiency and complexity of retrieving and analyzing data for its conversion into information. For example, overlaying a number of maps with a certain file structure may be a trivial effort while another structure may require extensive programming.

The simplest data organization is called sequential in which one piece of information directly follows another. An early structural variation on this which is seldom used today is direct storage where enough space must be saved for the largest variable even if all other variables have small data volumes. As well, if data is sparse a large number of potential slots may be wasted.

Inverted file structures create separate tables for each entry (see Figure 1.16). Data is

Figure 122.21
Sequential Storage
of Grid Data

1,1	yes		0.2	......	1
1,2	-	9.6	0.5	......	2
1,p	no	5.0	.92	......	6
1,N	-	17.2	.64	......	-
2,1	no	-	-	......	4
2,2	yes	4.1	-	......	-
2,N	-	1.3	-	......	3
3,1	no	4.2	-	......	-
M,N	no	-	.81	......	2

Figure 122.22
Inverted Storage of Grid Data

A

1,1	yes
1,p	no
2,1	no
2,2	yes
⋮	
3,1	no
⋮	
M,N	no

B

1,2	9.6
⋮	
1,p	5.0
⋮	
1,N	17.2
2,2	4.1
⋮	
2,N	1.3
3,1	4.2
⋮	

.....

K

1,1	1
1,2	2
⋮	
1,p	6
⋮	
2,1	4
⋮	
2,N	3
⋮	
M,N	2

Figure 1.16

Two Types of Grid Data Storage

(source IGU 1972)

stored by columns which wields a good deal of flexibility as well as being faster for operations such as overlaying but are also extremely complex in their arrangements.

The above structures are applicable to grid formats because of fixed lengths for each record but polygon data is difficult to arrange into standard size units. With polygons, grids are not used and data density from data type to type may vary greatly (i.e., census vs. soils data) and polygon size varies greatly, so in polygon systems there are generally both single short records and very lengthy complex records which must be combined, overlayed and analyzed. The difficulty arises because generally computer storage is linear. Approaches for polygon storage often involves just listings with data positions signified with various types of pointers. Pointers are values which are organized to easily locate data with complex structures. With pointers for referencing the actual data, items can be added, deleted, located and retrieved easily. Numerous pointers are required (one for every data element), but the trade-off for being able to easily reference variables has proven to be a satisfactory one.

In general, but with exceptions, the process of data update is simplest for simple data structures and more difficult as complexity increases. However, this depends on the nature of the required update. In many systems it is easier and faster to add a data item than it is to delete or change one. The cost of update can usually be decreased for most structures if the updating can be batched, that is, if the data to be added or changed can be collected for a period of time, and all changes made at once, during a single run of the update program.

From an information systems users the problem of data organization is to select and design a file structure which achieves the appropriate balance between considerations of flexibility, transferability, reliability, response time and costs. One should note that cost is composed of several factors: storage volume, amount of processing time, the required computer, costs of development and time costs. The last two costs can be decreased by using existing systems or parts of systems whenever appropriate.

With modern computers, there are a number of storage medis which differ mainly in their relative speed of access and cost. The basic relation is simple: the faster the access to data, the higher the cost. The two common forms of storage outside of core memory which is insignificant are discs and tapes. Tape provides a low cost storage medium which is slow because tapes must be mounted. A system oriented toward non-interactive inquiries, where reponse time between the initiation of an inquiry and production of results may run from hours to days, can make good use of tape storage. However, a system which is required to provide quick response to a wide number of users on an interactive basis must maintain a data base on disc in order to avoid major delays. Table 1.3 presents relative cost trends for on-line storage.

Table I.3

TYPICAL COST OF ON-LINE STORAGE TECHNOLOGIES
(order-of-magnitude figures)
Cost includes transport mechanism cost

Type of Storage	Cost per Bit	Size of Storage (millions of bits)	Access Time
Register	$10	0.001	0.01 microsecond
High-speed cache	$1	0.01	0.1 microsecond
Main memory	10 cents	1;10	0.5 microsecond
Large core storage	1 cent	10;100	5 microseconds
Drum and fixed-head disk	0.1 cent	10;100	0.01 second
Movable-head disk	0.01 cent	100; 1000; 10,000	0.1 second
Magnetic strip	0.001 cent	1000 to 10,000	1 second
Mass (archival) store	0.0001 cent	100,000 to 1,000,000	10 seconds

(After: Martin, 1976)

3.2.3.3. Data Manipulation Procedures

This set of procedures is required to establish a required bridge between data entry and data analysis. This step is necessary because it is often important to change data from a format which is efficient for storage to a format which is efficient for analysis. For example, data may be efficiently recorded by lines and vertexes but the system may require data to be recorded on the basis of cells for efficient analysis. A flexible system will contain capabilities for a number of manipulations to convert data to formats that provide efficiency or various analysis procedures.

Three types of manipulations are most important: (A) methods of projection change;(B) methods of mode change; and (C) the joining or separating of data records.

(A) Many map projections have been developed and a number are being used in spatial information systems. The ability to change data for a file from one projection mode to another is a desireable system capability. Generally, a system must be able to handle free x, y coordinates, geodetic coordinates, and plane coordinates. Projection change using the computer is becoming a relatively simple process. Transverse Mercator Projection, UTM Coordinate System, and the State Plane Coordinate System are the primary geo-referencing systems in the United States.

(B) Format or mode change involves the operations of reclassifying descriptive data, changing data and merging similar value records. Important aspects in this are changing line data to grid data. This operation has proven difficult because of the required complexity of programs for interpolating grid area boundaries from irregular polygons and uneven grid borders which result from polygon double digitization.

(C) For analysis, data arrays must frequently be broken down into sections due to limitations in computer memory size and analysis procedures. Records to be separated or joined are generally rectangular and the same size for keeping the math involved relatively simple. When data sets are grid rather than lineal or polygon the problem is trivial. For polygon data, data strings must be tested for boundary violations once joined and this is a relatively simple task. The entire process though relatively simple is critical because a small amount of mismatch can produce critical errors once separate maps are joined.

3 .2.3.4. Data Analysis and Retrieval

Retrieval refers to locating a file (data set, series, maps, photos) within the storage of the computer and extracting desired data from the retrieval file. Data analyses are the set of operations performed on retrieval data to produce information. Retrieval and analysis procedures include:

- retrieval from storage
- analysis on locational identifiers
- analysis on descriptive data
- analysis on locational identifiers in concert with data elements

The retrieval and analysis parts of computer aid spatial information systems appear to be quite varied, but there are several generic classes of computer aided operations which will be discussed.

Basic procedures involve data classification and aggregation. Though systems should ideally accept data completely unaggregated, the amount of data would be staggering. The basic principle central to having workable amounts of data in generalization. This onvolves the categorization of different data elements into large more homogenous areas of differing levels of specificity. (For example, a statewide soils data base may contain soil types which have to be aggregated into soil series for analysis because of data volumes.) Class intervals, defined categories and defined locations are all tools for data aggregation.

Some important data analysis procedures which are common to many systems are:

A) Search - this is used to locate single items; sets of defined items; locate undefined items; locate items based on data linkages; locate items based on criteria in the data set; and search by location.

B) Contour - is the process of drafting area value maps from point data.

C) Overlay - consists of matching different data sets for the same area to produce composite overlays. Grid overlay is generally used but polygon overlaying is developing and most systems currently being implemented or designed are polygon oriented.

Other procedures which are often readily applied are contour, smoothing, interpolation and extrapolation, trend analysis, path search analysis, and area calculations.

3.2.3. Data Output and Display

The output from spatial information systems need not always be a spatial display for numerous simple land resource associated models need nothing more than a tabulation of numbers. For example, a comprehensive land resource information system might use the following outputs:

A) Production of graphs, histograms, etc., showing relationships between variables and indexes. An example is a graph comparison of existing land uses at different time intervals. Systems should be able to produce such information interactively.

B) Spatial plots defining existing and historic conditions (economic, environmental, etc.)

of anywhere in the geographic data base.

C) Density plots showing information such as population or levels of agricultural production.

D) Overview plots wherein topography is shown and land uses are plotted on the topography.

E) Controus or isotherm plots to illustrate parameters such as degree of air pollution.

In summary, map and tabular displays form the primary building blocks for geographic information system utility. The advantages and efficiency of various displays are well documented; however, the versatility and timeliness of computer graphics will earn an expanding role in the presentation of spatial information. Good computer displays have to only present the most pertinent information since additional information can be retrieved from a good system quickly and inexpensively. (See Tomlinson, 1972, for extensive review of system graphics.)

3.2.3.5. Technical Procedures Summary

Both polygon and grid cell encoding are widely utilized. The polygon form is generally used in systems where accurate cartographic retrieval is important and the grid system is more widely used where the ability to overlay efficiently is essential and where a range of demands on the system are generally well defined. Table 1.4 displays the advantages and disadvantages of the geocoding options. Interrelationship between the level of data aggregation, system requirements and analytical tasks related to data use and geocoding are presented in table 1.5.

One of the major problems in encoding spatial data is the lack of measures which can assess the effectiveness of the encoding. Comparative measures need to be developed that can make quality control determinations in the process which extends from source document, through capture, to use. As well, there are few effectiveness measures that relate to processes of updating, edge matching and retrieving data.

Table I.4

CHARACTERISTICS OF GEOCODING OPTIONS

Geocoding Options

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

(source Duecker 1975)

Table 1.5

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

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

Key:

SYSTEM REQUIREMENTS

S1 General purpose computer
Card and tape files
General data management software
No specialized peripheral equipment

S2 General purpose computer
Digitizer, plotter
Specialized software for data capture

S3 Dedicated computer with digitizer, CRT, plotter
Specialized software for data capture

DATA LEVEL OF AGGREGATION

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

ANALYTICAL TASKS

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

(source: Duecker 1975)

3.2.3.3. Users of the System

The utility of a spatial information system in meeting perceived resource management needs is the justification for a system's existence. However, even when systems do meet perceived needs, there are additional obstacles to complete system utility. These obstacles include: the potential users perception of the system; user's inability to express problems in quantitative terms; familiarity and bias toward more traditional techniques; and perceptions of costs and benefits of the system when compared to other alternatives.

At the absolute minimum, a users relation unit should be established for any system which should be managed by persons actually associated with system development. User relations include user needs, user access arrangements, user training and formal arrangements for the system to react to user feedback. The following elements should ideally be a part of any system's dealings with its users.

User Needs

- perceived data elements
- perceived manipulative capabilities
- required outputs

User Access

- direct user access
- technical assistance
- formal user organizations

User Training

- system use
- approaches to quantitative problem solving
- expectations of the system
- use of interim products during development

User Institutions

user committees for system evaluation

user input into policy formulation

on system development, modification, and operation

User access to the system is the primary interface since this is where the level of actual system utility is determined. Depending on user experience, the interface can be through computer oriented technician and the sophistication and complexity of the user system interface must increase with the sophistication and complexity of the information system itself.

A user training program must include system access methods and the user must be trained to approach problems with a quantitative perspective in order to understand the range of capability being offered by the system. The problem with most systems currently operational is that system development programs have not operated in time frames that allow extensive user training. As well, meaningful user feedback should follow from the various aspects of the user interface and training.

3.2.3.4. Management of the System

One of the most neglected aspects of spatial information system development is system management. This is not limited to managing system development but includes the entire management perspective for complete operation of the system.

The institutional setting for a system is important because many systems have started with external funding and many funding sources cannot be relied upon to ensure survival of the system.

General policy statements are important and must address issues such as mode of operation (i.e., schedules, priorities, classes of service), interagency agreements, and data confidentiality.

Several kinds of planning activities are important in the management of information systems. Planning program staff and staff changes is important. A major concern should be detailed job descriptions identifying all the skills and capabilities required for operating a system.

A long range fiscal plan is needed because obtaining funds for system initiation has proven to be relatively easy and funds for system continuance or upgrading are generally much harder to obtain. Three vital areas must be addressed in a fiscal plan: total funding; development costs separate from operational costs; and fiscal continuity.

The problem of development costs versus operating costs must be examined rather closely as the distinction bewteen the two is not always clear. However, a distinction between the two should be stated, even if arbitrarily, and should be based on funding source, system operating policy, and expected user community considerations.

In the past, as well as presently, funding continuity often hunges on the delivery of interim products, either for actual use or for system publicity.

In summary, good documentation for all aspects of system management is essential and should include:

the system development plan
the staff plan
the fiscal plan
operational cost information
technical software descriptions
user education programs
user manuals
system use records
all interagency agreements.

3.2.3.5. Procurement of Hardware and Software

The final element of geographic information systems is the procurement of hardware and software which generally involves preparing detailed specifications and evaluating alternative ways of performing the technical tasks

Procurement of the data handling equipment and associated software can involve the following:

Hardware
- competitive bidding
- evaluation of proposals

proposal check-out procedure
performance records
maintenance records

Software
acquire existing packages
software testing
development of new software.

3.2.3.6. Summary of Section III

Section III has presented a brief introduction to the structure, scope, and potential utility of geographic information systems. As well, its purpose is to present terms and definitions which will be used in the remainder of this report.

Essentially the chapter should show the reader that geographic information systems entail far more than computerized cartography. Many planners have looked to simple computer mapping as the answer to their problems when in fact far greater problems with data collection and management may be facing them. Geographic information systems in their fullest sense provide cartographic capabilities but also the opportunity to construct efficient logical frameworks for data base development and management and the coordination of natural resource information users.

SECTION 3A

THE DESIGN MEHODOLOGY

3A.1 Introduction

With geographic data handling systems, most of the important developmental work should occur prior to the actual design of the system to allow systematic and logical approaches to problem solving if anticipated system capabilities are to ever be met. In the early stages of development the approach should be independent of actual data content and the constraints of hardware and software so that when developed, the system can be applied to any data that responds to the anticipated needs of anticipated users of the system.

Geographic information systems are not ends in themselves but are only tools for planners, managers and decision makers. The economics of such systems are a function of the intended purposes of the system and the actual level of use of the information produced by the system. For a system can reflect ultimate technical efficiency and not be a success unless the needs of users are being met at a reasonable cost.

The most difficult aspect of system design to accomplish, because it has been essentially ignored in the past, is the identification of the critical aspects of the interface between the capabilities of a system and the various decision making processes of potential system users. To better understand these processes design must focus on the concurrent development of a number of system sub-systems–data acquisiton, data processing, data analysis as a response to user data analysis needs, and integral management planning of all aspects of the systems operation.

All of the many stated purposes of geographic information systems can be generally summarized into one primary objective–the conversion of raw data into usable information that is actually utilized. The use of any information from a system begins with the final output being transferred from within the boundaries of the geographic system to outside the system-the domain of each individual user of the system. Therefore, system design should deal with issues which are actually beyond the scope of the system in order to understand user decision making, user data and information requirements, and required desired and information uses which will be central to the activities of the various user communities. For the true benefits of using an information system is determined by improvements in the performance of user decision making which are achieved directly by use of the system.

3A .2 System Economics and Design

Marble et. al. (1972), state that "the economics of geographic information systems use is quite simple: are the benefits worth the cost? " Costs are basically a function of what the systems capabilities are and how well those capabilities are executed. The components for determining true system costs are twofold: the direct monetary outlay to design, build, and operate the system; and, secondly, are the intangible costs related to the legal, political, and social realms of system functions. Such costs simply relate to the question: are there potential benefits to be derived from implementing no geographic system at all?

For the first cost, data collection and inputting costs, system development costs, system operating costs, and system maintenance costs are all generally quantifyable for entry into the decision making process. The second cost, however, presents a number of considerations which at best can be only examined and estimated and perhaps only inferred, but which are significant nevertheless. Related to these costs are assessments of information system use and potential use once implemented, which are extremely involved and complex assessments which are possible only through understanding the decision making processes of the various potential users. However, to date geographic information system use has generally not advanced to where these user oriented parameters are generally observable and quantifyable in terms of actual reliance on the various systems currently in operation as vehicles for imporvement in user decision making.

This previous discussion is important because the design approach to be outlined and initiated in the remainder of this report is based on the assumption that the economics of implementing a geographic information system are conditioned either positively or negatively by the design and development process. The first step toward maximizing positive system economics is the development of a systematic, long term, and interative design process. The end result of such a process is ideally a matching of information system output capabilities and the requirements of the community of system users.

3A.3 Initiating A Design Process

The basis for the development process presented here is basically an interative design process in which a range of alternatives are thoughtfully derived and then tested. Alternatives are weighed against system objectives before significant commitments are made prior to actual system design. Quite logically such a process begins rather generally with decisions of greater technical complexity and detail required in successive iterations of the process. This report in the context of the proposed process is regarded as being part of a first iteration and broadly covers the various concepts of perceived system need, a general review of data and other resources which are available, general system specifications derived from new continuing research priorities. Later iterations will represent the first real visible attempts to establish a geographic information system for Maine, should this process continue. By this time, the apparent needs and objectives for development should begin to surface to the potential system sponsors and advisory planning groups which should be actively involved in the design process once it is truly underway.

3A.4 Design Methodology

The proposed framework for the design of a geographic information system is based on the assumption that such a system consists of: (1) all potential system participants; (2) a processing capability that often involves computer hardware and software; and (3) data. System performance, recognizing these three elements as being part of the system, can be affected by: (1) the management and administration of the potential system; (2) the legal-institutional status of the system; and (3) actual political response to outputs from the system.

Recognizing these system components a three stage information system design methodology (see figure 2.1, Calkins 1972) which includes six integral steps is proposed. These steps include: the determination of system objectives; an assessment of data availability; an assessment of procedures, hardware and software which are available; a compilation of general system specifications; generating alternatives; and selecting a system or specifying a set of procedures that will satisfy the objectives of the system to the maximum possible extent.

Each step in the design process involves several types of decisions which have to be made, each of which relate to one or more decision variables. Such variables are regarded as decisions which are made in the process of designing a system. For example, for each variable at every step there are generally several options open (i.e., storage or data on disc, cards, or tape and selection of map scales for inputting-1:100,000 to 2:24,000) and each decision must be made explicitly recognizing the numerous interdependancies between such decisions.

If the six primary steps from the process outlined in fig. 2.1 can be aggregated into the three stages then this report represents a first iteration of stage 1. The systems approach delineated in figure 2.2 presents the general steps which must be a part of each iteration in the design process. Because the process is viewed as being systematic there are required feedback mechanisms which must be eventually structured and monitored as part of the process. Figure 2.3 presents a general

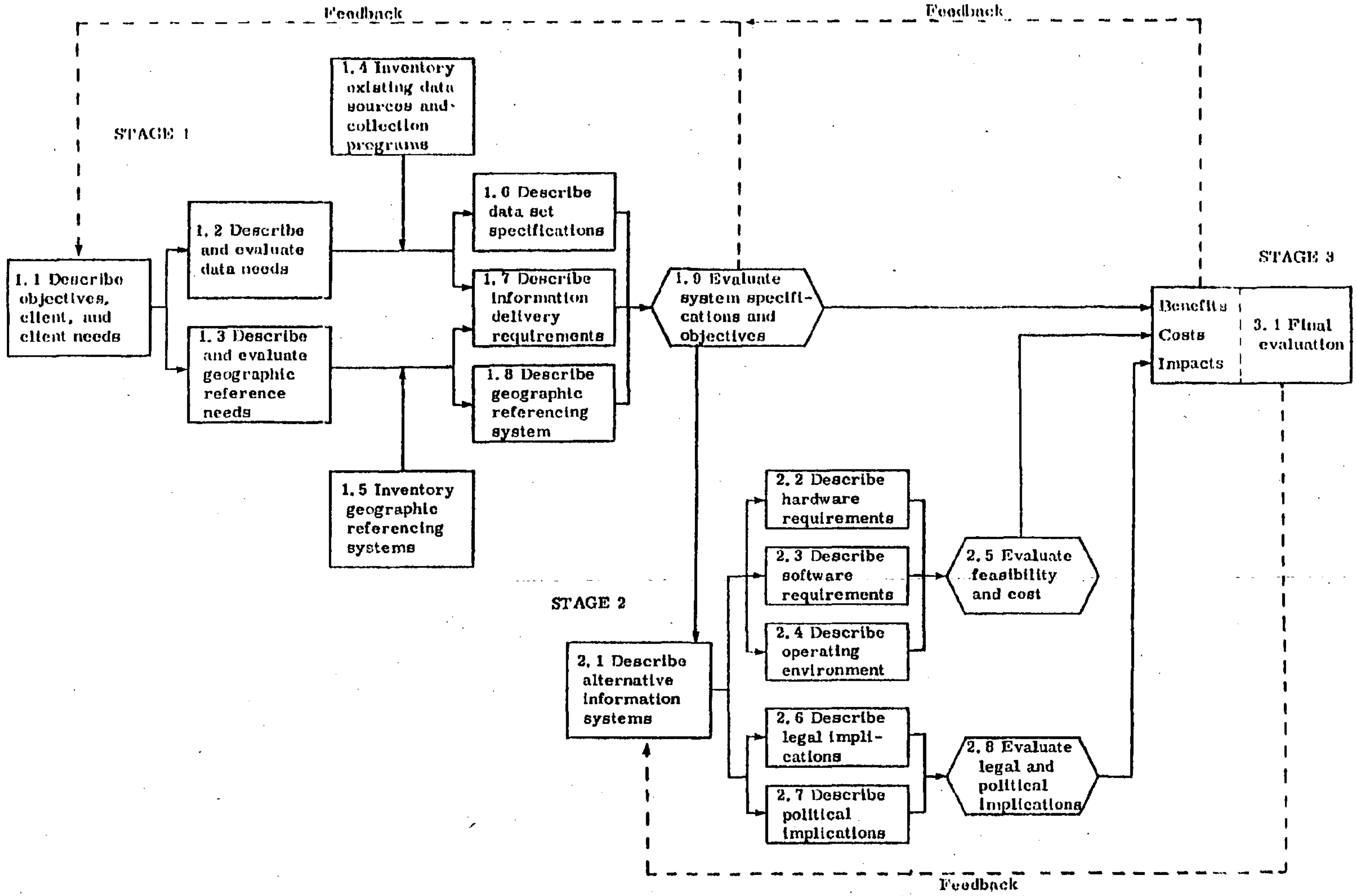


Figure 2.1 Geographic Information System Design Methodology

(source:Tomlinson 1976)

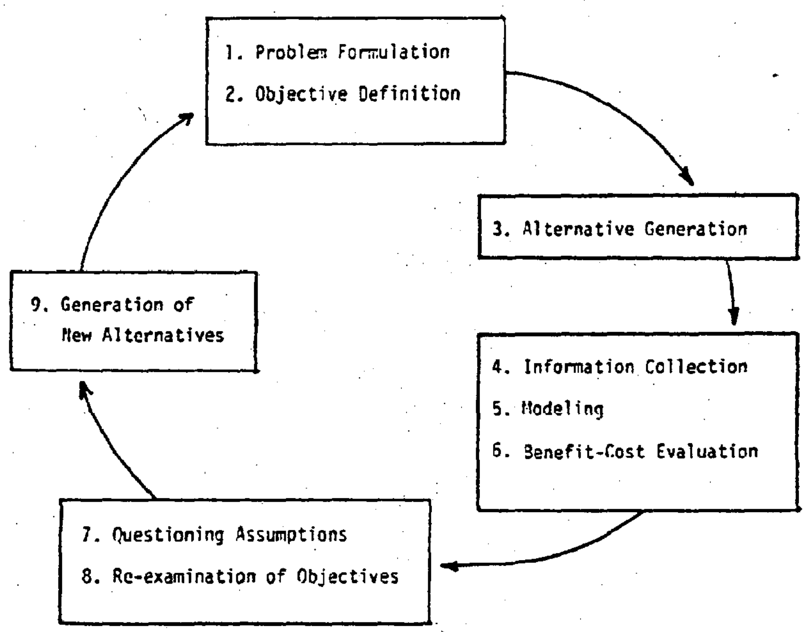


Figure 2.2
A Systems Approach

(source:Calkins 1972)

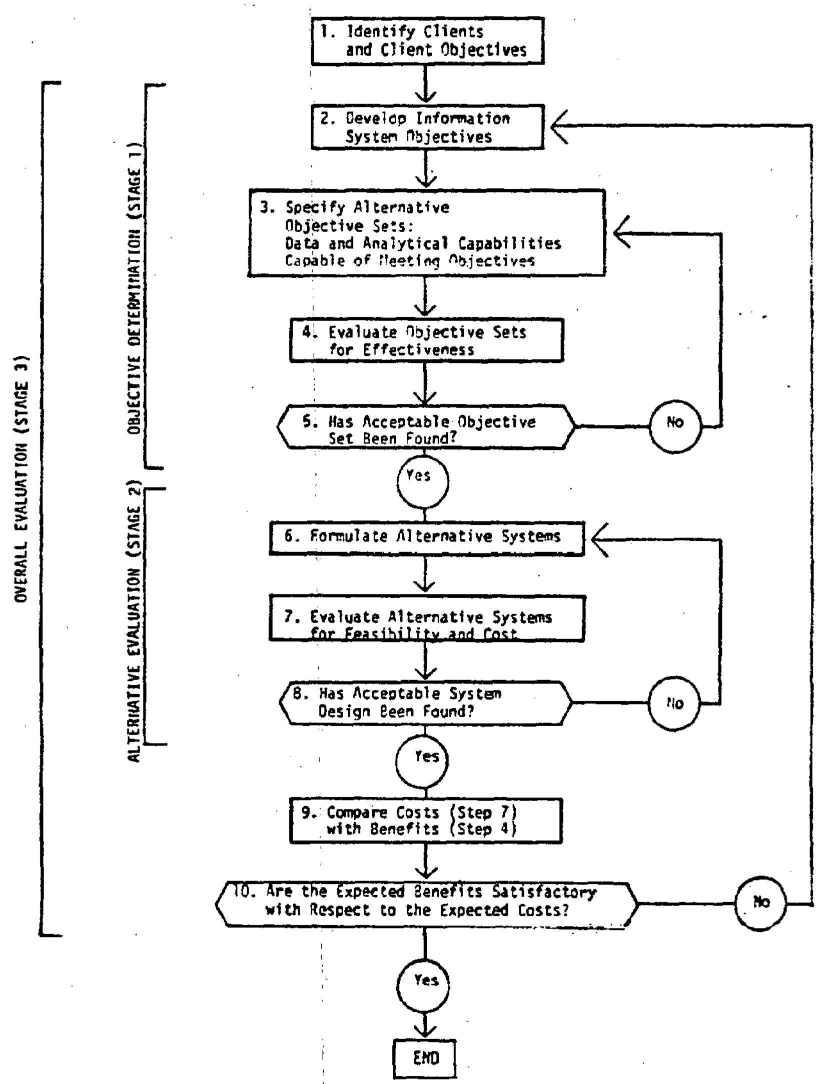


Figure 2.3

GENERAL MODEL FLOW AND FEEDBACK LOOPS

(source:Tomlinson 1972)

outline of how such feedback can be tied to each iteration of the stages in system research and development.

3A.5 Linking the Design Methodology to Geographic Information System Components

The previously introduced tasks which are proposed to be part of the design program can be related to the geographic information system components which were discussed in the second half of chapter 1. The relationship of the primary components (i.e. data base management, data storage and retrieval, etc.) to the primary tasks of system development is presented in Table 2.1. Table 2.2 illustrates a similar relationship for the design methodology to the various components of system support. The value of these displays is that they illustrate which topical areas need to be carried out simultaneously or separate from other aspects of the process.

3A.6 Linking the Design Methodology to a State Resource Data Planning Program

Many of the tasks identified as being integral to the proposed system design methodology are also central to a developing statewide data planning effort for which the need is widely recognized and preliminary formats have been proposed. As the basic intent of data planning is provision of an organized framework for coordinating state agency data related activities, the line between data planning and information system planning is for the most part negligible.

Data collection planning, data collection, data interpretation, data analysis. data storage, and finally data distribution have been initially identified as required components central to the development of both efforts. Data planning provides for inventory natural resource information, assessing user needs for various data types, and preparing background reports on the utility of these data types. Data planning is also tied directly to Federal collection activities, the collection activities of other state and regional agencies and possibly data collected and

STEPS OF THE DESIGN METHODOLOGY	GEOGRAPHIC INFORMATION SYSTEM COMPONENTS		
	Data Base Management	Data Retrieval and Processing	System Support
<u>Task 1</u> Determination of System Objectives	Identification of data elements Identification of geographic location identifiers		Identification of user and user's needs
<u>Task 2</u> Assessment of Data Availability	Identification of available data Documentation of available data Investigation of geographic referencing system		
<u>Task 3</u> Assessment of Existing Procedures, Equipment, and Manpower Resources	Survey of available procedures Survey of available equipment	Survey of available procedures Survey of available equipment	Availability of trained personnel Legislative and/or administrative basis for creating an information system
<u>Task 4</u> Determination and Evaluation of System Specifications	Data specifications Geographic referencing system	Geographic referencing system System products	
<u>Task 5</u> Determination and Evaluation of Technical Operations Alternatives	Transfer and storage alternatives	Retrieval, analysis, and display alternatives	

Figure 2.4 Relationship of the Design Methodology and the Components of a Geographical Information System (source: Tsachanz and Kennedy 1975)

STEPS OF THE DESIGN METHODOLOGY	SYSTEM SUPPORT TOPICS		
	Use and User Access	Management Requirements	Procurement Procedures
<u>Task 1</u> Determination of System Objectives	User access arrangements User training User needs		
<u>Task 2</u> Assessment of Data Availability			
<u>Task 3</u> Assessment of Existing Procedures, Equipment, and Manpower Resources	User training	Long-range staff plan Institutional setting	Software Procurement Hardware Procurement
<u>Task 4</u> Determination and Evaluation of System Specifications	Formal institutional arrangements to accept, evaluate, and act on user feedback	Documentation Product publications plan	
<u>Task 5</u> Determination and Evaluation of Technical Operators Alternatives		General operating policies Long-range fiscal plan	Software Procurement Hardware Procurement

Figure 2.5 Relationship of the Design Methodology and System Support Topics

(source Tsachanz and Kennedy 1975)

archived in the private sector. Interpretation and analysis primarily is designed to provide technical assistance in the use of surficial geology, wildlife resource, water, soils vegetation, and marine resource data. As well specific analysis such as assessments of development suitability may be undertaken when required. Data storage, retrieval, and distribution planning will certainly be required to examine the storage of imagery and mapped information, manual vrs. computerized data handling, and alternatives in distributing this information to the user community. Appendix 2 of this report presents the outline for the proposed data planning effort. Table 2.3 presents a summary interpretation of the linkages between the geographic information system design methodology and the proposed data planning program.

THE DESIGN METHODOLOGY	DATA PLANNING PROGRAM		
	DATA PLANNING	DATA COLLECTION	DATA STORAGE, RETRIEVAL and USE
SYSTEM OBJECTIVES	INTEGRATED USER NEED ASSESSMENT	LOCATE DATA SOURCES FOR G.I.S.	AMOUNTS, TYPES AND DEMANDS FOR DATA
DATA AVAILABILITY	PLANNING TO INSURE SYSTEM CONTINUITY	DATA ACQUISITION FOR G.I.S.	MAXIMUM EFFECTIVENESS IN TERMS OF DEGREE OF AUTOMATION, COVERAGE, RESPONSE TIME, ETC.
HARDWARE AND SOFTWARE	MAX EFFICIENT UTILRATION OF COMPUTER HARDWARE	DATA CONTINUITY FOR HARDWARE	EFFICIENTLY UTILIZE STATE OF THE ART SOFTWARE & HARDWARE
SYSTEM SPECS	PLAN SYSTEM SCOPE	TIMETABLE FOR ACQUISITION & UPRATING	PLAN FOR BEST MATCH BETWEEN REQUIREMENTS AND CAPABILITIES
DESIGN ALTERNATES	SELECT AND IMPLEMENT BASED ON EVALUATION CRITERIA	LEVEL OF AUTOMATION	PLAN FOR BEST MATCH

TABLE 2.3

DESIGN METHODOLOGY DATA PLANNING PROGRAM STRUCTURE LINKAGES

SECTION IV

ELEMENTS OF SPATIAL DATA SYSTEM DESIGN

4.1 DETERMINING SYSTEM OBJECTIVES

4.1.1 Introduction

Determining the objectives to be met by the system is always the first major task in system design, yet it is often the most difficult part of the design process. Resource planning and management, because of its strong interface with political decision making, often does not lend itself to clear, concise statements of objectives. Despite this difficulty the process must be initiated with an attempt to determine objectives by specifying the various data requirements and products applicable to the perceived resource management problems of potential system users.

The process must be iterative to the point that it should be carried out during and after a system has been implemented to allow performance evaluation to determine if changes are required or desired to increase the utility of the system. With such a model being used, the process is applicable to initial system design and subsequent modification as the system should never be viewed as a static tool but as one which must be able to be altered quickly to respond to changing land use and resource planning demands.

It is proposed that the objectives be determined by defining: (1) the system objectives for each potential user; (2) the data items and analysis capabilities which should be included in a system; and (3) using the components of 1 and 2 in an integrated manner for a total system evaluation framework.

4.1.2 Identifying Users and User Needs

This first step deals with: (1) identifying the decision makers which will use the system; (2) identifying the line agency users of the system; (3) identifying user objectives for using natural resource data; and (4) estimating what the impacts of using the system will be on user decision making.

Tomlinson et al. (1970) point to using decision matrices (Fig. 3.1) as a method for identifying proposed system users to describe their roles, actions, and information requirements. Using this tool through a number of iterations it should be possible to define classes of potential users. For example:
primary users may be identified because their aggregate need will essentially specify the scope and type of system implemented;
secondary users will have most of their needs met by the system but not by virtue of these needs forming actual system specifications;
tertiary users are all other potential users whose needs are implicitly met (that is, the system is valuable as a tool but tertiary user needs did not weight heavily in the design of the system).

The matrix is aimed at assessing user objectives by identifying the methods available to each of them for implementing planning and management decisions. This determination is valuable because information need is based directly on these actions and examination of potential implementation tools allows an approximate assessment of the impact of having or not having various information types or an information system. Also information that is not directly linked to some user activity may be of interest but it is of little utility and therefor should be of low priority for inclusion into any system data base.

4.1 .3 Identifying Data and Data Hanlding Requirements

This task is important for developing data requirements primarily in terms of content. Data which is to be identified must be directly related to user objectives which have been defined. The end product of this assessment should be described in sufficent detail to allow the preparation of relatively simple system specifications. Parameters such as scale, classification schemes, currency and accuracy must be adopted with standards adopted. Figure 3.2 presents an idealized diagram of the elements which should be a part of this determination. This is amodified tree diagram which is utilized to summarize the relationships between user objectives, information requirements, data content, and data specifications. It is also potentially valuable to prioritize information as economic considerations will surely limit the amount of data that can be included in the system. Weighting techniques can be applied for prioritizing information. Even counting the number of potential users requiring the same general data and information types places relative improtance on different data and information types.

4.1 .4 Determining Geographic Identifiers

Manual systems rely on maps on which locations of various data types are preserved and directly observable. However, if data are to be encoded for computer processing, directly observable information location is lost. Location identifiers are data items which specify relative or absolute locations (i.e., x y grids, latitude and longitude) which must be acquired, stored and manipulated. The process of encoding spatial data for computer processing thus incurrs at least on additional data item for each data set and the loss of easy to observe visual data displays.

This process as previously stated begins long before actual system design commences and is continued for the entire life of the system. A number of important parameters should be accurately estimated with acceptable accuracy before system specifications are finalized and the design process begins. It is the four following determinations which should most incluence user needs inputs into specifying the elements of system scope and structure:

1. For all users current and potential data needs, use, and handling requirements should be identified;

2. Potential future data needs should be examined;

3. Potential future data handling problems should at the very least be identified;

4. Desired future data items and data handling capacities for all system users should be identified.

4.1.5 Identification of System Users

Information system users will come from a number of operating environments: planning, management, research, and operations. Geographic information systems are generally useful to all functions except operations functions which rarely require spatially referenced information. Planning and management agencies and organizations are usually the primary users with research generally secondary or tertiary in terms of priorities of need and accessibility. Thus, in looking for potential geographic information system users those persons and agencies engaged in environmental planning and management will form the nucleus of the user class for determining initial information requirements.

The first iteration of the design process, which will be described in a later section, will look chiefly at state agencies making up the Land and Water Resources Council. This group was selected for a number of reasons. First, it is the Geographic Information System Coordinating Committee which initiated this study and therefor can be identified as a group of primary users. Secondly, a number of other potential users (i.e., local communities) may gain access to the system and be involved in actual system design through state departments such as the Planning Office. Thirdly, given the time constraints on this project identifying the complete potential user community would be prohibitive.

With this in mind the agency users to be identified and examined in terms of looking at data and data handling needs and requirements are:

Department of Energy Resources
Department of Environmental Protection
Department of Inland Fisheries and Wildlife
Department of Marine Resources
Department of Transportation
Department of Conservation
State Planning Office

Within each major office the needs of individual programs and divisions will be examined separately. For example L.U.R.C., Forestry, Geology, and Public Lands all of which are in the Department of Conservation will be examined separately.

There have been very few attempts to systematically assess needs for spatially referenced data in Maine. Efforts in the past have been associated with the Maine Information Display and Analysis System Users Group and with development of the Minor Civil Division Data Base within the State Planning Office.

The MIDAS users group (MUG) operating from the early to mid 1970's. With the Technical Services Division assuming responsibility for maintenance of MIDAS in 1972 this group was made up of a number of agencies who at the time were using MIDAS computer data files. This group regularly reviewed data availability, data needs, new data files and new data analysis techniques. The actual focus of much of the system was socio-economic natural resource civil division referenced data which could actually be retrieved for a number of spatial referencing requirements. The Departments of Inland Fisheries and Wildlife and Marine Resources did maintain resource data but this data generally related to permits and licenses and was stored and retrieved in formats generally unsuitable for mapping and direct spatial referencing. User assessments were generally informal without specific forms and questionnaires but reliance was placed on the regularity of the user meetings. A thorough review of MIDAS related activity for the period of 1970-1975 revealed little information on determining user needs and objectives which is of direct utility to this study because MADAS was a statistical data management tool and this study is addressing geographic data systems.

In 1976 the Technical Services Division of the State Planning Office began to work on a revised socio-economic non-spatial data base. By this time MIDAS coordination was absent and data file maintenance and use was extremely splintered. Technical Services, because of expertise and time and manpower constraints began to focus on socio-economic profile data. This data base is being developed specifically for internal SPO use and to maintain the information as a service for other data users. The data base includes elements most frequently requested within State Planning and by other state agencies. Design was started with additional iterations undertakne to aid in shaping the systems growth and expansion. User needs of other state agencies was not the single overriding factor but were important. It is also important to remember that the data base is being put together primarily to satisfy internal requirements for socio-economic information. No attempt is being made to address spatially referenced natural resource information because of the potentially large data amounts and the complexity of the data handling and analysis software which are involved with such information are beyond current capabilities of Technical Services within the State Planning Office. In summary, because this data base is being developed to provide growth, social, economic, housing and expenditure profiles it is viewed as not being applicable to developing objectives for a natural resource spatially referenced data system.

This apparent lack of user needs assessments for natural resource geographically referenced data has led the Natural Resource Division of the State Planning Office to initiate an extensive user needs study for natural resource information. This effort (summer of 1977) is part of the Statewide Natural Resource Data Collection Plan and it will be available to serve as part of the foundation for both a state data collection planning effort and continuing Geographic Information System studies.

4.1.6 User Issues to be Addressed

Based on figure 3.1 the following elements were identified as being important to an initial data related surveying effort:

1. Statements of agency missions and mandates.

2. Data types currently utilized by each agency to satisfy current planning and management needs.

3. Data which is collected and archived by agencies primarily for use by other data users.

4. An assessment of current and projected data manipulation capabilities and problems.

5. The need and methods to link statistical, spatial, and cadastral data systems.

6. The implications of 1 ... 5 on geographic information system design and development planning.

Constant updating of these parameters should continue through system implementation getting more specific in each succeeding survey to be able to accurately estimate:

1. Updated data and information requirements for system users.
2. Current data analysis techniques
3. Required data needs which are not being met at present.
4. Realistically useful data types and analysis capabilities.

4.1.7 The User Survey

The material to be presented in this section was derived from a user needs questionnaire, a follow-up telephone interview and a personal interview. Material is viewed as being preliminary as this is only the first iteration of a multiple step process which must be undertaken to fill in the matrices which were presented at the beginning of this chapter. (A copy of the questionnaire appears in Appendix 3 of this report.)

This section is intended to only yield a brief overview of spatial data handling activities of the various members of Maine's state resource planning and management system. The mission statements which are presented give general overviews of the fundamental aims and directions of data related activities. There are also concurrently data related objectives which are not directly formulated by those framing the various agency missions. The spatial data handling required to support the activities of these agencies can be described in terms of four broad categories:

1. Data gathering and dissemination;
2. Data production (i.e., map cartography);
3. Scientific research; and
4. Land resource evaluation and classification and administration of permits, leases and regulations.

Within these categories there are a wide variety of spatial data applications. Generally data amounts are relativly large and data handling is sufficiently complex to warrant the consideration of explicit state-wide policies directed toward the solution of data associated problems which can be identified.

The volumes of data which are needed and the complexity of data analysis procedures are aspects which greatly influence costs associated with data use.

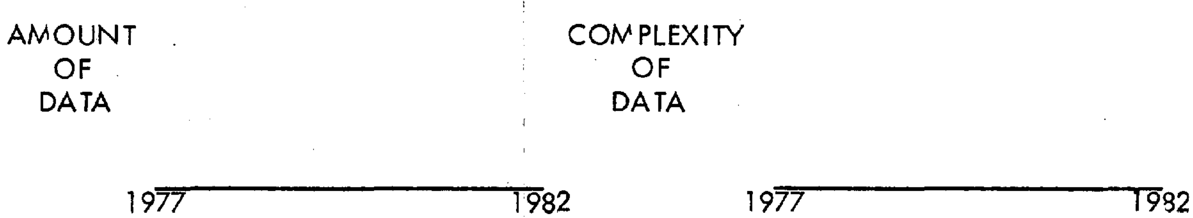
4.T.7.1 The Survey Results

As resources planners at the state level are engaged in diverse management, enforcement, protection and identification activities (Wright 1977) a composite view of their data needs must reflect this diversity. The key to a successful data needs survey and subsequently a complete data planning program is the identification of levels of data need and use commonality. The range will extend from agency specific data (i.e. deir fertility rates) to relatively general data for planning (i.e. land cover types).

Inland Fish and Wildlife

Inland Fish and Wildlife currently relies heavily on internal computer data processing capabilities for data use and storage. Much of this effort is associated with updating their data bases for individual species plans, liscensing, and regional modelling for wildlife planning. Approximately 20 Midas data files are maintained for these efforts using standard geographic referencing and data types coding.

The agency's data handling system is well established and needs appear to be well prioritied. Personel were able to estimate two important data related trends (amount and level of complexity in data use for problem solving) relative to present and projected activities.



Aerial photography and various map data types are widely utilized. Major data needs are detailed land use and cover types mapped and statistically summarized for each minor civil division and organized territory. In summary this agency is handling current data types well, and appear to be well situated for handling future data as it becomes available or is collected. Specific data handling tools such as computer graphics could be well utilized because of current agency reliance on an automated data base.

Department of Transportation

the Department's Bureau of Planning maintains a continuing review process for all transportation related activities with respect to environmental impact assessment. It also formulates environmental policy and administers the scenic highways program.

Primary data files are the statewide system of highway maps, various vehicle related parameters, road inventory data, accident records, speed data, and commodity movement data. The Planning Bureau relies on general data such as forest types, topography and slope land use change, housing population, and wetland maps as basic input for their planning activities.

The Bureau also does utilizes automated data files and digitizer recorded data for a portion of their total detailed data base.

In summary Bureau personnel felt that most of the data they require is available and being utilized, however, personnel did recognize that significant improvements can still be made in their techniques for data storage, and retrieval. Evidence such as increased digitizer encoded information attests to their attempts at streamlining data activities.

It certainly appears that a number of the previously discussed general data types (i.e., soils) could be held in a central automated data bank and completely satisfy current and projected Bureau needs and be equally available to other organizations with similar data needs (see Wright 1977).

Bureau personnel did not feel they had a sound basis for projecting data and data handling needs for 1977-1982 so a projection is not included here. Such projections are therefore omitted from the remainder of this report because this specifically appears premature.

<u>Department of Conservation</u>

Bureau of Public Lands

The Bureau manages public reserved lands under principles if multiple use, assembles land in larger tracts, inventories the naural assets of public lands and does agricultural and timelier management planning for these areas.

Primary data files and data types required for planning are: The Coastal Islands Registry; The Public Lands Resource Inventory; and general data types such as forest stands, agriculture, wetlands, soil types, topography and slope, parcel boundaries, wildlife resources and lake water data. Most pressing needed data are detailed timber and soils mapping. Where appliable the Bureau does utilize standard MIDAS and state codes for all data records.

The Bureau relies heavily on data which could be efficiently retrieved from a common data base and aggregated for detailed analysis using standard geographic information system capabilities such as overlay, area, search and regression analysis.

As mentioned, Bureau personnel felt unprepared to accurately portray projected data volumes and data handling requirements for 1977-1982. But they did generally feel both data amounts and the complexity of required analysis would increase sharply.

Bureau of Forestry

The Bureau carries out timber improvement studies and woodlot reconaissance, provides technical assistance for L.U.R.C. permitting and shore land zoning, and provides assistance to timber operators and processors. There are a number of specific task related divisions within the Bureau.

Primary data files are the Forest Insect and Disease Survey; Fire Danger Records, Fire Reports, stumpage sales, lists of logging firms and processing reports. General data types used are topography and slope, vegetation and soils. Currently the bulk of forest related map data and statistics come from the USDA Forest Services Forest Inventory which are conducted at 10 year intervals. Major data needs are detailed forest cover typing, forest suitability data and parcel ownership information. The Bureau has unsuccessfully partially relied upon MIDAS files for data storage and retrieval. System problems interfered with the data reporting and analysis requirements.

Bureau personnel felt they could in the future adequately plan with a statewide data base of forest types, soils, parcel boundaries, and other general data types. Such a data base could be readily stored in and retrieved from a larger statewide geographic information system, and formatted and aggregated for specific bureau purposes. More specific data types could be accomodated in such a system as well or housed managed, and analyzed inhouse by the Bureau. Personnel felt they had no solid basis on which to make data and data handling needs projections.

Bureau of Geology

The Bureau maps, interprets, and publishes geologic information and provides interpretive information for planning and regulatory agencies; studies are being conducted on surficial deposits such as sand and gravel, groundwater conditions, beach erosion problems and dune management plans. It works in wetland protection permitting, evaluating dredging permits and is involved in a study for nuclear plant and industrial facility siting. Primary data files include surficial and bedrock geologic mapping, groundwater, and well data.

Bureau personnel detailed the need for data analysis such as contouring from point data, and statistical analysis, and histograms construction. As well the need for retrieving and formatting data for activities such as landfill application review is also a major area of concern.

The Bureau is involved in the development of surficial geolgic and other data types which are essential to land planning and management activities in Maine. A number of data types used by the Bureau are relevant only to the bureau's activities but generally the agency does and will play an éssential role in the development of a central statewide general planning data base, because of its role in the development of most geotechnical data for the state.

Bureau of Parks and Recreation

Parks and Recreation purchases land, and designs and develops recreation facilities and boating access sites. It prepares the Maine Comprehensive Outdoor Recreation Plan, conducts scenic rivers studies and provides technical assistance for local recreation programs.

Primary data bases are Private Commercial Recreation Areas, Quasi-Public Conservation Areas, State and Federal Recreation Areas and Municipal Recreation Areas. Primary data related activities relate to long range planning and facility design. Other information which the Bureau requires are forest types, land cover, wetlands, soil types, landform topography and slope, population, utility locations, fish and wildlife resources, lakes, watersheds marine resources, and floodplains.

Data needs appear to stem from the lack of personnel for extensive raw data interpretation (i.e., from aerial photographs) rather than from the complete absence of raw data

Land Use Regulation Commission

LURC is responsible for collecting and maintaining data needed to regulate and guide land use activities; protect natural resources; and plan the future of the unorganized territories and plantations.

With respect to applications review the data is primarily gathered and provided by the applicant. Existing conditions discriptions and discussions of impacts are required. LURC generally must augment this information with land use, social historic and regional natural resource data. Two efforts, planning to gain an overview picture and establish basic information foundations, and development review which generates a significant amount of data are proceeding together. Most heavily used data types include wetlands, soils, topography and slope, land use, wildlife resources, watersheds, lakes and floodplains.

LURC planners expect the amount and types of data required to increase in the future. Availability of new data types such as LUDA cover mapping will change information related patterns.

Primary data needs are detailed cover typing, complete soils coverage, accurate wetlands maps and periodic high altitude photography. Again, LURC planners have identified data types. Often associated with geographic information systems as being primary in terms of use and need as high priority data types. Data analysis and interpretation activities are projected to increase in complexity and extent in the near future along with the amounts and specificity of this required data base.

Department of Environmental Protection

Bureau of Land Quality Control

This Bureau controls all significant development activity which may have an adverse impact on the environment; regulates location of solid waste disposal sites through the Municipal Waste Disposal Act; has started an extensive program of groundwater monitoring around existing dumps; regulates septage sludge and water treatment plant sludge disposal and protects Maine waters against discharge with emphasis on potato wastes and pesticides; conducts programs aimed at reclamation of mined areas and administers the Site Location Act.

Primary data types utilized are: land cover for watershed work; agricultural non-point sources; soil types and landforms; groundwaters, and topographic and slope analysis mapping.

Also land use, urban housing, population, industrial locations, fish and wildlife resources and watersheds and floodplains are used. With few exceptions their current data required for planning is general resource data aggregated for specific purposes according to the analysis undertaken (basin nutrient budget estimations).

Data needs stem from both manpower shortages and from available data which must be organized or aggregated to satisfy specific analysis requirements. Good runoff and climate data, flow and discharge and water quality data are identified as primary data needs. Water quality data enables agency sampling so expansion will be determined by future bureau capabilities. These agency specific data problems at least initially must be viewed as being beyond the realm of a statewide data planning effort.

Office of Energy Resources

Energy Resources prepares the Comprehensive Energy Plan for the State; developes energy Policy; assists individuals and organizations in drafting proposals for funding by Federal research and development agencies; and maintains a well stocked energy related library.

Current data related activities focus on creating an energy needs data base consisting of needs; storage locations, surplus and transport data for the state. The agency is also involved in the New England Energy Management System which may eventually provide a vehicle for storing and analyzing spatial and statistical data relevant to envery resource planning.

Specific data trends are difficult to project because much of the Departments work is response to specific problems and conditions which may change over short periods of time. In general data types related to population, land use, and transportation will continually be used and specific data types such as fuel oil supplier locations will be activated and deactivated as their need arises and falls. Although incomplete, GIS related general data is valuable to the Departments planning programs and goals when utilized with requisite energy related data parameters which may be too specific for inclusion into any general statewide data base. But, this data should be available from Energy Resources to parties through designed data networks determined by a larger statewide data planning program.

STATE PLANNING OFFICE

H.U.D. Land Use Element

Introduction

This program is designed to conduct policy related studies developing recommendations concerning improvement of Maine's planning and managment framework and complete program applications for H.U.D., Current work elements include state resource use and constraints analysis, conducting policy related studies, and describing Maine's resource managment system.

Current Data Use and Requirements

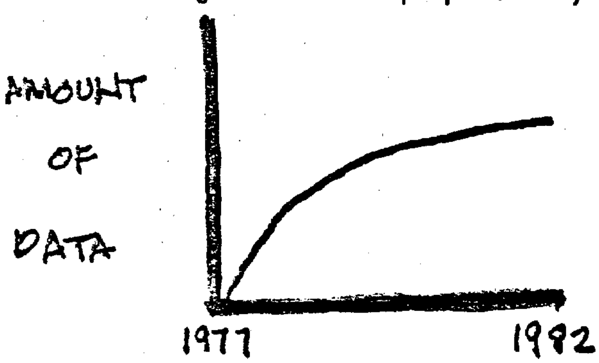
Because the program is state overview oriented, in the past there has been little compilation of data, instead the focus has been on various policy related activities. However, the program is currently involved in the development of a "State Overview Opportunities and Constraints Analysis." This effort involves the preparation of 1:500,000 scale maps, tabular statistics, and supportive text for a number of natural resource and socio-economic land use parameters. The result will be a binder type publication for easy periodic updating of the maps and tabular information types. Specifically, the report will include 20-25 basic inventory maps, initial resource and socio-cultural trend maps, and general capability mapping and analysis. Mylar bases will be archived for all of this information to allow the distribution of print copies to all interested parties.

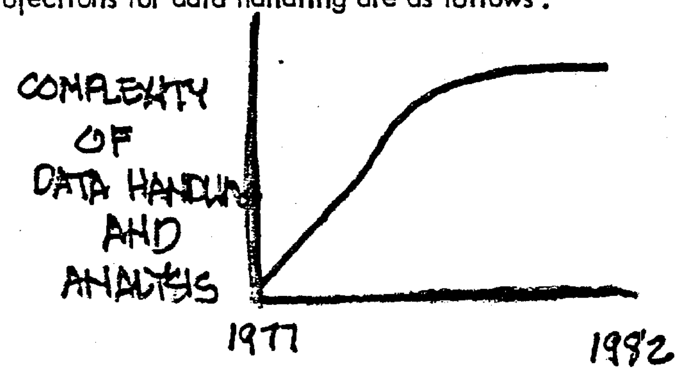
A second thrust, which is still in infant stages of development, is a geographic regional specific resource allocation mapping and planning project. It is projected that this effort will involved SPO, LURC, and the New England River Basins Commission. Again a number of natural resource and land use associated parameters will be mapped and analyzed at scales of 1:100,000 to 1:250,000. Products will include basic single factor maps and a number of composite analysis maps such as special resource areas and locations with high development potentials. Implementation of study findings will be unbinding on regional and local actions but should help shape federal and state resource oriented activities.

A third data group which will likely be a part of this program will be the LUDA Program ouputs available from U.S.G.S.. These 1:250,000 scale maps will include information on land use and cover types, watersheds, census tracts and public land ownership. Computer compatible magnetic tapes for all this information will also be available. It is projected that this data will be archived for both in house uses and for public distribution to other state and regional agencies as well as private data users.

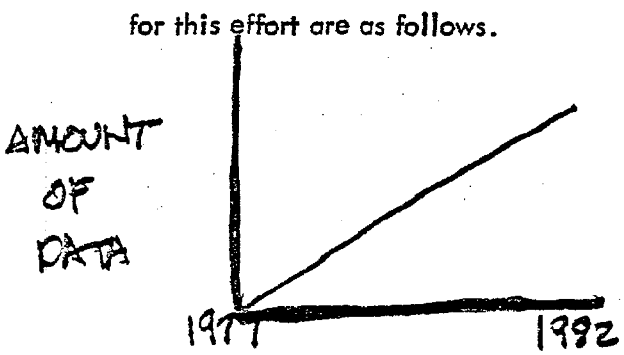
Data Handling Requirements

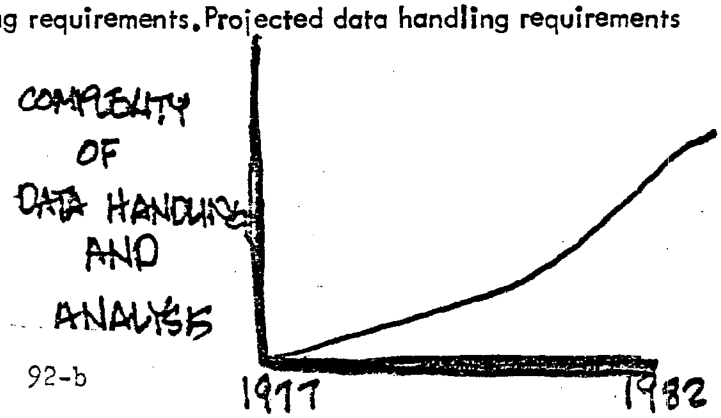
The Overview Opportunities and Constraints Analysis will involved archiving the base materials for composite analysis work and for [redacted] distribution. Eventual automation of this data base is desired for extending analysis and cartographic capabilities. Whether automated or manual this will be the first such overview data centralization for Maine and as such data handling is difficult to project. Five year projections for data handling are as follows.





The larger scale geographic region studies will involved extensive amounts of map data and associated tabular information. For this effort, due to its current initial stage of development it is also difficult to estimated data handling requirements. Projected data handling requirements for this effort are as follows.





L.U.D.A. associated data handling is not viewed as a complex data-handling problem unless there is significant demand for use of the digital data, which is doubtful.

Evaluation of Data Needs and Requirements

The HUD Land Use Element will be involved in compiling, analyzing and distributing a significant amount of state overview natural resource data. As data amounts increase data analysis activity will also be rapidly increasing so by 1982 a substantial data base and data handling problem will have been established.

The data needs questionaire and interview seem to present a clear understanding of the future direction of the program in terms of data amounts and handling requirements being recognized by those in the program. Eventual automation of at least parts of the data base does also seem likely to increase data analysis capabilities.

State Planning Office

Coastal Program

The coastal program provides information for private and governmental decision makers so informed land use decisions can be made. It also provides educational and technical assistance in coastal matters. Major program thrusts are: technical and financial assistance; developing recommendations on major coastal issues; data collection for resource planning; and the publication of resource maps and planning handbooks. The Coastal program has been operating under 305 CZM funding. As such its major role has been one of data provision for planning. A Coastal Scenic Inventory, Coastal Atlas, Coastal Socco-Economic Inventory, Coastal Lakes, Recreation Facilities and Climatologic Reports are the primary data bases.

The program will soon be operating under a 306 CZM funding mandate. As such the primary thrust will become the provision of funds for local initiatives and coordination of local planning activities.

Primary activities are associated with data compilation and distribution. The Coastal Atlas is currently in demand having to meet up to fifty data requests a week.

As well, a number of publications are widely requested and present a moderate handling problem.

The Coastal Scenic Areas Inventory is essentially nonfunctioning at present. The program must also handle a number of random requests for nonprogram generated data which is archived by the program. The building of complete indexes for all archived maps and aerial photographs and compilation of the socio economic activity are also important current activites. It is felt that information handling requirements are currently peake and will begin to taper off.

The 306 program will require no or little generation of new data. A part of the grant does go to statewide coastal planning issue studies but much of this work will be done within other state line agencies. Previously generated data will be the primary data base.

Data handling activities are currently peak requiring full time intern assistance. Should no major changes occur data handling problems associated with coastal program information will fall to line agencies and regional and local planning and management activities.

State Planning Office

Critical Areas Program

The Critical Areas Program consists of two phases: Area registration and conservation implementation. The primary objective is to inventory and compare the value of significant natual features of the state and decide which are really of state wide significance. Major work elements include evaluation and registration, report preparation, maintenance of a registry, a data storage file, and provide assistance to data users.

Most data handling currently relates to the area registration process. Once subjects to be inventigated have been established priorities are formulated on which areas will be looked at. Planning reports which present what types of areas should be included are prepared and then recommendations forspecific areas are made. The Critical Areas Board reviews the nominations.

Data handling activities relate to the publication and distribution of planning reports and information on specific areas. Data currently exists on 62 registered areas; 15 areas currently being registered and 150 areas which should be registered by the end of the year. It is projected that 100-200 areas will be registered a year.

A second data oriented activity deals with areas which are being studied for application of appropriate conservation measures. This activity is infant and data needs are unknown.

As well the areas are archived on $7\frac{1}{2}$ minute guadrangle maps and in descriptive computer files.

Data handling consists of collection, archiving and evaluating data on proposed or nominated areas. With current staffing and funding, given projected registration, data handling should not be a major problem.

Data handling trends show registration should continue at about current rates. As the program moves into implementation greater data analysis and distribution may be required for areas for which conservation implementation is proposed. As well the total amount of data archived by 1982 will be significant as 600-900 areas will have been registered.

Data needs arise with specific area nominations. Primary data collection by the program is not significant, as consultants and persons nominating areas often provide basic data. The quadrangle maps for area mapping are the only map type requiring state wide coverage.

Water Resources Program

This program is involved in comprehensive water and related land resource planning. Primary objectives are intensive water supply evaluations, data acquisition efforts, flood damage reduction planning, future water use projections, working towards water quality goals, and monitoring state regulations in regard to water planning. Current work elements include water supply assessments, data acquisition for river basins, critical areas work, specific pollution studies and watershed measurement projects.

Many of the activities of the program are not specifically data related. However, the program has created and maintained a number of water related data files. These have included the Lake Name Index (MIDAS 906Z); the Dams Inventory (MIDAS 543W); the River Names Index (MIDAS 905Z). Because of lack of technical support these automated files have been turned over to Inland Fis and Game. In addition a Listing of Great Ponds; A Lakes Studies Inventory, and A Map Overlay System (U86S 1:250,000) have been developed through the

the program.

Much data currently used is generated by line agencies such as Inland Fish and Game and Environmental Protection as the programming assumes a planning-coordination role rather than one as a primary data collector.

The handling of automated files had proven difficult so files were turned over to another agency. The program is still central in the map overlay system for the state.

A specific data handling need which has been identified is the inability to access and retrieve existing automated information by river basin for basin planning studies, as an effort to establish basin by basin data needs is also underway. The indent program does not have actual projections for levels of data collection, handling, and analysis for the 1977-1982 time frame.

Note: Department of Environmental Protection, Bureaus of Water and Air Quality, Department of Agriculture and Department of Marine Resources are not included because interviews could not be scheduled while project was underway but will be included in all future interations.

4.1.7.2 Composit User Needs

Wright's (1977) survey of data use and needs for Maine yielded soils, watershed and land use/cover to be critical data elements for statewide planning.

The Data Priorities survey and this work reveal that the most universal problem facing state resource planners and managers stem from shortages of funding and personnel. Because only agency specific goals currently/planners, guide planners procure their own information often resulting in essentially redundant data collection. (exerpted from Wright 1977)

Table 3.2 Data Priorities Based on Data Use and Needs. Map Scales Use for State Level Planning

Resource Category	MAPPING SCALE 1/1200	1/2400	1/6000	1/12000	1/20000	1/24000	1/50000	1/62500	1/125000	1/500,000	Point	Report Tabular
Land Cover (gen.)										X		X
Forest Stand Types				X	X	X		X				X
Other Vegetation												X
Agriculture				X	X	X	X	X				X
Wetland (salt)				X	X	X		X			X	X
Wetland (fresh)				X	X	X		X	X		X	X
Soil types				X	X	X		X		X		
Bedrock Landforms				X	X	X				X	X	
Glacial Land Forms					X	X		X		X	X	
Ground Water Locations			X		X						X	
Topography & Slope	X	X	X	X	X	X		X		X		X
Land Use					X	X		X		X	X	X
Land Use Changes				X	X		X					X
Urban Housing												X
Population Densities												X
Industrial Locations					X						X	X
Parcel Boundaries										X	X	X
Utility Locations										X		X
Transportation Facil.				X	X	X				X		X
Recreation				X	X					X		X
Fish & Wildlife Resources												X
Critical Areas					X						X	X
Marine Resources												X
Watersheds			X	X	X					X		X
Lakes			X								X	X
Flood Plains			X		X						X	X

Source: Wright 1977

Table 3.1

Resource Information Priorities

Resource Category	State Planning
Land Cover	4
Land Use	
Forest Stands	2
Agriculture	
Wetlands - Fresh	
Soil Types	3
Bedrock Landforms	1
Glacial Landforms	1
Topo and Slope	2
Land Use Change	
Housing	
Population Density	1
Recreation	1
Fish and Wildlife	1
Critical Areas	3
Marine Resources	
Watersheds & Water Resources	5

Table 3.1 presents Resource Information Priorities base on commonalities of response and need. Table 3.2 presents the scale differences which were noted for these data types. (A system centralized common data types would therefore be required to meet differing requirements of scale and aggregation).

4.1.7.3 Summary of User Needs Assessment

It is generally accepted that geographic information systems should have user needs well defined and these needs should be primary determinants in system coverage and developments (Tomlinson, et. al. 1976). Although generally accepted, there are no easy methods to acquire information about user demand. In the past it has been often only assumed that large scale geographic systems sponsored by public agencies are designed to serve a broad range of users, and its assumed that the system products will be automatically desired and used.

Tomlinson et. al. point out that user demand surveys are difficult to conduct and almost impossible to interpret. The principle constraint is that user needs investigations require responses concerning proposed products rather than real products. In the future various methods of determining user needs must be used (such as interviews and questionairs). As well, once initiated a system program must contain an extensive user interraction program to explain data products to potential users, to provide users with prototype products and to receive from users an assessment of the utility of data products.

4.1.7.4 Required future activities

Three efforts in regard to user need assessments should become regular parts of activities associated with data and information system planning. These are:

1) continuing use of interviews and questionaires in user needs determinations
2) assessment of the need to directly link natural resource, socioeconomic and cadastral information.
3) The provision of example system outputs.

4.2 Assessment of Data Availability

4.2.1 Introduction

Some of the most critical decisions to be faced in developing spatial data information systems concern methods of data acquisition. For example, if there are no constraints on time or cost, most system developers would choose to collect their own primary data. However, numerous constraints generally require agencies, at least in some part, to rely on other sources for much of their data. Significant among these constraints are dollar costs of basic data collection and the fact that similar data are collected and maintained already by other agencies.

Use of secondary data has immediate appeal because of reduced data associated costs, but the requirements for data handling and updating over a long time may create large initially unseen costs. Generally though, information system development is feasible only with the extensive use of secondary data sources which may be supplemented with primary data only when necessary. Therefor, any strategy for implementing a geographical information system must include an assessment of existing data sources (federal, state, local, and private). Such assessments must be long term, systematic, inerative and thoroughly documented, because the continued existence of a useable system will depend on the quality and reliability of the sources of data. Three specific subparts of data planning need to be undertaken to satisfy this need: identification of available data; documentation of available data; and investigation of spatial referencing systems.

4.2.2 Role of a State Data Plan in the GIS Data Availability Assessment

As previously introduced in Chapter 3, the proposed state data collection planning activity and the continuing geographic information system study should proceed on a parallel course. This duality in developing the two systems should allow development of linking mechanisms for the three basic components of data planning - user needs assessment, data availability assessment, and provision for storage, retrieval, and distribution of information. As well, it is proposed that this data related effort could also involve providing planning reports on specific data types such as capability and suitability analysis utilizing GIS associated capabilities.

A major part of the data assessment effort will focus on inventorying resource information for Maine. A second major thrust derived from assessing user needs in terms of data handling will be the establishment of standards for new data collection. Scales, level of detail, accuracy, coding and formats will all be examined to meet the requirements of the greatest number of potential data users.

Federal collection activity will be monitored. The state would submit statements of priorities to the various federal collection agencies to provide maximum consistency with user needs and geographic information system requirements. Formal liaison would be established between the Maine Land and Water Resources Council and the various federal collection programs. All state and regional agencies proposed data collection activities would be reviewed. Agencies would not be required to be consistent with the plan but recommendations as to where it appeared that consistency could be achieved without extra cost would be integral in this effort.

4.2 .3 Data Identification

Introduction

This section presents a general review of the data available for state resource planning in Maine. Later sections include discussions of issues related to data acquisition, and characteristics of data that affect both their handling and use. Emphasis is on implications for data handling rather than on specifics of data use.

Resource management and planning have environmental, natural resource, social and economic dimensions. Diverse types and considerable quantities of information should be considered in order to review the current status, past trends, and possible future directions in these areas. Timeliness and availability of information are the key considerations since decisions are often made on the basis of information which is available, without waiting for collection of additional data that would be thought relevant.

Although numerous data exist, they are scattered through the files, libraries and information systems of federal, state, and local public and private organizations. The data are often difficult to find and often exist in formats which are not of direct utility.

Availability of data is limited both by the extent of federal and state data programs and by the capabilities of current information systems.

Related to this, states are facing massive problems in developing data bases for resource planning. Data series on land use, soils, and geology are not complete for Maine as is the case for much of the country. In summary, data generally are not available for the majority of tasks for resource management, and it is both appropriate and necessary to identify and review information sources, distribution programs and efforts aimed at data indexing. This effort is initiated as part of this work not to serve as a compilation of data sources, but to identify the agencies, programs and individuals who are responsible for decisions concerning the availability of natural resource data.

4.2.3.1. Federal Data Sources

4.2.3.1.1. Indexes of Natural Resource Information

There is no catalog or directory of all the data sources and series of either the federal government or any state government. There have been movements toward such compilations at various levels of government but these are incomplete and generally stop short of listing data series item by item. However, two types of efforts are of interest to resource planners which will be discussed here in detail. These are: efforts to compile directions of data collected by federal agencies;and the establishment of the National Cartographic Information Center.

Several federal departments have undertaken compilation directories to the data and information systems within their organizations. These are identified to establish contacts with the indexing efforts which will most likely play important roles in the evolution of a state data collection plan and allied information system efforts. The feasibility of establishing formal ties to important data indexing efforts will be examined as part of a continuing effort.

Environmental Data Index (ENDEX): ENDEX contains computer-searchable descriptions of interdisciplinary files of environmental data on many levels. Approximately eight large environmental data files may be searched through the ENDEX System. When these files are large, detailed inventories are also provided. Specifically, ENDEX has three major components: (1) descriptions of data collection efforts; (2) descriptions of data files; and (3) detailed inventories of large, commonly used files. An ENDEX data file description lists the types and volumes of parameters available, the methods used to measure them, when and where the data were collected, the sensors and platforms used, data formats, restrictions on data availability, publications in which the data may be found, whom to contact for further information, and the estimated cost of obtaining the data. Individual ENDEX data files descriptions will be updated every 2 years. ENDEX services and products include: (1) access to specialized indexes of environmental data, grouped by geographic areas, institutions, or disciplines; (2) on-line, interactive searches of the indexes to answer specific questions concerning the availability and whereabouts of data files; (3) a quick-response determination of the costs of retrieval from large data files; and (4) data catalogs from large NOAA environmental data collection projects.

Oceanic and Atmospheric Scientific Information System (OASIS): OASIS is a computerized information retrieval service that provides ready reference to the technical literature and to research environmental sciences and marine and coastal resources. It provides computerized searches of both NOAA abd non-NOAA data bases containing references to technical publications. Approximately 33 major environmental data bases may be searched through the OASIS system. OASIS offers access to major meteorological and oceanic bibliographic information files not available anywhere else in computer-searchable form.

USDA Data Inventory. Office of Information Systems, USDA. 1973-1974. 7 vols.

The USDA Data Inventory provides a concise listing and description of the data files used or produced by agencies within the Department. Each of the first six volumes contains the program data requirements for one or two of the missions of the Department. (A mission is a grouping of Department-wide goals that characterize the Department's role in solving broad, national problems. The 10 missions are comprehensive and include all activities of USDA.) These volumes are: (1) Agricultural Exports. Foreign Agricultural Development; (2) Rural Development. Encironmental Improvement and Resource Development and Use; (3) Support for Non-Federal Governments and Institutions. General Administration and Program Support; (4) Food and Nutrition. Consumer Services and Human Resource Development; (5) Agricultural Production and Marketing Efficiency; and (6) Farm Income. The seventh volume is a combined subject index which lists all subject terms contained in the preceding six volumes and the associated program data. Each of the first six volumes contains five sections in addition to the Introduction. Section II is the actual inventory of program data requirements reported by each agency as being used or produced in support of a program for which they have responsibility. They are listed alphabetically by title within a Department mission. Each entry in this inventory provides descriptive information about the data requirements as follows: data inventory number, title, the agency that reported use of the data, the time period to which the data relate, the accessibility of the data, the mode of processing currently being used, Program Data Requirement Identification number, subject index terms which describe the general subject matter contents of the requirement, an abstract describing the contents of the data included within the requirement, and the division, branch, or unit that can respond to questions concerning the data.

Automated Systems Inventory: All major automated systems (operational and developmental) within the Department of the Interior are described in the Automated Systems Inventory which is maintained by the Office of ADP Management.

Resource and Land Unvestigations of USDI has produced several federal agency wide data indexes and a 3,000 page draft directory of existing data within USDI. This work was completed in 1975 and is being updated in 1977. It will be available to the State Planning Office and other statewide users in Fall 1977.

National Cartographic Information Center: This center maintains a data system containing information on the availability of aerial and space images, maps, charts, geodetic data and related digital data produced by federal agencies. The center also maintains information regarding the status of on-going cartographic data collection efforts. A formal tie for continual updating of center activities is available through a quarterly newsletter which they publish.

Earth Resources Observation System Data Center: The EROS Data Center maintains, reproduces, and sells to the public digital and photographic data acquired by satellites and high altitude aircraft. Input is received from LANDSAT 1 and 2 as second generation film negatives, 16 mm microfilm of LANDSAT scenes and the NASA Skylab Earth Resources Experiment Package in various film formats. Also offered as standard products are 16 mm microfilm for the 470 color composite scenes for the World Reference System, 16 mm black and white microfilm of USGS photo indices, 16 mm microfilm of NASA aircraft missions in color and black and white, and 16 mm microfilm of the NASA Skylab missions in color and black and white. A number of states (i.e., Texas) have established direct computer ties with both the Cartographic Information Center and the EROS Center.

4.2.3.1.2. Federal Agency Data Bases and Data Systems

This section briefly describes federal data sources for spatial information and the system's the data bases are a part of. The important aspects of this information are not the data specifics presented but rather the agency programs, plans, and contacts which should all be regularly monitored to ensure continued availability of important listing data types and the timing on planned data collection and storage activities for planning future availability of data. Data systems of important are summarized, as well, for locational information and not for system specifications. Again, it is proposed that this data and activity monitoring can be best accomplished as part of a formal state data planning framework.

DEPARTMENT OF AGRICULTURE
Rural Development Office

This office is responsible for the development and administration of various farm oriented loan programs and for coordinating a nationwide rural development program utilizing the support services of a number of other federal agencies. The office holds no specific spatial data bases with coverage in Maine but it does operate an information system of potential interest. The Federal Assistance Programs Retrieval System provides data to local community rural development programs for how communities may meet basic eligibility requirements. All programs within the system are keyed to the Catalog of Domestic Assistance.

DEPARTMENT OF AGRICULTURE
Forest Service

The Forest Service is primarily responsible for the managment of federal forest reserves. As well however, the Service has responsibility for nationside leadership for forest managment and planning applications and research efforts. The service also cooperates with state and local governments, agencies and organizations, forest industries, and private land owners in the protection, reforestation, managment, and utilization of millions of acres of forest lands and lands associated with vital watershed protection areas.

Important Data Bases and Information Systems are:

1. An overwhelming majority of data bases held by the Forest Service pertain to the lands which they directly manage. Their single data base with complete national coverage in the National Forest Inventory. This in an extensive ten year incremental look at the forest related resources of the country. For every state a large data base consisting of a number of forest related parameters is collected, stored, and analized. Most of the data are geographically referenced and spatially mapped.

The last survey was completed in 1970. This information was collected and is now archived without the use of computers. Plans are tenative to automate the 1980 survey using an accepted Forest Service data managment system which is currently under development.

2, Although it actually holds little data of interest the Service has been in the forefront in the development of computerized spatial data handling capabilities. Presently because of the proliferation of such systems within the Forest Service(11 major systems), the Service is engaged in an effort aimed at taking the best aspects of this existing work and developing two standard systems for nationwide use. The two systems would include a grid mapping system and a polygon mapping system. Because the final design of the systems is nowhere completed a number of their important existing systems are briefly outlined because of their particular relevance to particular data handling problems.

A. Total Resource Information System(TRI)- is a natural resource data base information system. The system data base as it has been developed by the Forest Service is divided into a number of subsections covering data under the headings of recreational data, land cover data, seenic resource data, and aquatic and terrestrial resource data types. TRI is a basic grid cell system. It can perform statistical computations such as cross tabulation but it cannot graphically overlay two different data sets for the same geographical area so from the graphics point of view its usefullness in resource planning is limited.

B. Polygon Layer Overlay Technique(PLOT)-is a polygon line system which stores coordinates for polygons and the identifiers which describe them. The principle utility of the PLOT system is in the overlay of at least two different mapped data sets creating a new polygon map from the composite of the maps overlaid. A number of statistical routines and mapping variations can be applied to both the source maps and the composites. PLOT is also a useful system in that it can accept input at any scale and can process and extract data for overlay interactively.

C.LIM, R3MAP, and MIADS -are basic grid mapping systems. These systems are of utility if speed and ease of application are required and visual clarity of the maps, fidelity of area boundaries, and the precise evaluation of areas may be sacrificed to varying degrees. All three systems are structured in that all only have line printer output, not scale change for data inputting and few interactive capabilities.

D. The Geographic Locator System(GELO)-is a line segment system where each input polygon is broken down into line segments. Each point along a line segment must be entered into the data base with a right and light identifier code to identify polygons for analysis and retrieval.

The data processing features of the system are:(1)store and plot back maps from any resource data type;(2)overlay one resource map on another and produce composite maps;(3)join adjacent maps to create different data coverages for single analyses;(4)calculate and report for a number of statistical parameters;and (5) produce master catalogs of all spatial data files in the system. Geographic referencing is by latitude and longitude and scale change for inputting and outputting data is also a basic capability of the system.

DEPARTMENT OF AGRICULTURE
Soil Conservation Service

The Soil Conservation Service is engaged in a number of activities and programs which involve collecting, storing, analyzing and distributing vast amounts of spatially referenced data. Primary amont these are the National Cooperative Soil Survey, The Conservation Needs Inventory, Flood Hazard STudies, Inventories for Resource Planning, Land Inventory and Monitoring, The Storage and Retrieval System for Soils, Woodland and Range Data, Snow surveys and water supply forecasting, and their watershed and river basin planning programs.

1. The National Cooperative Soil Survey-is generally composed of both general and detailed soil surveying activities which are carried out within state field offices.

2. The Conservation Needs Inventory-provides soil, water, and land use statistics for the nations land. The inventory was carried out in the late 60'd and early 70's using a random statistical sample that ranged from 2% to 10% of the total area of the various states. The system is automated and summaries and unaggregated data are available. The survey is currently being updated(1977-1978) under the new title of "The National Erosion and Sediment Survey."

3. Snow survey and rainfall data are collected and archived by state sources.

4. Iowa State Statistical Laboratory Data System-contains soil names and descriptions for 95% of the soils of the United States. The SCS Cartographic Office in Hyattsville, Md. along with the Remote Sensing Task Force and the Resource and Managment Information System Task Force are developing an Advanced Mapping System which is utilized for automated mapping of soils information. As well, efforts are underway to archive a digital soils data base utilizing the systems capabilities. As well interpretative capability mapping is part of the system. The system currently relies on manual digitizing and raster scanning for data inputting and a Gerber Plotter for output mapping. The system at the moment is primarily being utilized in the Important Farmland Mapping Program.

In terms of long range development plans SCS is planning the development of a regional office accessib le data system which would be required to interface SCS internally generated data with the topographic, geologic, and socio-economic data of other federal agencies. It is projected that the system will be completely interactive in the regional offices. The system will also directly interface with a number of aerial imagery sources. Direct picture processing and image incorporation into the data base is a projected capability. Data banks of the most currentl LANDSAT imagery and the most current without cloud cover will be included and readily retrieveable. The purpose of this system is to permit SCS planning personel in the field offices to be able to model the landscape overlaying and compositing soils informaltion with land cover, land use, and topographic data files. No specifics of the system are currently available as it is still in the early stages of planning and development.

Department of Agriculture contacts are:

Mr. R. Bacelius-Remote Sensing Task Force and Land Inventory and Montoring Program of the Soil Conservation Service

Mr. Paul Holm-U.S.D.A. Office of Information Systems

Mr. James Lables-U.S.D.A Office of Information Systems

Mr. John Kennedy-Office of Data Managment of the U.S.D.A. Forest Service

DEPARTMENT OF COMMERCE
Bureau of Census

Geography Division

The Geography Division of the Bureau prepares maps and materials for use by the other data collection and analysis of the Bureau. For all statistical census data types(i.e.,census of agriculture,economics,population,ect.,)geographic referencing files are available and such files have been incorporated into both federal and stateside geographic information systems.

Spatially based census files currently held are:

1)Master Enumeration District List-which summarizes the codes which identify the geographic areas presented in the 1970 census files. Counties,census tracts, and SMSA's are all coded for eventual cross referencing.

2)GBE/DIME Files which are street inventories of the urbanized portions of SMSA's relating address ranges to the various geographic codes.

3)Zip/Tract Cross Reference File is a correspondencefile between zip codes and census tract codes.

4)DIMECO-contains latitude and longitude coordinates describing the boundaries of all counties in the United States.

5)ADMATCH-is a computerized address matching system for relating local area descriptive data files with the geographic referencing files such as GBE/DIME.

6)GRIDS-is a computer mapping program for displaying grid oriented data.This is the primary program for graphically presenting ADMATCH analysis information and composite maps.

DEPARTMENT OF COMMERCE
Economic Development Administration

Business Research and Analysis Office

This office in conjunction with the Water Resources Council and Regional River Basin Commissions has developed a water use forecasting information system. This system "The National Water Assessment Model" has four primary components. First are base use estimates for all industries per EDA district. Second, future water use practices and economic growth are modelled for each industrial type for each region. Third, forecasts are made using the previously derived estimates and finally forecasts are summarized to produce composite regional water use totals for the industrial sector which are analyzed and displayed both in tabular and graphic formats.

Important data files include:

1. Actual monitoring of industrial water use for a national sample of 10,000 industrial plants nationwide.

2. Regional economic pattern summaries derived from other census information sources.

3. Technical hydrologic parameters such as recirculation rates, consumption and ground water base flow rates and surface storage capacities.

All data and summary reporting are disaggregated only to a subdivided major river basin. Modelling is done at this level but data is available for individual industrial plants.

Geography Program

As EDA's primary function is long range economic development planning focusing on areas with severe unemployment and low income this program is designed to tie the Administration's socio-economic data to a spatial component for information storage and analysis.

Important data files and systems are:

1)The Industrial Location and Retrieval System-which holds compilations of community resource profiles derived from data sources generally originating with regional and state planning and economic development agencies. Also a industrial-manurfacturing census which is spatially references and retrievable, an industrial location modelling system and a summary of different state geocode systems are included.

2)Composite Mapping System-has the capability for merging numerous sets of digitized data (Grid and polygon but analysis is done only on grid data files) with weighted combinations for composite socio-economic characteristic mapping. The system currently stores data at any scale in a 120 by 120 array of grid cells. As currently used by EDA each grid cell corresponds to about four square miles which is suitable for statewide and multistate data analysis problems. Some standard analysis outputs of the system are: zones of accessibility to linear or point facilities (highways or refineries for example), isopleth analysis of continuously variable data such as climatic parameters and locational mapping depicting a range of socio-economic and land use characteristics.

National Oceanic and Atmospheric Administration

National Environmental Satellite Service

This service provides satellite observations of the environment by establishing and operating a national satellite system and it conducts an integrated program of research and services related to oceans and inland waters and the lower and upper atmosphere.

Primary data files include:

1)Varying nationwide satellite coverages stored digitally in recently developed mass storage systems.

2)Environmental modelling and composite computer mapping projects of which coverage is very localized and spotty depending on research priorities and cooperative agreements.

Extensive satellite data for all states will be digitally archived and regularly updated for public distribution.

National Marine Fisheries Service

This service conducts surveys of the living resources of the oceans, analyzing in particular, economic aspects of fisheries operations with an eye on improving mans ability to use and conserve these resources. The agency is also examining alternatives to ocean dumping and provides leadership in promoting wise and balanced managment of the coastal zone.

Primary data files are archived in the National Oceanographic Data Center which is summarized at the end of this section.

National Ocean Survey

The National Ocean Survey has responsibility for producing nautical and aeronautical charts. It conducts surveys and prepares charts of the harbors and coastal and offshore waters of the United States. NOS currently has three major computer mapping projects underway or planned.
1)development of a completely automated nautical system by 1980. System will include data acquisition, storage, production and distribution.
2)Automation of the production of aeronautical charts with a complete data system by 1980.
3)Establishment of a digital data base for geodetic control and North American datum by 1979.

They are on a day to day basis producing digital maps of cloud cover, daily weather maps, aeronautical charts, and specialized mapping projects related to the continental shelf, defining state boundaries more accurrately and metric conversion studies in computer mapping.

National Weather Service

This service within NOAA is responsible for reporting the weather of the United States and its posessions and provides weather forecasts to the general public, issues warnings against destructive natural events such as hurricanes, tornadoes and floods. It provides special service is support of aviation, marine activities, agriculture, forestry and urban air quality control.

Within Maine the service has approximately fifty reporting weather stations which are supplying data readily being utilized by a number of state agencies as inputs into their planning and managment activities.

Important data files held by the service are:

1)Annual compendium of daily river stages.

2)Local and national climatic summaries.

3)Monthly storm damage data.

4)Crop-climate correllation summaries.

5)Annual statistics on climatic extremes.

6)Specialized data on various hydrologic and climatologic parameters with only local coverages.

Much of the data distribution, storage and archiving by various NOAA divisions and services are through agency environmental data services which are summarized below.

National Climatic Data Center-maintains nationala rchives on weather data collected since 1841. Allmaterials are computer archives on microfilm and magnetic tape.

National Oceanographic Data Center-maintains the computerized National Marine Data Inventory which stores the records of thousands of oceanographic cruises and projects. Data are stored so they can be retrieved by geographic limits, the kind of platform involved, by country, state, county, dates, institute and subject. All material is reproduceable for users on tape, punched cards, microfilm, or printed hard copy.

National Geophysical and Solar-Terrestrial Data Center-deals with data related to the ionosphere, variations in the earth's magnetic fields, solar activity, cosmic rays, and marine geology and geophysics. Computerized indexes are held for such things as available photographs of the sea bottom by depth, location, and time, and punched card and tape records of undersea explorations. There is also a computerized annotated bibliography of undersea geologic sampling efforts. Tape stored records of seismographs from stations around the world are held with the magnitude and location of all earthquakes and follow up data indexed for retrieval. In addition, a complete geophysical data base for use in large scale land and ocean surveying and mapping is held.

Agency Contacts

general:

1. NOAA(general)-Robert Gelfeld who is in charge of the ENDEX indexing system.
2. NGDC-Daniel McGuire

specific:

1. The National Water Assessment Model-John Klein ADP Operations
2. EDA Geography Program-David Portch-EDA Systems Division
3. The Environmental Data Centers-Thomas Austin
4. Geography Program of Census Bureau-Larry Caubaugh-Data Users Service

DEPARTMENT OF DEFENSE
Corps of Engineers

The Corps is primarily responsible for major Federal water resource development activity which involve engineering work such as construction of dams,reservoirs harbors,waterways and locks.This work is generally intended to provide flood protection fpr cities and major river valleys,reduce the cost of t5ansportation, supply water for industrial and municipal use, orovide recreational ac facilities for the nation,regulate use of the rivers and protect the shores of the oceans and lakes.The Corps also provides planning assistance to states and other non-federal entities.Most data associated with these activities are collected and archived specifically to an individual project orientation.

Primary data files are:

1.The Construction Engineering Research Laboratory Data System contains extensive data with nationwide coverage on rare and endangered aquatic and terrestrial species of insects,plants and animals.

2.The Fort Belvoir System for coastal wave dymanics analysis contains coverage on a national basis.

3.The nationwide flood plain map data file is a manual graphic and statistical data file.

4.Environmental Reconaissance Inventory-is a manual system for storing and distributing 1:250,000 scale maps on land and water resources,forest types and wildlife habitats, historic sites,lanhdscape enhancement and reclamation potentials.Four states(not including Maine) have been done on a pilot basis and coverage for Maine is planned but has not been confirmed.

5.System of Information Retrieval and Analysis-contains a national master file with regional breakdowns on agricultural census data ,population and housing ,manurfacturing locations, commercial patterns ,use of public recreation facilities,and environmentallly sensitive wildlife areas.

This data is maintained centrally at the Laurance Berkeley Laboratory and is accessed remotely by regional Corps offices.

6. Hydrologic Engineering Center-contains automated procedures and data files for hydrologic engineering and planning applications, flood hydrograph data, water surface profiles reservoir system analysis, streamflow simulation, rating and flood routing, basin rainfall and snowmelt computation and urban storm runoff. The center actually maintains little region or statewide data but uses the system for specialized hydrologic analysis.

7. Water Control Managment System -is an automated data system with nationwide coverage with an emphasis on hydrometric data and river basin modelling utilizing automated satellite information (LANDSAT).

Agency contacts are:

Robert Thompsom--Fort Belvoir
C.P. Marks --Waltham Regional Office for New England

DEPARTMENT OF HOUSING AND
URBAN DEVELOPMENT

Division of Housing Managment

Primary programs of H.U.D. are New Communities, Community Planning and Development, Housing Production, Housing Managment, and Federal Insurance, Primary data banks relate to open space, residential and public facilities, water and sewer systems, urban rehabilitation, and federal insurance statistics. Much of this data also resides in an aggregated automated form within the Technical Services Division of the Maine State Planning Office.

H.U.D. does utilize a computerized information system with mapping capabilities for data bases related to subsidized housing unit types and locations across the country. Data in this system desrcibe projects according to census tract, latitude and longitude, number of units the each project, numbers of bedrooms, occupancy rates, and the specific types of program funding associated with the units. Reference for future planned and developing H.U.D. data systems can be found in HUD Long Range Plans for Data Automation 1975.

Agency contact:

Horace Bussell--Office of Managment Information

DEPARTMENT OF INTERIOR
Fish and Wildlife Service

The primary mission of the Fish and Wildlife Service, which is responsible for wild birds, mammals, inland sport fisheries, and specific fisheries research activities is the assurance of maximum benefit from wildlife resources for the American people. Biologic monitoring, surveillance of pesticides, heavy metals, thermal pollution, studies in fish and wildlife populations, ecological studies, environmental impact assessment through river basin studies, stream channelization, dredge and fill permits, and impact statement review, and area basin planning of river basins and watersheds are the primary activities of the Service.

Man of the data files held by the service are also held within the individual states. A number of files are associated with data systems in Washington.

1. The Migratory Bird and Habitat Research Lan maintains a nationwide computerized data base on waterfowl and inland game birds. Banding, recovery, and survey data from general statewide and region specific studies are utilized to compute recivery rates which are utilized in the planning of a number of hunting and harvesting associated parameters.

2. Waterfowl Survey System involves a large number of computer files and analysis routines developed to analyze quantitative information on waterfowl population ecology. The system is utilized by a number of bureaus within the service in the preparation of environmental impact statements.

3. National Wetland Inventory-will be undertaken in Maine 1978- 1979. The survey will result in a digital and map data based on wetland types and locations.

4. Habitat Evaluation Information System is a projected automated data base and system which will contain information on coastal aquatic and terrestrial animal populations and habitat characteristics.

5. A final significant data base is projected to result from a U.S. Fish and Wildlife Ecologic Characterization of the Maine Coast (Biological Services Office). The relevant part of this study will include both graphic atlases of mapped information and a digital data base.

Data collection activity will be documenting data on coastal ecology, physical characteristics, pollutant residuals, land use parameters and economic data. This information will all be indixed and the information will be available. The collection of data is being carried on in three directions. Federal, state, andregional sources are being contacted to collect data which has already been assembled within these sources to identify major data gaps. There will also be an attempt to identify obscure data sources (i.e. scientific research, thesises ect.).

Energy Resources Company the primary contractor, utilized computerized analysis procedures in a majority of their work. Availability of this data will depend on their data analysis strategy. They do project the availability of the following data types in digital formats: physiography, geology, groundwater, marsh habitats, upland habitats, land activity, soils, discharge sources, critical and unique areas, lobster and clam areas, endangered species, hazard areas, climatology, land ownership and archeologic and historic sites. Digital format has not yet been finalized.

Primary contacts are:

Warren Blandon- Migratory Bird and Habitat Research Lab

Ron Beck-Energy Resources Company

Ralph Andrews-Fish and Wildlife-Regional Boston Office

DEPARTMENT OF THE INTERIOR
Geologic Survey

Conservation Division

Current responsibilities of the Conservation Division do not generally require the storage or distribution of extensive spatial data bases. It is projected, however, that the division will be holding extensive amounts of spatial data generated by other divisions and departments because of a long term significant commitment to expand mineral resource estimation programs and production monitoring activities.

Important data bases with national coveraged currently held within the division are:

Well History Files(large sampling for all states)
Pipeline Data Base
Field Reservoir Reserve Estimation System(all significant mineral resource reserves)

Land Information and Analysis Office

This office exists as part of the directors office rather than as a separate division. The primary programs of the office are: Earth Resources Observations Systems; The Geography Program; Earth Resources Applications; Resource and Land Investigations; and The Environmental Impact Analysis Program.

The most important data base held within the office is associated with the Geographic Information Retrieval and Analysis System of the Geography Program. A nationwide data base for land cover, political units(county boundaries), watersheds, census subdivisions and federal land ownership is the basis for the system. Mapping scales are 1:100,000 and 1:250,000. All maps will be in polygon format. 1983 is the projected completion date for the entire country and digital tapes will be available as areas, regions and states are completed.

Cover types and land use compilations are based on the classification scheme and definitions for level 2 land use and cover as outlined in Geologic Survey Professional Paper 964. Political boundaries will only include county and state boundaries. The census county subdivisions map provides a graphic depiction of Census Tracts in SMSA's and MCD's. The Hydrologic Unit map provides a geographic reference base for all statistical data held within the Water Resources Division of the Survey. Federal Ownership (all sources) and possibly State Ownership will be the final base map for the system. Cooperative federal-state agreements can be undertaken and the information will be then available after 1 to 1 1/2 years.

Two other data collection and storage efforts are of potential interest to state data planning and information system efforts. The first is the "Image Data Inquiry System" which is held cooperatively with the NCIC and it provides access to indexing for over 6.5 million frames of aerial photography. Secondly is the "Cartographic on-line interractive digitizing and display editing system" which is involved in research into problems associated with creating large digital data bases. The primary focus is currently working on the measurement of potential data volumes and the impact of these volumes on digitizing methodologies.

The programs of this office are generally coordination oriented so there are few actual data bases held but rather the division is involved in developing protypic data bases and data handling systems.

Geologic Division

The activities described under the Geologic Division are primarily oriented toward research and development. Although many of the associated data bases are extensive most of the divisions activities require or result in point geologic data. Most of the data bases are oriented toward solving rather specific problems -usually at the project level.

However the division is working with the Bureau of Land Managment on methods to better reference the location of minor civil divisions in large scale computer oriented mapping. The division is also at work on a system which will allow the retrieval of geologic point data by political boundaries and deed descriptions.

Primary data files of the division include:

Radiometric Age Data Bank
Paleomagnetic Data Bank
Earthquake Information System
Geothermal Data Bank
Digital landslide suseptibility determination.
Oil and Gas Files
Geomagnetic Analysis
Rock Analysis Storage System
Computerized Resource Information Bank
Geologic Synopsis Program
Computer Composite Mapping(detailed mapping done by special agreement only)
Applications of Computer Cellular Mapping to Planning(research oriented Geographic System)

Topographic Division

The primary objective of this division is the provision of basic cartographic data for the United States as required by other federal agencies, state and local organizations and the general public. The division is concerned chiefly with the production of any map series which requires national coverage or national mapping standards. Topographic and planimetric maps of the division include:

1)the standard mapping series(non-digital)-1:24,000; 1:62,500; and 1:250,000 which by far have been the tradition focus of the program.

2)a variety of feature and color separates based on the above base information.

3)a new 1:100,000 intermediate scale which is currently only planimetric but is projected to be extended to include topographic information.

4)1:50,000 and 1:100,000 county maps.

The division only on special agreement will provide other cartographic products derived from existing data bases-slope class mapping, aspect mapping and cultural feature mapping.

The primary existing computerized data bases of the Topographic Division are:

1. Computer generated shaded relief mapping
2. Aerial Photography Quadrangle File System
3. Geographic Names Information System

Perhaps most importantly is the recognition of need by this division for a national cartographic data base in digital form. This recognition is for both topographic and planimetric data types. The objectives of this proposed "National Cartographic Data Base " focus on the provision of appropriately coded clean data types for a number of different data types and parameters.

The data currently projected to be included are:

1. geographic referencing-coordinate base maps
2. hydrography-streams and rivers, lakes and ponds, wetlands, reservoirs and shorelines
3. boundaries-political jurisdictions, national parks and forests, military lands and similiar entities.
4. rectangular survey system-the public lands survey network
5. transportation systems-roads, railroads, trails, canals, pipelines, transmission lines, bridges, and tunnels.
6. hypsography-contours, slopes and elevations
7. geodetic controls-monuments, markers, landmark structures
8. geographic names
9. all man made structures
10. vegetation

The plan is to have a prototype system in operation by 1980 and the data base completed for the entire country within another 10 years (1990).

The data elements identified for inclusion in the system will be separated into two categories:

1. digital terrain data files

2. planimetric digital cartographic files

This separation is a function of data acquisition methods. The data base will use the 7 1/2 minute topographic quadrangle series of maps as the basic working unit for data acquisition, entry and storage. Present plans are to use stereographic photogrammetric models for all areas unmapped at the time of potential data entry.

The digital elevation data will be recorded as one file per 1:24,000 scale quadrangle on magnetic tape. Within each elevation file, data are to be arranged in sequential logical records. Three types of record formats are utilized. The first contains general characteristic information about the digital model including quadrangle name, latitude and longitude of the lower left corner and UTM coordinates. The second is a variable length file used for the actual elevations. The x and y coordinates are identified and the location of the remaining points can be computed or estimated from pattern and spacing of points, rows and columns, and visual checking of the elevations against graphic source materials. All elevations are in feet. The last record contains the number of elevations for the complete topographic map record.

The planimetric files will contain information derived by digitization. This will be done by manual digitizing, interractive line following and possibly raster scanning. Planimetric cartographic files are stored on magnetic tape and consist of two main sections in the proposed archival format. The first section contains header information for the file among which is a count of the number of overlays in the file. The second section is divided into a number of subsections, each subsection for each independant network covering each cartographic data type.

The header section contains three logical records. The first includes the name of the map, local x and y coordinates for the four corners of the quadrange sheet, map projection codes and the latitude and longitude for the southwest corner of the quadrangle sheet. The next subsection contains the number of networks, lines, nodes and areas. Every line segment on the map is coded with a starting node, ending node and left and right parcel identifiers. All coordinate values are to .001 inches. The third section consists of a number of parameters which are important for internal data base managment routines.

The graphic section consists of node, area and line elements. The node list is fixed in length and contains the major and minor fields for the particular attributes being addressed. The second list contains positional coordinates which define the shape of the particular attributes. The third data list is reserved for descriptive texts.

Water Resources Division

The Water Resources Division is responsible for the Survey's appraisal of water resources and for research in hydrology related to the occurrence, distribution and quality of surface and ground water. It is responsible for the coordination of water data collection activities of all federal agencies and for the design and maintenance of data collection networks and for providing information on the availability of water related data. This particular requirement is met by the previously discussed National Water Data Exchange(NAWDEX).

Water data collected from hydrologic data stations, areal hydrologic information, and various types of interpretive studies are stored in the divisions National Water Data Storage and Retrieval System (WATSTORE). In addition the division is performing digital processing of remote sensing data for a number of purposes-predominately in a national wetlands mapping program, spatially mapping relationships between water quality data and land use change and in the assessment of various types of mining activity on water quality. Specific data files associated with WATSTORE include: streamflow data; reservoir contents; river stages, water temperatures; conductance; sediment concentrations; sediment discharges; annual maximum stream flows and river stage data.

Agency contacts:

1. Geography Program-Eric Anderson, Bill Mitchell,
2. RALI Program-Olaf Kays, Ethan Smith,
3. LUDA Program-Richard Witmer
4. National Cartographic Information Center-John Swenmorton
5. EROS Data Center- Allen Watkins

DEPARTMENT OF ENERGY
(ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION)

The mission of the Energy Research and Development Administration is to consolidate federal activities relating to research and development of various sources of energy in order to increase efficiency and reliability in the use of all energy resources. As well ,an idealized long terms goal is to make the nation self succificent in energy and to advance the goals of restoring, protecting and enhancing environmental quality and to insure public health and safety.

Office of Environmental Information Systems

This branch of ERDA has the responsibility of initiating and coordinating research efforts into relationshio s between energy planning and comprehensive environmental planning. The major thrust of the effort(conducted within the eight national research labs) is focused on developing environmental data bases and automated data handling capabilities for utility in compreshensive energy-environmental planning. The major systems and data bases of relevance are:
(Specific systems are not detailed but information on them is being compiled by the State Planning Office)

1. Brookhaven National Laboratory--construction of an automated system for modelling relationships between land use and air, water and noise pollution. There are extensive data bases on air and water pollution for Maine much of which was initially gathered by in-state sources.
2. Laurance Berkeley Laboratory-contains no relevant data bases but is involved in extensive research and development in interactive compyuter graphics and grid based landscape analysis systems.
3. Pacific Northwest Labs-are involved in the development of a polygon mapping system.
4. Los Alamos Scientific Laboratory is involved in extensive studies incorporating demographic analysis procedures into spatially referenced information systems. As well, the direct processing of satellite da ta is a focus of this labs research.

Agency Contact: T. M. Miller, Office of Environmental Information Systems

ENVIRONMENTAL PROTECTION AGENCY

The EPA has programs in air and waste managment, water and hazardous materials, legal enforcement of environmental regulations and research and development. EPA coordinates and supports research and antipollution activities by state and local governments and reinforces efforts within the federal government aimed at accounting for imact impact assessment in their various operations. It also makes public its written comments on environmental impact statements.

Most data bases held by EPA originate from state by state reporting sources, so most data types already reside instate. Information systems are listed because in some instances data may be accessed more rapidly from EPA than by the compilation of instate data which is archived there.

Data systems with national coverage are:

1. Standards and Regulations Information System
2. Pesticide Enforcement Managment System
3. Discharge Compliance Data System
4. STORET
5. Pestidies Registration System
6. Pesticides Analysis and Control System
7. Inventory of Public Water Supplies
8. Storage and Retrieval of
9. Survey of Municipal Waste Water Treatment Needs
10. National Emissions Data System
11. Air Pollution Information System
12. Energy Data System
13. Solid Waste Information System
14. Environmental Assessment System
15. Noise File
16. Population Studies System
17. Models for Fresh Water Ecosystems
18. Eutrophication Study System

Agency contact: Office of Environmental Information Systems EPA.

4.2.3.2 Identification of State Data Sources

The purposes of this effort should be numerous. First it should present a baseline assessment of the data which are currently and projected to be available from state sources. This permits both an assessment of data availability from state sources but also function to measure met and unmet data needs when data available is examined along with the user needs assessment.

It must be reiterated that data availability assessment is viewed as a continuing process and the information presented here is intended to function only as initial baseline information.

Currently there are no complete indexes of natural resource information for the state of Maine. However, TRIGOM (The Research Institute of the Gulf of Maine) has been instrumental in the compilation of indexes for specific data types and in an effort currently in early planning stages to develop a complete interactive automated resource data index for the state. Previous efforts have included the KWIC Index of socio-economic and environmental data and a Maine Rivers Bibliography. Based on information and experience gained while preparing the New England Endex, TRIGOM has proposed the development of a complete data indexing system. This system will allow any data user needing references to documents or files to go to a computer terminal where he can search a bibliography using specific key words. Modules will be developed separating the complete data base for simple, fast, and economical data searches. The modules will separate the data base by data type (i.e., map vs. tabular), data coverage (i.e., marine resources), and geographic coding (statewide, local specific). It is projected that if implemented, the Maine Index will form a cornerstone in the data planning and geographical information system efforts.

A second index of primarily spatial or map information will also be prepared for Maine during the summer and fall of 1977 by the Research Institute of the Gulf of Maine. Both hard copy and interactive automated retrieval are planned. This index is being designed according to a similar index recently completed for the State of Wisconsin. All spatial information with coverage in the

State of Maine will be documented according to the following parameters: data type, dates of collection and publication, detail, aggregation, quality, timeliness, sources, lengthy description and contacts for information and acquisition. It is planned that once the base is established it will be updated annually as part of a state data planning effort. Once completed, the GIS feasibility effort could utilize this information to develop reports of various data sets and how these data relate to problems in establishing a spatial digital data base. (For example, data documentation and evaluation for specific data types.)

4.2.4 Data Documentation

As data availability assessment continues, based on data types identified by user needs assessment and/or from system design specifications, the characteristics of pertinent data should be fully explored to determine potential utility to users within the system. Data residing in specific agencies being utilized only by those agencies with no other apparent users need not be evaluated as part of this process. A systems central available data base is intended for information and data types for which either a number of agencies have shown a need or for single agency which does not have the expertise or resources to analyze the data in house (L.U.D.A. Cover Mapping).

An effort involving the construction of a matrix (Figure 4.1, which is an example from the Texas Natural Resource Information System) which delineates data flows within Maine could be easily accomplished and would yield a concise picture of data use and sources for the state. State Planning Office Technical Services Division and TRIGOM data indexing makes such a matrix easy to assemble for review by the various user agencies.

4.2.4.1. Documenting Existing Non-Automated Spatial Data

If the specifics of data collection mode and resolution of data are acceptable for utility spatially referenced data may contain several types of errors. Prudent use of data indicates systematic

SUMMARY OF NRIS DATA/INFORMATION NEEDS

AGENCY	NRIS FILE NAME	USER AGENCIES														TOTAL	RANK
		TWRC	TDA	TSWC	TRRC	TSDH	TPWD	THD	TWQB	TIC	GLO	TFS	ACB	BEG	TWDB		
I. GENERAL LAND OFFICE																	
	Submerged Lease Data File				3		3		6		3			6	6	83	
	Rental Master File				6		3				2			6		87	
	School Land Master File		5		5		6				2	6		5		85	
	Veteran's Land Master File		5		5		6				1					87	
	Veteran's Land Statistical File		5				6				3				6	90	
	Unsold School Land File						3				3	6				89	
	Land Classification Master File		3				5		6		3				5	85	
II. TEXAS FOREST SERVICE																	
	TFS Woodland Management Activities		3				3	6				3		6	6	83	
	TFS Southern Pine Beetle Operations		2				3					1				83	
	TFS Wildfire Reports						4					3			5	89	
	TFS Fire Weather Reports		5		6		2				3	3		6		81	19
III. TEXAS HIGHWAY DEPARTMENT																	
	Air Pollution Data		6		5		2	3	6				2			80	18
	Bridge Log			6			2	2	6					4	3	79	17
	Structure Inventory and Appraisal			6			6	2						5		89	
	Road Life Data		6		6		2	3					4			84	
	Automatic Traffic Recorder						4	1		6			3		6	83	
	Truck Weight Data							3					4		6	90	
	Speed Survey						5	2		6			4			87	
	Origin-Destination Survey		5				4	4		6	3		3		6	80	18
	Roadway Characteristics, State System		5				2	3	6	6						85	
	Roadway Characteristics, County System		5				2	3	6	6						85	
	Traffic Log		4			6	3	3		6			3		6	80	18
	Rail-Highway Grade Crossings						4	3								91	

Figure 4.1 Example Data Needs Summary for the Texas NRIS

(source: Texas NRIS Report 1974)

data checking which should uncover errors if they do exist. Several errors and methods for documentation are presented below.

A. Horizontal accuracy - because of time and economic constraints only relative determination of positional accuracy is generally feasible. For example, a simple visual comparison with a separate related data set at the same scale is a good test of positional accuracy. Another simple method involves duplicate compilations for limited sample areas using different methods or draftsmen. Another method is to relate cartographic representations to aerial photographic images to obtain checks by coordinate comparison. As well, it is often valuable to coorelate information found on maps of differing scales.

B. Map contents - should be checked because data items may be classified incorrectly or ground conditions may change between the time of data compilation and time of projected use. Generally, any information checking must depend on a sampling scheme. Random sampling and ordered sampling are two such techniques. It is noted here that many map products which will be available have such determinations of content accuracy associated with them. Variance and standard derivation are used to quantify the validity of the data.

C. Inaccuracies of base changes - is important because enlargement of map scale, perhaps from 1:250,000 to 1:100,000, enlarges line data and aggravates the displacement of any features. An acceptable map at a given smale may be deficient at another scale. As well, when maps are reduced a different set of problems result. Excessive detail often clutters the new map; information may be repetitious and difficult to read; and places where adjacent records meet may be disjointed.

4.2.4.2. Documentation for Automated Spatial Data Files

Data may be acquired and recormatted into a number of formats and investigation is usually required to determine which format or mixtures of formats will provide the best possible data base. The task for computer oriented data is one of assessing the quality and quantity of information and determining the sata transfer processes that will permit computer storage and manipulation of data.

The data format, record format, and volume of the data to be handled are the prime determinants of any data handling capability and generally have a great impact on computerized data handling techniques. If the data is not directly useable and understandable at the point of acquisition some initial processing will be required.

The various classification schemes used for computer storage must be examined. Verification includes checking the logic of particular systems to determine if the characteristics which have been encoded can be perceived on source documents and maps.

In terms of recoding and formatting choices must be made among the various coding methods because the potential classification of data manipulation routines and file structuring are strongly impacted by coding and format.

4.2.5 Spatial Referencing

A separate study is required of available spatial reference systems which can be either implemented as a standard for an information system or used with available data that requires locational identifiers. This task must examine and evaluate base maps, coordinate reference systems, and other methods of identification such as street address systems.

This step is critical because there exists an unsolved problem of designing a spatial referencing system which can operate efficiently at both the regional and state leval and the urban development level. The problem must be addresses and it is likely that a resulting system will be multi0format capable of manipulating and relating data encloded in different formats and different scales. The interface between urban development and typical rural uses is generally of most interest to planners and it is accurately portraying this interface which is of great difficulty in spatial referencing systems.

4.2.6. Data Availability Summary

For Maine a preliminary review of federally available data, the status of current in-state data acquisition, and the characteristics associated with those data suggests that:

A. Data relevant to the needs of users exists in a number of media, aerial photographs, maps, tabular summaries, and computer storage devices.

B. Few efforts (i.e., the coastal program) have begun to pull these data into a consistent format and to make the data available in a coordinated way.

C. Data series (i.e., soils) are available in varying degrees of completeness and timeliness.

D. Perceived data needs differ among agencies with even similar responsibilities.

E. General standard analytical framework do not exist to guide collection of data for state resource planning programs.

F. For federal data sources, most series are currently incomplete for Maine or are available at smaller than desired scales.

The data formats and definitions of individual series (i.e., surficial geologic mapping) have not been related to needs for planning and decision making being based instead on scientific and technical criteria.

SECTION V

COMPUTER HARDWARE AND SOFTWARE AVAILABILITY

Chapter V

Assessment of Software, Hardware, and Manpower Resources

5.1

5.1 Introduction

Investigating existing software, hardware, and manpower resources which are currently available, or which will be available is required to develop a listing of resources for system development, operation, and maintenance. This step is not to select equipment or software or to determine personnel requirements but rather what is or will be available, what it costs, and how various components may be obtained. This assessment is vital for two basic reasons:

1) The equipment market is highly competitive and vendors frequently make claims which are difficult to substantiate, and

2) Necessary funds must be available for the time when equipment or software will be required and personnel hired.

5.2 Survey of Software

When data manipulation, analysis, and output requirements are defined the available software must be located and thoroughly documented. The software survey must also list hardware requirements for each software item. The question of transferability must be explored in detail. Within this effort important software sources have been identified and information collection is and will continue to be underway. Some of these sources include:

Federal Agencies

USDA Forest Service

USDI Geologic Survey

Energy Research and Development Administration

States

Federation of Rocky Mountain States

New York

Maryland

Montana

Hawaii

Alabama

Arkansas

North Carolina

The Canadian Geographic Information System Research

University of Iowa

Simon Fraser University

New York University at Buffalo

University of Saskatchewan

The National Labs

Marshall Space Flight Center of NASA

As well, the effort will procure and evaluate a number of software compilation and indexing efforts:

1) International Geographic Union's 800 page compendium of computer grapic software. (Hard copy available winter 1978)

2) New York State Geologic Survey is currently engaged in a comparative geographic information system study examining a number of statewide systems. (1978)

3) American Society of Planning Officials is evaluating the comparative advantages and disadvantages of a number of statewide and regional systems for the RALI Program of the Geologic Survey (1978)

5.2.1 G.I.S. Software Issues

Software for Geographical Information Systems is currently in a state of flux as the emphasis changes from grid-cell-oriented systems to polygon-oriented systems. The greater flexibility of the polygon approach and the elimination of many scale-effect problems inherent in a grid system have provided the driving force for this transition. Due to the computational complexity of some operations involving polygons, these procedures are often approximated by grid-cell methods. As better polygon-oriented algorithms are developed and as hardware advances (cheap parallel processors, array processors, networked micros, and software support for microcode facilities) are brought about, G.I.S. software will come to execute all primary operations directly on polygon-representing data structures. Many fundamental issues in the analysis of spatial data remain to be answered by scholars. This is reflected in the paucity of analytic tools in all types of G.I.S. facilities. The meaning of spatially distributed data must be imputed from the G.I.S.'s graphic & tabular outputs by human, not programmatic means. G.I.S. software will continue to evolve in the years ahead as more complex hardware and more sophisticated analytic tools are brought to bear on the more difficult phases of computer-carto-graphics. Over the next couple of decades, pixel-oriented data from remote sensing systems will be integrated into G.I.S. systems. Breakthroughs in data storage technology and substantial advances in the mathematics of picture grammars will allow an eventual integration of polygon and pixel (picture) systems.

The process of selecting potential Geographical Information System software should continually be modified to take into consideration the State's changing needs for output (manipulation and analysis included) and the changing tableau of resources being offered by the several software sources. This on-going survey should assess not only the capabilities of the software offerings, but also such issues as transportability, ease of implementation, and maintenance, quality of source code (correct & robust algorithms and lucid style), execution efficiency (computational algorithms & disk I/O mostly) & hardware and software requirements. Within this effort important software sources have been identified and a continuing information collection process has been initiated.

Most of the G.I.S. software systems monitored will be in the public domain and will be available at a nominal charge for copying the transmittal tape and for documentation. In some cases technical assistance and training might have to be acquired also. That is likely to be billed at commercial rates where the activity does not fall under an existing inter-governmental assistance program.

Additional information may be collected on a continuing basis by scanning NTIS & COSMIC (NASA), publications together with GEO ABSTRACTS & COMPUTER and CONTROL ABSTRACTS. The National Cartographic Information Center is also likely to be of some assistance in monitoring G.I.S. software developments in Federal agencies.

Detailed evaluation on a continuing basis will require 350-1000 hours a year by personnel knowledgeable of GIS and the applicable programming languages. The lesser estimate would be appropriate where management policy restricted the scope of the monitoring process by eliminating from consideration certain classes of G.I.S. systems such as those written in PL/I, those based on grid cells, or particular data structures, or those written for systems that would be difficult to convert for use on State hardware.

Beside the basic survey of G.I.S. software several other monitoring efforts will be required. No G.I.S. system will provide all the analytic capabilities required by line agencies for "production" tasks or by researchers for more erudite endeavors. Software compatible with the G.I.S. hardware must be monitored in the following areas to augment the core G.I.S. capabilities: statistics, geography, graphics, modeling & data base management.

The most likely candidate for a staistical package is SPSS (from SPSS, Inc.) because of its ease of use, its professional maintenance, and its availability on many computer systems. Other packages such as BMD(P), OSIRIS&SAS shouls be profiled also, but the dearth of sound statistical packages in the small to medium systems market and the sales potential (several tens of thousands) should lead to the introduction of several significant new statistical packages in the next few years. IMSL's offerings should also be noted because of the very high quality of the subroutine package and its potential (either as parts of user-written programs or unified by a data managing "driver") for use in tandem with a G.I.S. facility.

Auxiliary geography programs are available from several sources i.e., University of Kansas, University of Michigan, Bureau of the Census, and Harvard University. The programs available from these sources are usually special purpose packages that provide a single or only a few functions. The input and output of these sundry programs are not, in general, standardized but they provide a ready source of "raw material" that an adept programmer can meld into an existing G.I.S. facility at low cost.

Another class of software that should be preiodically examined is conventional computer graphics. Packages range from basic systems to control graphics equipment such as plotters and image displays to sophisticated systems for interactively developing complicated graphic prsentations. Although the data structures normally used in computer graphics differ from those required by G.I.S. and cartographic applications, the graphics software will be applicable (at least at a low level) in many situations. Charts and graphs for reports, blocks of code for shortening the development time for some G.I.S. modules, and most aspects of a computer-assisted manual cartography system can be derived from existing computer graphics packages. The three best known systems are GINO-F from the Computer Aided Design Center, DISSPLA from Integrated Software Systems Corp., and GCS from the U.S. Army Corps of Engineers. A detailed profile of such systems found in 1977 ACM SIGGRAPH publications provides a good review of existing systems.

The model-making software tools which will be required to simulate natural and human systems fall into four general catagories (1) discrete (SIMSCRIPT from CACI) (2) continuous (DYNAMO from Pugh-Roberts Associates for systems dynamics modeling and CSMP from IBM) (3) hybrid (GASP IV from Pritsker Associates) & (4) specialized subject-oriented programs written mostly in Fortran. SIMSCRIPT is currently undergoing conversion to several mini systems, and is currently available on a number of mainframes. DYNAMO exists in a transportable Fortran-based version (although it is not the latest version), and GASPN is written in ANSI Fortran and will run on almost any system. The status of these basic systems and the availability of other packages should be monitored in "Simulation" and the annual ACM roster of programming languages and by vendor contact.

Data base management systems should be profiled to determine if available systems provide any of the capabilities required to maintain the geographic-oriented data bases of a G.I.S. facility or to interface with socio-economic files. Although any respectable G.I.S. is in part a data base management system, this capacity may not be conveniently available to perform such activities as editing or inquiry activities and it may not be adaptable to the maintenance of other types of data files. Most DBMS's tend to be expensive, difficult to transfer from one type of system to another, and available on maxi (mainframe) systems or large minis. Both independents and computer manufacturers offer DBMS, but with the exception of possibly only MUMPS, none is available across a wide range of systems. Standards for DBMS exist but are only loosely applied in practice. Information on this topic is best obtained from the hardware vendors of potential systems, trade publications (the sellers of this commodity advertize heavily), and organizations such as the MUMPS users group.

The language in which the chosen G.I.S. is written and the language used for in-house programming will affect many aspects of a Geographic Information System operation. At the moment, only one language, Fortran, should be considered for either role.

The basic requirements for selecting a primary site language are: (1) ability to do the job, (2) few hardware selection restrictions, (3) application software availability, (4) no major software transfer problems, (5) ease of writing, reading, & maintaining programs & (6) availability of trained programmers.

Of the large number of languages currently in use only a few merit notice as the primary language in a G.I.S. context: Fortran, Assembler, PL/I, PASCAL, APL, and Basic.

Fortran meets all criteria reasonably well. It is a scientific language well adapted by design philosophy and compiler optimization attempts to function well in a "number crunching" application like graphics. It's major weakness is the lack of sophisticated data type capabilities as part of the language itself. Data structures of any type may be created, but all the details must be seen to explicitly by the programmer. The language does not relieve him of any of these clerical tasks. Despite this drawback, most G.I.S. systems are based on Fortran. Almost all computer systems have Fortran compilers, so few hardware selection problems would result from the choice of Fortran as a local standard. A vast amount of quality software is available in Fortran in the areas of graphics, cartography, statistics, numerical analysis, simulation, and other fields related to a G.I.S. situation. Some transportability problems exist in spite of approved standards for the language, but it is the language of choice (of most programmers) when transportability is the primary issue. Fortran code, the not inherently modular or structured, may be written in a comprehensible and maintainable form. (Of the languages under consideration only PASCAL offers significant advantages in clarity.) A ready supply of Fortran programmers exists as most colleges use this language as the primary vehicle for instruction in scientific computing.

A new Fortran standard is about to be introduced. A number of the language's weak points will be corrected, but a flexible data structure facility will not be included. Existing Fortran IV programs will run under the new Fortran 77 (also called Fortrev) with few if any changes.

Assembler is not an appropriate choice for a primary G.I.S. system language. It is hardware dependent, is not transferable in general & possesses no store of applicable software. Assembler programs are difficult to maintain and few programmers are skilled in its use.

PL/I merits some consideration as the primary language in a G.I.S. situation. It is designed for scientific computing (amongst other things) and offers more data structure handling capabilities as part of its syntax than does Fortran. PL/I tends to be somewhat slower than Fortran and does not directly offer all the data structure capabilities required by a G.I.S. implementation. The language is available only on a small number of systems although that number is growing. Only a modest amount of G.I.S.-oriented applications software is available. The language is reasonably easy to use if one stays away from its arcane features and takes proper advantage of its block structuring. It is somewhat difficult to find good technically-oriented PL/I programmers.

PASCAL merits future consideration as a basis for G.I.S. systems, and current consideration as a research tool. Its structured programming orientation and user-definable data structures make it a nearly ideal logical framework for a G.I.S. system. Because most implementations are interpretive or semi-interpretive, execution is slow. Versions utilizing firmware "target machines" or compilers should relieve the speed problem. PASCAL is currently available on a small number of machines but its rapid rise in popularity and the relative ease with which it is implemented will likely change that situation. Little application software is available in PASCAL and few practicing programmers are familiar with it. Experimental versions exist which are oriented towards parallel task execution.

APL and BASIC have both been used in graphics situations (often by the addition of graphic "primatives") but their usual implementation as interpreters makes them too slow for the heavy computational load presented by geo-graphics. While BASIC is nearly universally available, APL is available on only a few systems. A significant amount of applications programming exists in BASIC, but in general, it is not sophisticated enough for a scientific/G.I.S. environment. Very little applications programming is available in APL. Some transportability problems exist with both languages. While BASIC source code (for medium and large programs) is not particularly easy to maintain and modify, APL programs are distinctly worse. A good supply of BASIC programmers exists, but few programmers are conversant with APL. These two languages are best relegated to the position in which they will best serve; as adjunct languages in a G.I.S. setting they offer a very quick way to write short programs that will be used once.

Software transportability is a major consideration in configuring a G.I.S. facility. It would not be financially possible (and certainly not reasonable to attempt to write all G.I.S. and ancilliary programs in house. Many quality public domain and proprietary packages are available for a fraction of their development costs. To adequately make use of this enormous resource (and in some small way to contribute to it) one must have a grasp of some of the basic issues of software transportability in a G.I.S. context. Object code (programs essentially in the form that the computer executes) is not covered here since it is usually transportable only between a limited number of identical or nearly identical hardware systems. Source code (programs in the form that is written and understood by humans) for many languages is much more easily tranfered from one type of system to another. Programs known as compilers or interpreters (peculiar to each computer) translate the source code of a high-order language such as BASIC or FORTRAN into the "unique" form accepted by a given computer's processor. Given the constraints of a G.I.S. environment transportability needs to be considered only in terms of a restricted number of languages.

PASCAL is not widely used currently and little practical experience exists in transferring large applications packages from one implementation to another. PASCAL does have the advantage of formal definition that is a defacto standard. The most frequently encountered implementation is a transportable compiler which produces code for an imaginary "target" machine. About the only system dependent feature is the interpreter which then decodes these instructions in terms understood by the particular computer. This approach is less likely to produce compatibility problems than is the unique (machine-dependent) compiler approach. The rapid increase in popularity being experienced by PASCAL will likely generate development activity that should be monitored as part of a G.I.S. planning effort.

FORTRAN is the most easily transported of nearly all languages. This is in part due to the existence of a formal standard for the language (1966). Beside this official standard, a kind of unofficial standard developed over the years as vendors added features (such as direct access, quoted hollerith strings, generalized subscripts, variable dimension sizes,...) that were commonly adopted by the industry. IBM G-level fortran is representative of this practical standard. A new official FORTRAN standard (1977) is in the works. It is not yet clear what affect the new standard will have on program transportability, but the problems are not likely to be severe.

Difficulty in converting programs written in one vendor's FORTRAN to run on another system often results from subtle deviations from the standard or alternative interpretations of the imperfect or incomplete description of the context by the standard. Details of implementation are not spelled out by the standard (and probably shouldn't be) and this too leads to unforeseen problems. Several different schemes exist for representing integers and floating point numbers. When a program being converted makes use of the bit patterns in a way that the facilities of the FORTRAN system would not, difficulties may arise. Computational situations, tho syntatically indentical may result in different answers on two systems because of different rounding, truncating, word length, and algorithm features. Differences in collating sequences can cause great problems with sorting and searching operations. "Do loop" parameters, subscripts, and subprogram arguments are other constructs commonly subject to subtle differences in performance. Some computers are mostly register-oriented, some provide extensive hardware support for stock operations, and others are essentially stack machines. These design differences give rise to variable capabilities for identical blocks of code. For instance a mult-tasking Fortran-written applications package developed on a stack machine and using recursive and reentrant features of the code will not run on a register-oriented machine that does not allow subroutines to call themselves and does not permit shared code. Many other features of general system design, compiler implementation, and operating system interaction may cause software transfer problems.

In practice the issue is considerably simpler. A program of "foreign" origin can fail in only a few basic ways. If it does not compile the problem is usually easy to spot and fix. If it does not execute properly, a few runs with a good test data set will usually find the difficulty. The remaining small percentage of conversion malaises should be referred to a competent programmer. The programmer should identify the troublesome code with debugging techniques and should rewrite portions of the program to isolate potential trouble spots. Rewriting stylistically poor code and inserting adequate internal documentation will eliminate many problems in this class. Another approach is to use the PFORT Verifier to isolate code which violates the (severe) PFORT restrictions. Once rewritten and tested against a sound test data set most problems will have been removed.

Writing FORTRAN programs for in-house use or for distribution requires stylistic dicipline on the part of the programming staff and sound documentation practices. Programs should be written using only those features of the local FORTRAN compiler that meet the standard. (The vendor can supply the National Bureau of Standards test results.) Non-standard or necessarily-clumbsy code should be placed in seperate modules and appropriately identified as should code which may be mathematically unstable (mostly problems revolving around finite arithmatic precision). Alternately the software may be written in PFORT or may be produced using the pre-processor techniques developed by IMSL. The latter is only practical in a large scale software development and distribution situation.

The lifespan of software is determined by a variety of factors including subject-oriented advances. Innovations in hardware in the areas of parallel processing, content addressable memories, and networking will generate future software regimes potentially much different from today's.

Many present systems will become obsolete because an inherently superior methodology will be available. The same ware of technical innovention that will obsolete some software systems and languages will also provide the means for easily maintaining old, but useful, software. Emulation, compiler generators, and other software techniques will allow these programs to be moved with some ease to a novel, non-native, environment. The computer industry is not likely to forget the problems of moving from second generation to non-compatible third generation systems. G.I.S. software is likely to be in a state of turmoil for the forseeable future. Better, or even just appropriate, algorithms will be found for such tasks as polygon overlaying vector to raster conversion & vice versa, manipulation of sparse matrices & statistical analysis of spatially distributed data. The whole realm of picture processing is undergoing rapid mathematical development and it will have substantial future impact on Geographic Information Systems. The near future is likely to see some limited forms of parallel processing become available at a reasonable price. Many G.I.S. functions are inherently parallel in nature (intervisibility analysis, gridding, contouring, overlaying, some types of sorting & searching,...) and the current sequential approach to these problems will not endure for very long (probably less than 15 years at the outside). Illustrative is the task of overlaying polygons which has given the USGS so much trouble. Multiple parallel processors could cheaply and quickly solve these problems and the software to do the task is conceptually simple. Parallelism in the form of a group of networked minis could even today offer a practical, bruteforce solution. In general the future of G.I.S. software is one of change. The best that can be hoped for in an operations sense is that the transitions are smooth.

Geographic Information Systems are in part a new dicipline and in part a novel integration of existing specialties. The care of the subject is computer science (data structures, data management, and algorithm specification and implementation) along with some basic cartography. Key personnel must be skilled not only in the centrol GIS concepts but also in the peripheral components drawn from mathematics and statistics. In practice the primary project individuals need to possess a basic understanding of the technical issues faced by the clients they will be serving in the natural systems and human services fields. The selection of personnel is the most important phase of initialing a G.I.S. capability but it is the most difficult area in which to prescribe optimizing action.

Top personnel must not only be good technicians but they must also have a very clear perception of where they and their G.I.S. system stand in the scheme of all things. Tho this technology offers powerful means of organizing, manipulating and displaying data and of building models to give insights into the workings of the world, it also offers subtle and alluring pitfalls. Not only is "garbage in garbage out" a data processing truism, but, in a G.I.S. context of applying simple computer models to complex natural systems, "goodies in and garbage out" is perhaps a more appropriate aphorism. G.I.S. project management must understand that their fancy system has the potential for being an expiditious means of doing very inappropriate things. They must also come anatomically equiped to convey this message to their superiors when requested to "turn the crank". Selection of personnel offers the interesting situation of few being knowlegeable enough about (important) G.I.S. subtilties to take the true measure of job condidates.

To aid the examination of personnel factors G.I.S.-related jobs are lumped into three categories: director-programmer, programmer operator, and data-entry specialist. (Three individuals could potentially operate a G.I.S. facility during the early development stages, but an operational system would probably require more.) The size of the G.I.S. staff is dependent on a host of factors not yet identified or fixed.

SKILL REQUIREMENTS

AREA OF EXPERIENCE OR KNOWLEGE	PERSONNEL CATAGORIES		
	DIRECTOR PROGRAMMER	PROGRAMMER OPERATOR	DATA ENTRY SPECIALIST
PROJECT MANAGEMENT	REQUIRED	HELPFUL	–
GEOGRAPHIC INFO SYSTEMS	REQUIRED	REQUIRED	–
FORTRAN PROGRAMMING	REQUIRED	REQUIRED	–
SYSTEMS ANALYSIS	REQUIRED	REQUIRED	–
SYSTEMS PROGRAMMING	SOME BKGD REQUIRED	SOME BKGD REQUIRED	–
COMPUTER OPERATION	REQUIRED IF IN-HOUSE SYSTEM	REQUIRED IF IN-HOUSE SYSTEM	HELPFUL
DIGITIZER OPERATION	REQUIRED	REQUIRED	REQUIRED
KEYPUNCHING	–	HELPFUL	REQUIRED
PLANNING	SOME BKGD REQUIRED	HELPFUL	HELPFUL
CARTOGRAPHY & GEOGRAPHY	SOME BKGD REQUIRED	HELPFUL	HELPFUL
SOCIO-ECONOMIC/CENSUS DATA ANALYSIS	REQUIRED	HELPFUL	–
STATISTICS	SOME BKGD REQUIRED	SOME BKGD REQUIRED	–
MODELING & SIMULATION	SOME BKGD REQUIRED	HELPFUL	–
FORESTRY & AGRICULTURE	HELPFUL	HELPFUL	HELPFUL
GEOLOGY & SOILS	HELPFUL	HELPFUL	HELPFUL
LIMNOLOGY	HELPFUL	HELPFUL	HELPFUL
MARINE & ESTUARINE SYSTEMS	HELPFUL	HELPFUL	HELPFUL
NATURAL SYSTEM ECOLOGY	HELPFUL	HELPFUL	HELPFUL
REMOTE SENSING	SOME BKGD REQUIRED	HELPFUL	HELPFUL

G.I.S. team members can probably be found in Maine (especially if advanced notice of position openings is given.) The director and some of the programming staff will probably not be current State employees. If the "right" people were already on the payroll, this report would have been unnecessary. Some programming staff members and the data-entry specialists should be drawn from current State service for their knowledge of agency operation and working familiarity with the (conventional) methods that will be replaced by G.I.S. functions.

User-agency personnel who will become involved with G.I.S. operations will probably not be selected for that role because of their G.I.S. expertise, but because of their need for G.I.S. services. This lack of supply of knowledgeable users must be dealt with by training and by the creation of specialized G.I.S. interface programs that "speak the language" of the particular diciplines.

Non-State personnel may occasionally be required in a consultancy capacity. A small number of individuals associated with the University and with private concerns are potential G.I.S. advisers. Remote sensing capabilities are available from the University although little associated computer processing capabilities are in place. Regional science, mathematical geography, and cartography skills are at best in short supply. Survey and cadastral specialists are available from the University. The supply of application area (wildlife, forestry, ecology,...) experts will be somewhat diminished by problems of working in a new milieu, but those willing to adapt will be able to tap the substantial potential of G.I.S. systems.

5.3. Survey of Available Hardware

The equipment options must be identified and documentation of specific performance characterisitics obtained. The process should look for ranges and trends in cost and performance initially and only secondarily identify specific products and companies to monitor.

5.3 Survey of Available Hardware

The monitoring of the general purpose computer and computer graphic equipment market places is a necessary component of the planning process for the implementation or a G.I.S. This on-going process will be required to gauge the price trends (for the most part downward), price-performance ratios (usually decreasing), and available units of function (some increments may not be a convenient "size"), and to give early notice of the impact of new systems and emerging technologies. The pace of innovation in this field is truly astonishing. A system chose either on the basis of cost or performance for G.I.S. purposes today would probably not be the system selected a year or possibly only six months from now. If G.I.S. equipment selection is not made with an eye to system "extensibility", the hardware will rapidly become (relatively) expensive and ineffectual in an environment of increasingly complex and voluminous demands and of decreasingly expensive newly-introduced equipment alternatives.

The following subsections provide a thumbnail sketch of general purpose computer gear and computer graphics equipment potentially applicable to a G.I.S. system. Machinery currently available or soon to enter the market is described first, followed by potential developments of note.

The heart of any currently conceivable non-research G.I.S. facility is a stored program general purpose computer. This appellation covers a wide range of systems from micro--computer systems costing a few thousand dollars thru mini-computers costing on the order of a hundred thousand dollars to maxi-computers costing millions. All of these types of computers are potential candidates for utilization in a G.I.S. system. The choice depends on the mix of tasks to be presented to the system. The most common maxi-computers are those manufactured by IBM, Honeywell, Burroughs, Univac, Control Data, and Digital Equipment Corporation. These machines are characterized by high performance, the ability to handle large problems, the availability of extensive systems and applications software, good maintenance service, and high price. The processors of these systems typically perform .5-15 million arithmatic operations per second on "words" of 32 to 60 bits. They have the ability to rapidly access large amounts of data stored on magnetic disk and tape subsystems, to 'simultaneously' serve many users, and to respond to hardware and software errors in a sophisticated manner.

Over the next few years maxi-computers are likely to change slowly in price and capability. IBM, whose market position sllows it to strongly influence the profitability of other computer companies innovative products, has apparently adopted an evolutionary transition from its current 370 series to its as-yet-unnamed successor line. Most other large system vendors will probably follow suit. The next five years will undoubtedly bring performance/price increases of 100-200% based on cheaper main memory and higher density conventional magnetic tape and disk secondary storage.

It is not likely that a 'maxi' system would be heavily used in any Maine State G.I.S. facility beyond the prototype stage. The amount of automated cartography and associated spatial analysis tasks is not great enough to warrent a dedicated mainframe and substantial economies are to be found in smaller systems. The cost of computation on a mini or micro system can be more than an order of magnitude cheaper. Such a smaller system offers a much simplified 'systems programming' and operational environment also.

Mini computers are, today, often mini only in price. They span a very wide range of capabilities which now overlaps a substantial portion of the realm of maxi computers. They usually operate on words of 16 bits, although 12, 18, 24, and 32 bits are also sometimes used. These machines will produce .1 to 6 million operations per second. In recent years the variety and quality of systems software (operating systems language processors, and utilities) available for mini systems has come to rival that available on maxi mainframes. Also, high performance peripheral equipment originally designed for maxi systems has been made available on minis.

Micro-computers, designed to operate on words of 8, 12, or 16 bits are now finding their way into data processing and computational situations which ten years ago were the realm of maxi-computers and which more recently were the baliwick of minis. They can perform up to 4 million operations/ sec. Micro-computers are manufactured as chips by the major semi-conductor firms and sold as systems level products by them and a large number of other firms. Increasingly powerful and flexible systems are being marketed including micro-computer implementations of existing mini instruction sets and bit-slice chip sets for flexible design of rather powerful systems.

The peripherals usually available for general-purpose computer systems include magnetic disks, magnetic tapes, punch cards, and punch paper tapes for data storage and line printers and character-printing terminals for hard copy. A wide range of disk subsystems exists from floppy diskettes, thru cartridge disks, to high capacity 3330, Winchester, and Storage Module devices. The floppy diskettes typically store .25 - 1. million characters of information (denoted 'Mb'), can randomly access any information of the diskette in about 300 milliseconds, and transfer data at a rate of 30,000 to 100,000 characters/second. They are most often found on micro, mini, and intelligent terminals and data entry devices. The cartridge disks usually contain 2.5-20 Mb, have an access time of approximately 50 milliseconds, and transfer data at a rate of several hundred thousand characters per second. Such systems are usually found on mini systems. The 3330, Winchester, and Storage Module type systems offer storage capacities of 40-300 Mb, access times of under 35 milliseconds, and transfer rates exceeding 800,000 characters per second. These subsystems usually have sophisticated controllers for maximizing performance. These high-capacity disk subsystems are available on most maxi, many mini, and even a few micro computer systems. Prices for a dual floppy system range from $1500 to $4500 with controller, cartidge disk systems run from $8.00 to $20,000 similarly configured, and the high-capacity systems go for $ 15,000 to $ 50,000 with a few actually under $7,000.

Magnetic drums and fixed, head-per-track, disks acting as a slow, but low-cost, extension of main memory are also available for many maxi and mini systems. They are usually special purpose subsystems with fast access and transfer capabilities. Recently three other technologies have been adopted to this task. Change-coupled devices, bubble memory systems, and a form of magnetic care memory.

The number of "disks" of any of the above types that can be attached to a given computer depends on the particulars of that computer. Typically for 'minis' four disks or disk subsystems is the limit. For 'maxis' the maximum amount of disk storage is usually much greater. Typical maximum disk storage capacities are .5-4.0 Mb for micros, 40-1200 Mb for minis, and several thousand Mb for Maxis.

The costs of disk subsystems vary widely. A dual floppy system with controller can range from $1,000 to $4,000; a dual cartridge system with controller runs about $18,000; a dual Storage Module system with controller goes for $40,000 to $70,000; and a Winchester-technology dual fixed head system with controller costs about $10,000.

Magnetic tape systems are usually based on one of three media-standard ½ inch tape, ¼ inch (3M) cartridges, or cassettes (similar to audio cassettes). The standard ½ inch tape is the basis for most tape systems for maxi and mini systems. A variety of recording standards exists with some systems working with seven tracks (200 bpi (characters per inch) & 556 bpi) and others with nine tracks (800 bpi, 1600 bpi, and 6250 bpi). The tape drives which read and write ½ inch magnetic tape usually process only a single mode and density with nine track 1600 bpi being the most common capability. The 6250 bpi and multi-density drives are most commonly found on maxi systems. The maximum storage using the standard 2400 ft. reels is about 40Mb (at 1600 bpi) and 160Mb (at 6250 bpi). Transfer rates vary from 20,000 to 200,000 characters/second. The cost of a single drive with controller varies from $10,000 to $40,000 depending on speed, flexibility, special features (such as automatic tape loading), and controller sophistication.

The 3M cartridges and cassettes are most commonly found on small-mini, and terminal equipment, although in some cases high-density 3M cartridges have been used on larger minis to replace ½ inch magnetic tape where transportability was not required. Generally a 3M cartridge will contain .3-3Mb and the cassette .1-.3Mb with transfer rates around 50,000 characters/second for the 3M cartridge and characteristically much less for the cassette. Prices for dual drive systems with controller are approximately $3400 for the cartridge and $ 2000 for the cassette.

Punch cards and paper tape facilities should not be considered for inclusion in a "modern" system unless the implementor requires them to interface existing equipment such as a keypunch pool's owned punches or lab instruments producing punched paper tape or if they are necessary for communication with "foreign" systems.

Alpha-numeric terminals for data entry, inquiry, & time sharing, produced either as printing or CRT (television-like) devices, are available from a variety of vendors. These terminals may be utilized proximate to the computer or remotely over telephone lines or other communications links at speeds of 11 to 960 characters/second. Many have elaborate editing features, local storage, some local processing capability, and ability to interface to a variety of different systems. Most CRT terminals offer a display of 80 columns by 25 lines while the printing terminals reproduce the line printer format of 132 columns. A few of the "dot matrix" and "daisy wheel" types of machines provide graphics facilities. Prices for quality CRT's start under $1000 while good printing terminals begin at $ 1500.

Beside these "conventional" peripherals, there are other devices that should be noted, especially in the G.I.S. context. Speech recognition and voice response units are currently available, functional in cases where limited vocabularies are sufficient, and potentially cost effective. A digitizing station utilizing these capabilities would probably be twice as productive as one without them. A digitizer operator needs to carry on a complicated series of status inquiry, status setting, and data editing operations but he often is not free to use a keyboard device. A voice facility would allow these functions to be easily executed. Commercial grade voice input systems cost about $12,000 but the existance of hobbyist system with nearly the same capabilities and a price of $ 300 bodes well for the future. Voice response systems cost from $500 to $7,500 depending on the technology employed, sound quality, size of vocabulary, and operational flexability. Very sedate dial capabilities should be available for $1,000 within a couple of years.

Integrated computer graphic systems marketed for automated cartography are produced by Broomall, Bendix, Calma, and Applicon. These systems tend to be more oriented around computer assisted drafting than data base management of graphic entities. In most cases the data structures utilized by these integrated systems do not provide the flexibility required for general planning cartography nor do they provide direct compatibility with likely GIS data structures. Due to the low sales volume or such systems their prices are high relative to the value of software and hardware components. Prices range from $125,000 to $ 300,000.

Integrated computer graphic systems marketed for general purpose graphic work offer better performance/price ratios than systems targeted for cartography. Packaged general purpose systems and locally integrated graphic peripherals are currently the most auspicious approaches to developing the graphic capabilities of a GIS implementation. Tektronix produces two systems of potential utility in a Maine system. The 4051 is a microprocessor (Motorola 6800) based graphics system employing a storage tube, cartridge tapes, stand-alone BASIC language capability, instumentation

face and PS communications. It provides adequate capabilities for an intelligent digitizing station and can also act as a graphic terminal. The Tektronix 4081 is a stand-alone and terminal emulating system based on an Interdata mini computer, large screen combination storage/refresh graphic tube, disk storage, and communications. It offers more capability than the 4051 but suffers from a lack of flexibility imposed by the proprietary operating system. Both IMLAC and Megadata sell refreshed high resolution graphic systems which avoid this problem. Each employs a mini computer running under the mini vendor's standard operating system and supporting all standard peripheral devices and standard software including Fortran. Comtal corporation offers on LSI-11-based pixel-oriented color picture processing system that has sufficient resolution (1024X1024) to allow it to function in both a line drawing and an image processing role. The prices for the above systems range from $15,000 to $50,000.

The following paragraphs provide a cursory view of computer graphic input and output peripherals that may be added to most graphics systems or may be used as components in a locally designed graphics facility.

Computer graphic input peripherals and data acquisition systems convert information contained on photos, charts, and maps and signals received from objects in the natural world into a form that can be processed by digital computers. In a GIS context the most common systems are tablet digitizers and remote sensing systems.

Tablet digitizers are available in a variety of forms. (some are essentially modified drafting machines and should be avoided because of mechanical problems and high inertia of the cursor assembly.) Most digitizers today consist of a flat surface on which the object to be digitized is placed, a cross-hair cursor or stylus for selecting points and generating status information, and an electronics assembly for control of degitizer operation and interface to a local computer, a local storage medium (cards, tape, floppies, cassettes, or punch paper tape), or a communications line. The working areas range from 11" X 11" to 42" X 60" with accuracies, resolutions, and repeatabilities from .01" to .005. Common options are multi-button cursors and menues for increased flexibility in entering command and status information, backlit boards for use with photographic negatives or for forced interpolation to a base map, & rear projection capability for extracting information from slides. Prices for tablet digitizers without interface or auxilliary recording media range from $3,000 to $12,000 depending on size, metric capabilities, and special features. Among the vendors offering tablet digitizers are Talos, Science Accessories Corp, GTCO, CEC, Summographics, Bendix, and Computer Talk.

Optical line following digitizers provide accurate, high-speed capture of line data. They require meticulously prepared graphics as input and are only partially automatic, requiring human assistance for classification of data and for rectifying pathological situations (perfidious nature, careless draftsmen, and stochastic dirt as still more inventive than the programmers hired to create automatic digitizing systems). These devices are very expensive.

Raster scan devices for optically digitizing source documents include television cameras (especially those having capacitive discharge recording systems), scanning microdensitometers, and Raster scan photo cell arrays. These systems render a map, chart, or photographic product as a two-dimensional array of intensity values. The raw output must be reorganized to represent a new array of pixels conforming to the metric of the database, or meaningfully aggregated into data structures representing line and areal entities. In practice this task is very difficult except for "pixel" oriented systems where input graphics have been manually produced in conformance with the system's metric.

Other devices used for graphic data input are light pens for use with raster scan tubes & joy sticks, track balls, and thumb wheels which are options on both refresh and storage tube systems for positioning of graphic cursors. Touch sensitive screens and function panels may also be used with both types of display devices to control graphic variables.

Remote sensing systems such as LANDSAT, spatially distributed environmental monitoring systems, and even (soon to be deployed) point-of-sale and electronic funds transfer systems can be considered generalized graphic peripherals. Real-time spatially distributed computerized monitoring systems are likely to proliferate in the near future, blurring the traditional distinction between the data collection process and the data analysis function.

Computer graphic output peripherals provide displays of point or line

line data temporarily on the face of electronic instruments or permanently on paper, drafting media, scribing material, and various types of "photographic" film. The two traditional systems, electro-mechanical plotters and cathode ray tubes, have been augmented in recent years by a large number of new technologies.

Two basic types of pen plotters, drum and flat-bed, are produced in a wide variety of sizes, capabilities, and prices. Drum plotters are usually found in 11" and 34" widths with the length of the plot limited by the length of the paper (up to 150'). The plotters' pen carrage may have as many as four pens. The types of pens usually offered include pressurized ball point and capillary action pens for use with paper, india ink pens for writing on paper or drafting film (mylar, etc.), and, in some more expensive units, scribers. Prices range from $3,500 to more than $15,000 for drum plotters without interface. Flat-bed plotters are produced in small (11" X 17") timesharing units and large (34" X 34+") high-accuracy devices. Ball point, fiber tip, and india ink pen systems are supplemented on large flat-bed plotters by scribers and photo plotting heads. The accuracies on both types of plotters exceed .0025" for moderately priced units. Plotting speeds run from 2"/sec. to 40"/sec., but for most cartographic applications speeds above 10"/sec. are not useful because of the high percentage of short rectors. (It takes a while to accelerate to maximum speed.) Prices for the small flat-bed plotters are around $4,000 and for the large models range from $15,000 to well over $50,000. Interfaces for plotters range from RS-232, custom mini interface cards, and time-sharing devices for code compression and hardware graphic generation, to off-line controllers utilitzing magnetic tape to supply plotting information. Controller prices range from $1,500 for RS-232 and some mini interfaces to more than $15,000 for off-line units. Vendors of pen plotters include Houston Instruments, Calcomp, Broomall, Zeta, Gerber,

& H. Dell Foster (K&E) amongst others.

Electrostatic plotters utilizing a single color pattern of dots (with spacings of .01 to .005) are available in a variety of widths up to 72". Plotting speeds are around 2 inches/sec but vector to raster conversion prior to plotting may consume a substantial amount of time. Prices range from $8,000 to more than $25,000 for the plotters and controllers are priced approximately the same as those for pen plotters. Electrostatic plotters are not as useful in small-scale cartographic operations as are pen plotters, but where fast, multiple-copy, monochrome plots are required or where frequent storage tube hard copies are required, they may be useful. The high cost of special paper and supplies also restricts the circumstances in which they are cost-effective. Versatec, Gould, and Varian are the best-known electrostatic (printer-) plotter manufacturers.

Inkjet plotters, available from Herz and Applicon, provide spectacular color capabilities on plots up to 22" X 34" utilizing a variety of media. Areal plotting (coloring) capabilities are considerably superior to line drawing capabilities, but for thematic mapping it is a nearly ideal low-volume device. Prices are approximately $50,000 with off-line controller.

Display screen devices in common use are storage tubes, CRT's, and gas discharge (plasma) panels. Storage tubes retain the image painted on the screen phosphers. This allows a complicated pattern to be built up over an extended period and permits the use of a low-speed interface to the computer. Changes to any component of a displayed plot, however, require the entire screen to be replotted. Common resolutions for these vector-oriented systems are 1024 X 1024 and 4096 X 4096 addressable points and screen sizes range up to 19" diagonal. Communication with the processor may occur at up to 9600 baud if local or transmission constraints permit. Prices run from $3,500 for minimal models to over $15,000. Hybrid devices having both storage tube and refresh characteristics exist as part of higher priced systems. The two primary vendors of storage tube devices are Princeton Electronic Products and Tektronix.

CRT graphic display tubes are produced with vector and raster display methodologies. Vector systems more naturally conform to the operations of cartographic line drawing while raster scan devices provide a more appropriate mode (especially in color systems) for areal definition. Vector CRT resolutions vary from 512 X 512 to 4096 X 4096 while raster displays usually have approximately 1000 scan lines for high quality units and about 500 or 250 for less expensive units. Both vector and raster CRT's must have any image on their screens displayed at a rate of 30-60 times a second or the observer's eye will be irritated by the flicker. This refresh capability requires that the "memory" for the display not be the physical properties of the screen phosphors (as it is for storage tubes) but be RAM or CCD either within the refresh graphic terminal or a short distance away in a computer and coupled to the terminal by a very high speed direct memory access scheme. Vector systems allow dynamic changes to be made to a line drawing which is a distinct advantage in an editing situation, but they also are limited in the amount of detail that can be shown at a given time (due to the amount of memory available for refresh and the rate of which refresh must occur). Manufacturers of vector-oriented terminals include IMLAC, Tektronix, Hughes, Calcomp, Vector General, General Turtle, and Megatek. Raster-oriented display are produced by Ramtek, Lexidata, Aydin, Data Disk among others. Prices range from $3,000 to over $30,000 depending on features and options.

Plasma display panels consist of an array of points of light that may individually be switched on and off. The screen does not require refreshing and thus may be used with low-speed serial interfaces to local or remote computers. Panels are normally arranged in a 512 X 512 point array with 60 points/inch in both dimensions. Plasma terminals are manufactured by Magnavox, Applications Group, Burroughs (alpha-numeric only currently), and several military systems vendors.

Line printers and printing terminals may be used as graphic output devices in circumstances where they offer sufficient resolution. Pen plotter emulation programs are available and many mapping packages exist with output designed for rapid line printer review as well as for high resolution slow pen plotting. Recently dot matrix terminals and line printers have come onto the market with impressive plotting capabilities. The only device capable of meeting the volume demands of a GIS situation is Printronix's 300 line/min. printer-plotter. This device plots 60 dots/inch horizontally and 72 dots/inch vertically at a rate of 17 inches/min. on standard line printer stock. (One-twelvth inch squares may be employed where equal spacing is required.) The price is approximately $6,000.

Other less common graphic output equipment includes dry silver and electrostatic devices for making copies of storage tube images ($4,000-$10,000), film plotters such as those produced by Dicomed, & laser & electron beam recorders manufactured by Gerber and several other firms. Several computer output microfilm systems have plotting capabilities in addition to their hardware generated character sets. Typically resolutions of 16,000 X 16,000 or 32,000 X 32,000 points are available. Both raster and vector oriented COM systems may be used, but vector systems offer very substantial savings for cartographic applications. Since these machines cost $150,000 to $300,000, a GIS facility would contract for this service.

A number of technologies offer future potential for GIS systems. No reliable projections can be made about performance characteristics, introduction date, or prices and certainly the citations are incomplete. Any continuing GIS implementation study or active GIS group should maintain an ongoing profile of emerging data capture, storage, processing, & display technologies.

In terms of GIS systems the most important development likely to be introduced in the next few years is the video disk. Experimental versions of both analog and digital systems have been produced and mass marketing of the analog system is imminent. The digital version should offer the same access and data transfer capabilities as conventional disks but are capable of extremely high storage densities and exceedingly low costs (due to the economies of scale provided by the closely-related home entertainment systems). One to perhaps 25 billion bytes can potentially be encoded on a single (less than $10) medium and read or written using a device costing substantially less than $5,000. Initial versions will probably be write-once, but substitution of optically-switchable recording media for holes burned in an aluminum substrate by a laser should eliminate that restriction. In late 1977 a lower performance photographic (central recording only) system was announced that was in many ways analogous to video disks.

Large screen displays based on liquid crystal, light-emitting diode, plasma, or other technologies may become practical within the next decade.

Fiber optic transmission systems, some already being field tested, will soon offer very fast data communication capabilities that will make possible networking, database sharing, load spreading, and other activities designed to maximize the performance of whole systems of computers and to provide alternatives when individual processors or peripherals fail. Point-to-point digital satellite service is scheduled for 1981. This will provide virtual site independence for components of networked systems. Potential transmission speeds for these systems are several orders of magnitude greater than todays best data transmission facilities. The technical issues in future data communications capabilities will likely be subservient to political and legal issues. The boundaries between computing and telecommunications and between the Bell Systems and the specialized carriers will occupy the time of Congress and the Courts for a long time to come.

Ovonic graphic media will come onto the market in 1978 with 3M's introduction of a dry micro-fiche system with the remarkable property of being erased or written a single frame at a time. The potential of this type of material for computer cartographic output is substantial. This material also has potential utility as a digital or analog (holographic) means of data storage.

Optical computers have a remote potential for very fast operation which may exhibit a high degree of parallelism. Lack of sufficiently high-grade optical components is one of the major impediments to progress in this area. Very fast processors and memories may eventually emerge from IBM's work on Joseph Junction devices and from Burroughs' development of ovonic circuitry. Most advances in computer power in the short run are likely to result from the incremental increases in speed as circuit density goes up and from various forms of parallel operation.

5.4. Availability of Personnel

The capabilities of personnel must be weighed against personnel requirements needed to develop, maintain and use a geographic information system. Good background on existing capabilities, including the ability of potential system users to deal with their problems in a quantitative manner will provide the basis for personnel planning.

5.5. Some Hypothetical Alternatives

If the State were faced with the immediate requirement of developing a Geographical Information System facility for itself and possibly also for the Regional Planning Commissions and the towns, it would not be possible to evaluate all of the possible organizational settings for such a service nor would it be feasible to exhaustively scrutinize the current and future software and hardware considerations. In such circumstances the course of action would have to be determined on the basis of the personal judgement of the State and consultant managers and technicians selected to carry out the mission as opposed to the slow, detailed, analytical approach espoused by the bulk of this document. A GIS system derived in the former manner would not necessarily be an inferior product. Indeed, a skilled manager operating on his 'instincts' will often produce a better-integrated product than will a committee, composed of diverse interests, which spends its time examining details. The following subsections describe one such scenario and discuss the costs and other parameters associated with securing the required GIS functions from different sources.

The personal opinion of the author is that there are two viable approaches to a near-term commencement of GIS projects: a low-level approach utilizing University or consultant services and equipment & a mid-level approach involving in-house or consultant systems. No high-level approach is suggested because of the costliness of such an operation, the rapid changes of performance/price ratios, and the lack of trained personnel. In the following subsections only the mid-level approach is outlined. The low-level approach would consist of a restricted subset of the mid-level set of capabilities which, though probably not as cost effective on a per-unit basis, might well provide the least costly path to an eventual full-blown GIS facility at a time when equipment was relatively cheaper and the technology was more mature. The choice of tasks that would constitute the low-level approach would depend on State priorities and the availability of particular graphics, computation, software, & personnel resources in the region. Costs for the low-level approach would run from $24-50K per year for personnel (1.5) FTE's computer time, equipment rental, etc. for research to keep abreast of the field & for the production of some of the more basic products cited in the mid-level profile. Additional requirements would result in corresponding higher costs. Several very specific GIS tasks might be undertaken on a project basis for under $25K each, but pooling such jobs in a continuing series of related jobs would probably be more fruitful. The choice depends on the particulars.

5.5.1 Purposes of a Maine GIS Facility (Mid-level)

A mid-range State GIS facility would provide a "local" capability for storing and analyzing spatial data for meeting State needs and anticipated Federal requirments for manipulating natural systems information. It would have the potential for becoming a cost effective replacement for or assistance to conventional cartographic operations. Given the drift of cartography towards an eventual digital constitution and, in particular, USGS's long-range plans, the implementation of a State GIS capability would commence the long processes of user training and infrastructure develeopment that would eventually allow Maine to enjoy the benefits of this new technology.

This State facility would have the ability to install the USGS's LUDA software and database for experimental and production purposes. It would also support the development of an in-house GIS or the modification of a "foreign" GIS to meet peculiar Maine needs. (Different data structures may be needed to optimally accomodate the local topology; different file organizations may be required to handle local inquiry and utilitzation patterns.) An existing system would allow the State to make "live" evaluations of software developed at other sites. This local evaluation is important as many GIS operations are "data dependent". (For instance, a system developed in the Mid West and tuned for the mostly "rectangular" land use geometry prevalent there may perform very poorly in a Maine context where irregular natural features are more likely to influence land use.)

Production of a number of useful planning and management tools would be possible with a computer-based GIS. Maps and other graphics such as cartograms, graphs, and charts may be rapidly generated from the GIS data base. Similiarly, reports and statistics may be very quickly extracted from the data base by a fully implemented GIS. A well-designed facility can handle most tasks of this kind hundreds of times faster than a draftsman or clerk and at substantially reduced unit cost.

Conventional cartography in the near-term can benefit from being piggy-backed onto a GIS effort in many cases. Change of scale, elimination of systematic distortion, change of projection, rectification of maps of varying "metric" and many other activities may be under taken digitally when maps have been digitized rather than, or in addition to being drafted.

Another piggy-back benefit of a mid-level GIS system is a general purpose computational and data storage capability. In a planning context this would most likely be utilized for maintaining and processing of socio-economic data bases, word prcessing (or an archieval interface to a distinct word processing system), running selective mailing lists, indexing documents, facilitating communication by routing information to agencies and citizens according to catagories of interest, and innumerable other general logical and computational tasks.

A mid-level GIS capability would include graphic data entry capability via on-line manual digitizer for low volume capture of original graphic information and for correction and update of the various geographic databases. The capability would also be present for future expansion of manual digitizing and for inclusion of automated digitizing equipment.

Another major purpose of a mid-level GIS system is the support of research. The modeling of natural and man-made systems is necessary to allow the mass of data potentially available to decision makers to be organized in a hopefully meaningful (dynamic) logical structure. The presence of a State controlled GIS would provide a common database for modeling efforts and potentially allow modeling efforts to be more useful than heretofore has been the case. University, contractor, and agency simulation, whether theoretical or practical, could all be consolidated in one place. (Contractor work could actually be done on the State system to guarantee proper performance of the software on the "target" machine and to monitor progress.) Many issues in theoretical cartography and spatial analysis are unanswered and a State GIS facility open to scholars would be beneficial to both parties. The research opportunities that would arise from any tag-along socio-economic database facility are likewise substantial. If the GIS's computer facility were also used as a means for routing technical (and other) information, a interesting opportunity would exist for research in organizational behavior and program evaluation.

5.5.2 Hardware and "System" Software Requirements for a Maine GIS Facility (mid-level)

Any computer system considered for the heart of the GIS facility should be able to pass the following tests:

1) Substantial computational capacity at low unit cost.
2) Ability to run very large programs as logical unities.
3) Ability to serve several simultaneous local and remote users.
4) Sufficient surplus capacity to allow a high-level language approach to programming and a minimum of systems programming.
5) Capability to run low priority monster jobs (such as polygon overlaying) in "background" at very low cost.
6) Unattended operation.

High performance is a must in the areas of graphics and simulation. The computer should be optimized for floating point computation. A fast processor, the availability of fast cache and scratchpad memory and the presence of user accessable writeable control store, 32 bit (vs 16 bit) ALU and memory data paths, memory inter-leaving, fast error-correcting main memory, and a high-speed floating point processor are all definite plusses. Several large mini computers come very close to this profile. Another requirement is the ability to execute a large program as a single logical entity. Geographic applications often require very large arrays that often can't be conveniently handled in parts and that won't fit in any resonable efficiency in a much smaller amount of real memory. The disk requirement would probably be met initially by a 40-80Mb drive, although substantial expansion may be necessary in time. Tape needs are for one 9-track 1600 bpi drive at first with the possible option of 800 bpi capability for compatibility with equipment at other sites. A 300 line per minute printer will suffice, with very long print jobs being handled at sites with high-speed printers after transmittal on tape. Some 300lpm printers also offer dot-matrix plotting for quick review plots and for CRT hard copy.

System software requirements include a virtual-memory time-sharing

System software requirements include a virtual-memory time-sharing operating system, language processors, utility programs, and (bending the nomenclature slightly) basic graphic, statistical & modeling packages. The operating system should be a single integrated facility for controlling all system hardware and software resources and for providing simultaneous access to these capabilities for several users who may wish to compile, edit, execute or debug programs in a variety of languages. The OS's support of virtual memory should not place any 64K limits on code or arrays, and the supervisor proprogram should provide for dynamic allocation of system resources, device-independent input-output, and a tree-like directory based named-file system. The operating system should privide password-controlled capabilities and file access for security against intentional or accidental intrusion or file destruction. Jobs should be able to be spawned by terminal users for batch execution, or they should be able to be triggered by time or the completion of other jobs. Unattended batch operation should be provided. User extensibility of OS features and control file capabilities are considered advantages.

The operating system should also possess an integrated set of utilities for sorting, merging, and copying files. A convenient software facility supporting the writeable control store hardware is necessary if the user is to be able to make use of this hardware. In addition, vendor commitment to industry-standard networking protocols will possibly provide future operational and maintenance advantages. Other system software requirements are for an Editor capable of handling large files and working from scripts, and for a symbolic debugger usefull to non-systems programmers.

Fortran is the language of choise for most GIS and related tasks at present. ANSI Fortran IV is adequate, but syntax equivalent to IBM G-level

Fortran is to be preferred. A macro assembler is needed for the coding of those few, frequently-used, routines which do not result in efficient code when written in a higher order language. PL/I, though not necessary, would be a definite plus if available by offering the potential for implementing such systems as Canada's CGIS. Likewise, PASCAL would be highly advantageous if it were available. It offers a nearly ideal means of dealing with cartographic data structures, tho its usual semi-interpretive implementation tends to run slowly. Statistical software needs can be met for the most part with SPSS, BMDP, and IMSL or SSP. Modeling efforts will require discrete, continuous, and mixed simulation capabilities which can be provided by (Fortran-based) GASP IV. Data Base Management Systems or MUMPS (especially if it runs under the regular operating system) may offer significant advantages in maintaining socio-economic and graphic data bases, generating reports from these data bases and providing a convenient interface between user programs and the system's data bases.

The computer system need not be an in-house one as there are significant potential economies to be derived from a shared system or from contracting for specific resources. A cooperative venture with University agencies with similar interests and with the capability to meet the State's GIS-related research needs might be worthy of consideration. Another possible means of savings would be to contract for such a system from a private source during normal working hours, and have commercial work run on the system during off hours. This would potentially reduce the unit cost and would avoid the politics of replacing an owned (but obsolete) system in five years. An off-site computer facility could be supported from in-house terminals over 4800 baud voice-grade phone lines or over hard-wired lines for an Augusta site in close proximity.

The graphics subsystem for a mid-level GIS implementation would consist of "conventional" computer graphics peripherals meeting the particular and sometimes demanding requirements of computer cartography. The basic components of the graphic subsystem include display, plotter, and digitizer.

The graphic display to useful in a cartographic context should offer a large viewing area (19" diagonal is ideal), high resolution (4096 X 4096 points), the capacity to display a large number of vectors (4000 minimum), and the ability to rapidly display changes in data being edited. Refreshed graphics displays offer more advantages in this respect than do storage tube systems, but the latter may function adequately if a very high speed interface is used. The plotter should be a drum or flat-bed "pen" plotter capable of plotting on paper or mylar with a minimum dimension of 34". High speed is not required in a plotter used for GIS work because the nature of the data being plotted seldom allows this high-cost capability to be adequately exploited. The digitizer in the system should be a large bed (36"X48" or 42"X60") back-lit unit operated on-line.

In addition to the actual graphics hardware, a capability should be included to allow preparation of plot tapes for off-site COM plotters, photo plotters, and inkjet plotters where their special features are required or conventional printing is desired.

The basic software required to support the graphic subsystem includes a standard graphics package (such as GINO-F, DISSPLA, or GCS) and device handlers for the various hardware components of the system. In addition, a contouring package should be procured and a series of graphic utilities should be written locally for editing, gridding, grid compositing, polygon generation from grids, and a number of other frequently used general functions.

5.5.3 Applications Software Requirements for a Maine GIS Facility (mid-level)

The general considerations in the initial development of a GIS applications software capability are to facilitate the production of <u>some</u> useful products at an early date and to preserve flexibility by not getting too locked in to an approach or software package. The overview of the subject that is sketched in 1977 is not going to be very accurate in a decade and it will be a hardy creature indeed if it is apt in five years. Adaptability and the opportunity to experiment with novel software approaches will be necessary until the technology slows its rate of change.

The computer consultant personally believes that the most advantageous choice of Geographic Information System is the USGS's LUDA (GIRAS) System. GIRAS is not the best system currently available. In fact it appears that in late 1977 significant portions of the projected GIRAS facility are (1) not yet implemented, (2) not yet integrated into the system, or (3) not yet ready for dissemination. Still, unless the LUDA program has fatal flaws, or unless Maine needs to move immediately, GIRAS should be the pick over other currently better-developed systems. GIRAS will receive large amounts of money and effort in the next few years & Maine can expect at least some Federal support in implementing and maintaining GIRAS. This system will evolve over the next few years into a well functioning entity or it will be scrapped and replaced by one that does work. In either case the package will be "supported" and will interface directly to the USGS developed data bases. If the State is able to attract quality personnel to its GIS effort and if it does not saddle them with administrative or production tasks it may be able to produce a superior home-grown system. This, however, is not likely to come to fruition. In-house GIS development & evaluation and implementation of other GIS systems be undertaken, though. The "best" system may come from any one of these sources in actuality. Prior opine should not carry more weight than the facts as they are reveallea in time.

The software for computer-assisted conventional cartography should be written in-house and should make extensive use of the system graphics package (GINO-F, GCS,...). This approach is potentially the quickest and neatest. Most of the dog work has already been done. Statistical programs will most likely be written in SPSS or in Fortran with BMDP or SSP routines patched in. These programs would be written in-house to meet local requirements. Similiarly, modeling software would undoubtedly be written by staff or consultants using GASP, Simscript, or Dynamo "tools" or possibly pure Fortran in an attempt to simulate Maine natural or human systems. "Imported" simulation programs (as opposed to building block utilities or general frameworks) should be used only with the greatest care.

5.5.4 Resource Availability for a Maine GIS Facility (mid-level)

The equipment profiled in the preceeding subsections is readily available from a variety of vendors. The computer system (processor, memory, disk, tape, communication interface and printer interface) is currently available from Prime, Digital Equipment Corp., and Harris. Other vendors such as Interdata, Texas Instruments, SEL, National Semiconductor, and Data General are likely to introduce systems in this class in the next few years. Upgrades to Hewlett-Packard's 3000 would allow it, also, to function in this role. Printronix offers the only 300 line per minute plain paper printer/plotter, but Tally and Centronics may be expected to produce similar equipment. Electrostatic printer/plotters are probably not cost-effective in this application although they do provide the dual functions. Graphics displays are available from Megatek, Tektronix, IMLAC and several other sources; appropriate plotters are offered by Calcomp, Houston Instruments, Zeta, and K & E amongst other vendors; and digitizers may be acquired from

Summagraphics, Talos, and Altek. Interfaces for integrating the graphics subsystem exist in the form of industry standard RS-232C and IEEE-488 interfaces and custom interfaces for particular mini systems.

System software is available from the computer system vendors or from software houses offering such products as SPSS, IMSL, GINO-F, DISSPLA. Other items in this catagory, such as GCS, SSP, and BMDP are in the public domain. GIS application software is readily available from the USGS and other public domain sources.

The availability of good personnel is the most serious constraint on developing a well-functioning GIS capacity. It will be difficult to assemble a compatible, appropriately skilled, core team at a respectable price. The minimum staffing for a Maine State facility would be a director-programmer, a programmer-operator, and a digitizer-key-puncher with one of the first two individuals having some knowledge of cartography. An initial narrow focus of interests might allow a group of this size to be productive, but volume work or a widened scope would require more personnel.

5.5.5 Operations of a Maine GIS Facility (mid-level)

Beyond the initial period of hardware integration and implementation of a basic GIS capability the activities of a computer cartography group would be dependent on the then extant priorities of the Executive. Some of the likely candidates for early action are (l) developing a computer-assisted conventional cartography capability, (2) putting up the entire Maine LUDA data base, (3) supporting 305 projects with special data base development, modeling, and reporting, (4) providing Forestry with a means of spatially and temporally analyzing spruce budworm data, and (5) creating industrial siting and development evaluation capabilities. The computer consultant suggests that any GIS capability be initially maintained by the SPO. The general planning perspective would provide a more appropriate environment for the development of a GIS facility than would the more restricted milieu of DEP, Geology, LURC, or Forestry. The needs of these other agencies should be initially met by a central facility under a cooperative agreement. The future may bring satellite systems, or even independent specialized units, but such alternatives should not be considered early in the development cycle.

5.5.6 Estimates of Costs and Associated Parameters by Source of Computing Resource (Mid-Level)

The following tables provide subjective estimates of the major cost components of a GIS facility. No total is given as that could be misleading if accepted at face value. Summing the major components would probably give reasonable approximations to true costs, but a much sounder set of figures could be derived by specifying the details left undefined in this hypothetical alternative.

COMPUTER SYSTEM (MID-LEVEL)
PARAMETERS

	COMMITMENT PERIOD	EQUIPMENT FACILITY PROCUREMENT METHOD	FIRST YEAR COMPUTING RESOURCE COSTS	SUBSEQUENT YEARS COMPUTING RESOURCE COSTS	POTENTIAL PROBLEM UPGRADE INCREMENTS	ABILITY TO DEMAND SYSTEM RESOURCES
CCS	?	CONTRACT	100K	100K	YES	NO
INHOUSE	LONG (5) 3 - 5 3 - 5	BUY LEASE RENT	100K 50K 65K	25K 50K 65K	NO	YES
UMO	3 - 5	CONTRACT	100K	100K	YES	NO
SSRI	3 - 5	CONTRACT	50K	50K	NO	POSSIBLE
PRIVATE OUT OF STATE	3 - 5	CONTRACT	100K	100K	YES	NO
PRIVATE IN STATE	3	CONTRACT	40K	40K	NO	YES

GRAPHICS SYSTEM (MID-LEVEL)
PARAMETERS

	COMMITMENT PERIOD	EQUIPMENT FACILITY PROCUREMENT METHOD	FIRST YEAR EQUIPMENT FACILITY COST	SUBSEQUENT YEARS EQUIPMENT FACILITY COST	SYSTEM UPGRADE
CCS	LONG (5) 3	PURCHASE LEASE	$ 50K $ 25K	$ 7.5K $ 25K	BY NEW PURCHASE IN MOST CASES
INHOUSE	LONG (5) 3	PURCHASE LEASE	$ 50K $ 25K	$ 7.5K $ 25K	BY NEW PURCHASE IN MOST CASES
UMO	LONG (5) 3	PURCHASE LEASE	$ 50K $ 25K	$ 7.5K $ 25K	BY NEW PURCHASE IN MOST CASES
SSRI	LONG (5) 3	PURCHASE LEASE	$ 50K $ 25K	$ 7.5 $ 25K	BY NEW PURCHASE IN MOST CASES
PRIVATE OUT OF STATE	LONG (5)	PURCHASE LEASE	$ 50K $ 25K	$ 7.5K $ 25K	BY NEW PURCHASE IN MOST CASES
PRIVATE IN STATE	LONG (5) 3 3	PURCHASE LEASE SHARE (Local site)	$ 50K $ 20K $ 12K	$ 7.5K $ 20K $ 12K	BY NEW PURCHASE IN MOST CASES

SOFTWARE (MID-LEVEL)
PARAMETERS
(one-time costs)

	"SYSTEM" SOFTWARE				APPLICATION SOFTWARE		
	OS & COMPILERS	GRAPHICS	STATISTICS	MODELING	LUDA	OTHER GIS	
CCS	$0	$0-15K	$0-2K	$ .5-10K	$0	NOMINAL, AND UP	
INHOUSE	$0-15K DEPENDING ON VENDOR	$0-15K	$0-2K POSSIBLE SPSS CONVERSION NEEDED	$.5-10K	$0	NOMINAL, AND UP	
UMO	$0	$0-15K	NONE, USAGE CHARGES FOR SPSS	$0-10K	$0	NOMINAL, AND UP	
SSRI	$0	$0.15K	NONE, USAGE CHARGES FOR SPSS	$0-10K	$0	NOMINAL, AND UP	
PRIVATE OUT OF STATE	PROBABLY NONE, USAGE CHARGES POSSIBLE	PROBABLY NONE, USAGE CHARGES LIKELY	NONE, USAGE CHARGES FOR SPSS, POSSIBLY OTHERS	$.5-10K	$0	NOMINAL, AND UP	
PRIVATE IN STATE	$0	$0	$0-2K	$.5-10K	$0	NOMINAL, AND UP	

PERSONNEL (MID-LEVEL)
PARAMETERS
(Annual costs)

	IN-HOUSE				SUPPLIER OF COMPUTER RESOURCE		
	PROJ. DIR.-PROGRAMMER-CARTOGRAPHER	PROGRAMMER OPERATOR SYSTEMS	DIGITIZER-KEYPUNCHER	OTHER	MANAGEMENT	SYSTEMS	RESEARCH
CCS	$18-25K	$13-18K	$7.5-10K	UNKNOWN	NONE	UNKNOWN	N/A
INHOUSE	$18-25K	$13-18K	$7.5-10K	UNKNOWN	N/A	N/A	N/A
UMO	$18-25K	$13-18K	$7.5-10K	UNKNOWN	LOW	$10K EST	ON PROJECT BASIS
SSRI	$18-25K	$13-18K	$7.5-10K	UNKNOWN	LOW	$10K EST	ON PROJECT BASIS
PRIVATE OUT OF STATE	$18-25K	$13-18K	$7.5-10K	UNKNOWN	$15-25K EST		NEGOTIABLE
PRIVATE IN STATE	$18-25K	$13-18K	$7.5-10K	UNKNOWN	$15K WITH THE POSSIBILITY OF REPLACING SOME IN-HOUSE PERSONNEL REQUIREMENTS		

SECTION VI

DETERMINING SPECIFICATIONS FOR A GEOGRAPHIC INFORMATION SYSTEM

Determining Specifications for a Geographic Information System

6.1. Introduction

The determination of system specifications occurs when the results of steps 1, 2, and 3 are deemed satisfactory to continue. Step 1 should set the ideal objectives for a system and steps 2 and 3 describe constraints or at least identify where additional work is required to overcome one or more constraints. System specifications generally consist of three main parts: identification of data elements, data formats, mediums for record archiving; a description of the spatial referencing system to be used; and information output requirements. The work to be summarized here is only initial as it results from only the first iterations of steps one, two, and three.

6.1.1. Data Specifications

Data specifications should consider balances between desired resolution, data availability, and costs of data acquisition. Desired resolution is a function of the minimum area unit that can be singularly recognized for use in the system, and the type and structure of the classification codes for the other data attributes to be associated with each spatial unit. Data availability and cost factors, if need be, can be adjusted to determine specific data elements, data formats, recording medium, and record format. This establishes data input requirements for the system.

6.1.2. Geographic Referencing

The next aspect of system specifications is a description of the spatial referencing system to be included. The concern here is with the structure of data for the entire geographic system and the need to have a number of referencing systems. Any spatially referenced data base has as its cornerstone a reliable planimetric base for geographic referencing.

6.1.3. Information Output Requirements

The information output requirements are concerned with the manner in which information can be derived from the system and presented to the user. This includes all necessary data manipulation and analyses required to transform stored data into useable information. All functions must be specified in such a way that it is very clear how each desired output is obtained. Attention to detail will prevent potential difficulties in development and use of the system. Test analyses should be conducted to determine feasibility and preliminary cost information.

The three parts of system specifications must be evaluated in concert to determine the extent to which original objectives will be met. This allows the explicit determination of system boundaries. System boundaries refer to the definition of the system in terms of several variables such as purpose, users, subjects included and total geographic extent. In situations where a defined users group is not well recognized in the design process, system boundaries result from interaction of the system sponsors and system designers (Figure 6.1). The major system boundary variables are:

1) type of system refers to the dedication of the system (i.e., totally operational or partially demonstrative and research oriented).

2) geographic area

3) subject - the general categories of data considered appropriate for the system such as land characteristics, economic, demographic, hydrologic, data, etc.

4) object - performance for the user including types of analyses and geographic displays.

6.2. Specification Decisions and Decision Variables

It is apparent that there are many decisions, both technical and administrative, which are made during the design and development of a geographic information system. Decision variables are

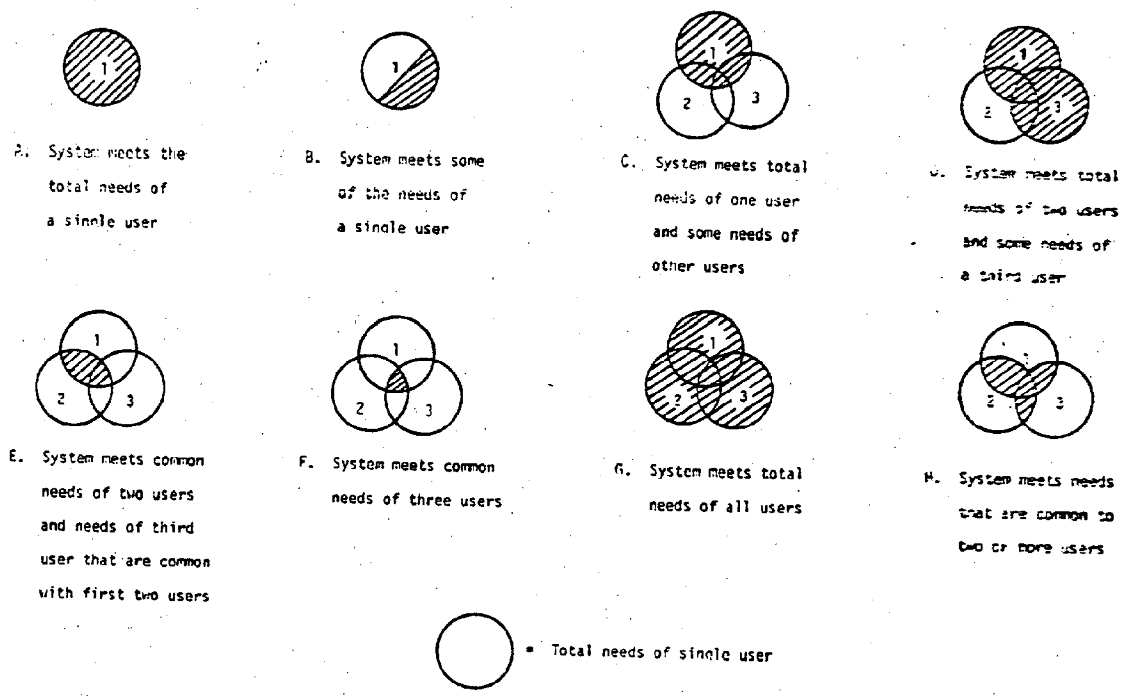


Figure 6.1
System–Users Relationships

(source IGU 1972)

a tool to examine the scope of these decisions in a relatively ordered way. Such decisions generally contain a number of options. As well, before identifying alternatives a procedure must be developed to reduce the number of possible alternatives without rejecting any major alternatives. A recommended process is to group the variables on the basis of interdependence and then evaluate each group separately (see Figure 6.2 and 6.3).

Figure 6.2. Example of Grouped Decision Variables

Data Volume and Handling

- scale
- data storage medium
- file structures
- software
- hardware
- automation

number of alternatives for each element

Such a systematic grouping is strongly recommended because it forces explicit examination of each decision variable and if executed it assures no major alternatives will be overlooked. Failure to use such a process can easily result in an early decision, possibly regarding scale, that becomes overly binding on further system development.

6.3 The Continuing Process

The final step of part 1 is a complete evaluation of data specifications in terms of effectiveness, consistency and feasibility. Effectiveness is simply evaluated by a comparison of final data specifications to system objectives. A consistency evaluation checks that initial system specifications are internally consistent and that they represent a solution capable of meeting objectives. Consistency between geographic referencing capabilities, data storage specifications, analysis specifications, and information delivery is the desired end result. Feasibility assessment is a broad somewhat subjective evaluation of the reasonableness of the specifications particularly with respect to data availability constraints and overall cost estimates for data acquisition.

To summarize stage 1 of the design process, the desired result is a set of consistent specifications representing system objectives. These specifications are viewed as being satisfactory for use in the generation of system alternatives. If specifications are unsatisfactory more iterations of stage 1, which this report has outlined a first iteration of, will be required.

Although a system design should not be totally constrained by available data, systems will of necessity be constrained by the availability of geographic referencing systems. So system objectives and specifications which have been developed will not only represent an ideal system but rather reasonable objectives warranting further evaluation. Some specific products of stage 1 should be:

- Data definitions
- Geographic referencing system design
- Specifications for data, retrieval, analysis and information delivery
- Cost estimates for data acquisition, and
- Subjective and explicit evaluations by a large segment of the potential user community of the work done to date.

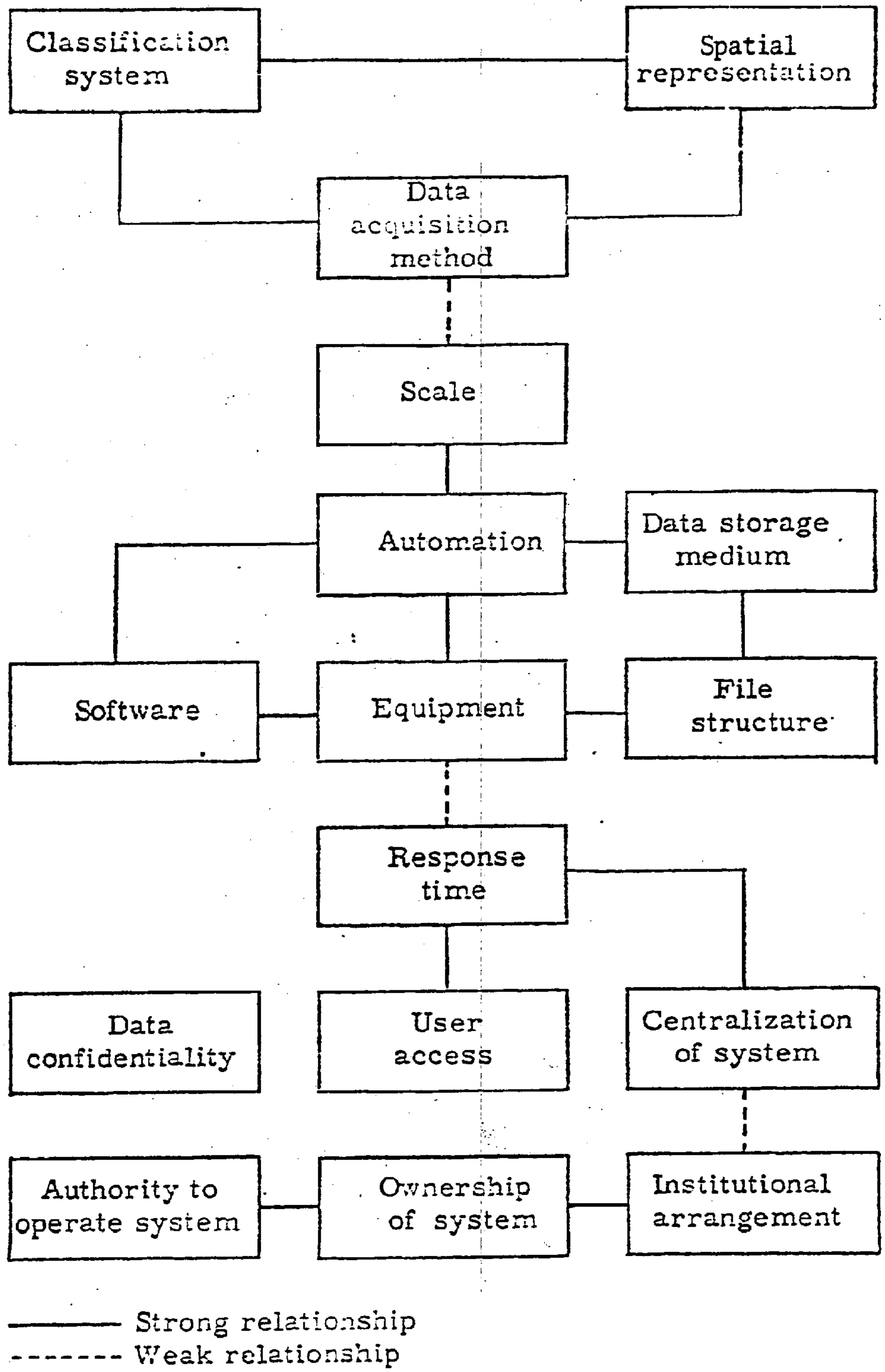


Figure 6.3 Grouping Decision Variables

(source U.S.G.S. 1975)

6.4

In summary, Phase 1 involves a long iterative design process which is probably best accomplished by an in-house user oriented process and not by the extensive involvement of outside system consultants. A state data and information planning program is viewed as an ideal vehicle for this process as such an effort integrally involves data gatherers, data users, and data user needs assessment.

6.5. Structure and Evaluation of System Alternatives

When the products resulting from Phase 1 are deemed satisfactory and the data required to meet these objectives have been identified and documented in terms of data elements, data formats, record medium and record format, alternatives to accept, store, manipulate and output can be specified and evaluated. For data handling this investigation includes factors related to software, hardware, operating policies and legal implications for each alternatives. Evaluation of alternatives is done to determine the technical feasibility of the data handling procedures and their associated costs. The formulation of alternatives is accomplished by selecting one decision variable value for each dimension of the system and if the resulting combination represents a feasible solution then the alternative is evaluated in greater detail.

After feasible alternatives have been identified, each alternative must be detailed in terms of hardware, software and an institutional framework for system operation. There are likely to be more than one possible solution per alternative as hardware requirements can be satisfied by a number of different machine configurations. Relative advantages of different hardware, software, and institutional framework will depend on factors such as time availability, information availability on the alternatives and the level of analysis which is underway.

The selection of the system to be implemented considers the feasibility and cost evaluations in terms of system specifications which have defined data needs and associated manipulation requirements. The alternative that best satisfies specified needs at the lowest cost should then be identified. If no alternatives are found to be satisfactory then either new alternatives must be constructed or system specifications may be modified. The decisions regarding change should depend upon the factors that contributed most toward rejection of an alternative.

6.6. Summary

In summary this general design process is based on a systems analysis approach (see Calkins, 1972, I.G.U. 1972, USDI 1975, 1977). It does assume each task should be completed in an explicit manner and that one task leads directly into the next. The system should be best utilized in an iterative manner with greater precision desired in each successive iteration. It must be noted, however, that due to extreme difficulty in obtaining good data for the various steps, it will often be necessary to continue with less certainty than this process with all its flow charts appears to offer. The only real advantage of such a process is that it forces greater attention on details where otherwise broad implicit assumptions would be made and it requires documentation of every step, decision, and investigation, which has been the thrust of this work and should be the thrust of the continuing GIS effort.

This report has presented a first attempt at determining system specifications. As outlined in Chapter II, this report only covers phase 1 of what is viewed as a three phase developmental process. Many of the inputs which would be required for true specification generation were not yet available so much of this work is left for later iterations of the various stages of the process.

FINDINGS AND RECOMMENDATIONS

This section will be prepared in cooperation with a subcommittee of the Land and Water Resources Council after the report has recieved a full review.

GLOSSARY OF TECHNICAL TERMS*

ACCURACY

The extent of freedom from error. That is, the closeness of a measurement to the true value, or to a value accepted as true.

ACCURACY, OVERALL

Accuracy over the entire range of an instrument.

ALGORITHM

A computer-oriented procudure for resolving a problem.

AUTOMATED CARTOGRAPHIC SYSTEM

Automated methods of producing maps and charts, in graphic and digital form, in order to reduce production time.

BACKGROUND PROGRAM

A program being operated upon by a computer system at low priority usually during times when the CPU is not in demand for interactive operation.

BLOCK

The physical unit of information within a particular data set.

BORDER

Exact term for the division between two mapped areas which is interior to the subset being bounded. Synonym: Boundary.

BOUNDARY

General term for the division between two mapped areas. Synonym: Border, Edge.

*Taken partially from Amidon, E., in I.G.U. Geographic Data Handling

CADASTRAL SYSTEM

System containing information on parcel boundaries by ownership of property.

CARD IMAGE

A representation in storage of the hole patterns of a punched card. The holes are represented by one binary digit and the spaces are represented by the other binary digit.

CELL

The smallest region in a grid.

CHARACTER RECOGNITION

The assimilation and analysis of pictorial information relative to numeric or alphanumeric characters.

CHOROPLETH MAP

Map showing discrete areas such as states or counties. These units are considered uniform with respect to the statistics collected within them.

COMPILATION

(A) Selection, assembly and graphic presentation of all relevant information required for the preparation of a map. Such information may be derived from other sources. (B) A graphic document produced by this process.

COMPILER

Computer program to translate a source language into an object language. or

Operator of photogrammetric machine who produces a two-dimensional diagram from the stereo-photographs. or

The cartographer who carries out the process of map compilation.

CONFIGURATION

The specific arrangement of system computer hardware and peripherals making up a system.

CONTOUR

Line joining points of equal vertical distance above or below a datum.

CONTOUR PLOTTING

The process of plotting (or drawing) contour lines from a data file consisting of a set of points representing a contour line. Synonym. Contour Drawing.

CONTOURING, AUTOMATIC

The process of calculating and plotting (or drawing contour lines with a computer and peripheral devices) using a set of points representing arbitrarily selected positions on the topographic surface, and defined according to an arbitrary coordinate system.

CORE

The fastest access storage part of a digital computer usually utilizing magnetic cores.

CRT DISPLAY

Cathode ray tube display.

CURSOR

Aiming device, such as a lens with crosshairs, on a digitizer.

DATA BANK

An information store usually in digital form organized in such a manner that retrieval and updating can be done on a selective basis and in an efficient manner.

DESCRIPTOR

The catalogue descriptive part of digitally stored cartographic data

or

The keyword used in library information systems.

DIGITAL

Representation of data in the form of bits. Contrast with analog.

DIGITAL IMAGE

A two-dimensional matrix which represents an area on a photograph. Each position of the array is assigned a grey level, which may be limited to two. Synonym: Digitized Image, Digital Picture Function.

DIGITIZATION, AUTOMATIC

The process of conversion of analogue or graphic data into digital form using automatic processors such as ALF, character recognition, pattern recognition, or scanning.

DIGITIZATION, MANUAL

The process of conversion of analogue or graphic data into digital form by an operator with or without mechanical or computer aids.

DIGITIZE

To use numeric values to represent data.

DIGITIZER, GRAPHIC

Machine that changes graphic cartographic information into a digital format for computer input.

DIGITIZER, LINE_FOLLOWING

Device which automatically tracks an individual line and at selected intervals digitally records its position with respect to an arbitrary coordinate system. Synonym: Automatic Line-Following, A.L.F.

DIGITIZER, POINT

A manually controlled cursor senses position, usually by electromechanical means. An operator must activate the recording of positional elements or other information.

DISPLAY

Any graphic presentation in hard-copy or as a transient image or a device (usually CRT) attached to a computer for the rapid display of selectable information in map or list form.

DISPLAY, LINE DRAWING

A display system (usually CRT) which produces an image from lines drawn as a series of dots or vectors on a screen.

DISPLAY, RASTER

A display system (usually CRT) which produces an image from scan lines in a raster format. Usually used with refresh type displays.

DISPLAY, REFRESH

A display system (usually CRT) which produces a transient image and which must be refreshed about 60 times/sec., in order to prevent flicker. Comparable with a standard TV display. Limited in data content because of time.

DISPLAY, STORAGE

A display system (usually CRT) which produces a stored image at high speed on a screen and which will remain unless erased for a period of at least one hour. Not limited in data content except from aspect of visual clutter.

DRAFTING UNIT

A digital computer-controlled X, Y mechanism capable of drawing lines to high accuracy.

DRUM, STORAGE

A rotating magnetic drum storage division for digital information used for relatively fast date access in computers.

EDITING

Detection and correction of the data obtained in graphic data reduction.

ENCODER

A device for converting linear or rotary mechanical motion into precise digital form. Widely used in digitizers and for position feedback for drafting units.

FILE

A variable number of records grouped together and treated as a main division of data.

GEOCODING

Geographic referencing or coding of location of data items.

GEODETIC COORDINATES

Latitude and longitude with reference to a standard spheroid.

GEOGRAPHIC BASE FILE

Coded network.

GRID COORDINATES

Euclidean coordinate system in which points are described by perpendicular distances from an arbitary origin.

HARDWARE

The physical components of a computer and its peripheral equipment. Contrasted with software.

ISLAND

Single-line boundary within a polygon.

KEYBOARD

A manual input device for function operation or alphanumeric entry.

LIGHTHEAD

Lightspot (and possibly symbol) generator - used by high accuracy drafting machines to record lines and symbols on photosensitive material.

LIGHT PEN

See Interactive Positioning Device

or

A device the size of a ball-point pen which is used for pointing to a location on a CRT screen. The coordinates of the location are obtained usually from a time lapse measurement within the raster image formation.

LINE PRINTER

A peripheral device for computers which prints a line at a time. At each position along the line a set of alphanumeric characters is available. Maximum line length varies by manufacturer, but, usually 120-130 columns are available on high-speed printers. It can be used for high-speed listing or, by spacing symbols, as a plotting device.

MAGNETIC TAPE, DIGITAL

A method of storing data by selective polarization of the surface of a ferrous-coated tape. A large reel of magnetic tape (2400 Ft) will store a large amount of data but sequentially.

MERGE

Combine two sequenced files into a single sequenced file.

or

Joining two or more lines and areas together. When this operation is not simple a combination of deletion and interpolation may be required to present a pleasing appearance.

MINICOMPUTER

A low cost computer with limited core capacity. Widely used for device and system control and data handling when large computations are not involved.

OFF-LINE

Processing is not directly under the control of the central processing unit.

ON-LINE

Processing is directly under the control of the central processing unit.

OVERLAY

Map of an area to be superimposed on one or more maps of the same area. The purpose is to find data combinations, or more exactly intersections and unions

or

Digital image of areas as in definition one above.

OVERPRINTING

Superposition by successive printing of line-drawn or continuous-tone information.

PLOTTER

An X-Y mechanism controlled by a computer generally for the recording of location information, e.g. symbols, names, etc. Line drawing may also be carried out but units capable of high accuracy line drawing usually are referred to as drafting units. Lines are drawn as a series of vectors.

PLOTTER, DRUM

A plotter where the sheet material is transported by a rotating drum which provides the motion for one axis of the mechanism by a forward or backward motion. Pens are mounted on a bar parallel to the drum axis. The bar movement provides the second axis.

PLOTTER, FLATBED

A plotter where the sheet material is fixed on a flat table survace and the pen or printer is carried by a gantry and trolley mechanism providing two axes of motion.

PLOTTER, LASER

A plotter in which line vectors are drawn by a laser beam deflected by galvanometers.

POLYGON

Plan figure consisting of three or more vertices (points) connected by line segments or sides. The plane region bounded by the sides of the polygon is the interior of the polygon.

RANDOM ACCESS

Process of obtaining information from or placing information into storage where the time required for such access is independent of the location of the information most recently obtained or placed in storage.

RASTER SCAN

A line-by-line sweep across a display surface to generate or record an image.

SCANNER

Any device which systematically breaks up an image into picture elements (or pixels) and records some attribute of each picture element.

SCANNER, DRUM

Apparatus which scans and records 2 or moregrey levels, usually of reflected light from a picture fastened to a rotating drum.

SCANNER, FLYING SPOT (CRT)

Apparatus which scans and records grey levels, usually of transmitted light, by electronic means.

SCANNER, LASER

A scanner plotter with the exception that the light source is replaced by a laser to give very fine resolution. Total map production time is normally increased appreciably.

SCANNER, PLOTTER

A device which normally consists of a continuously rotating drum mechanism carrying and photosensitive material and a variable intensity light source which linearly traverses the length of the drum.

SEGMENT

Subset of consecutive polygon points. Synonym: Link, Arc.

SEQUENTIAL OPERATION

The performance of actions one after the other.

SMOOTHING

Filling a line of observed data points by a continuous line.

SOFTWARE

Porgrams used to control the operation of computers, originally this term included program compilers but was extended to programs in general and even to reports.

SUBROUTINE

A computer program which uses data or instructions from another program. A means of dividing a large program into smaller routines and which may be used without modification in diverse applications.

UNIFORM GRID

Square rectangular or, more rarely, hexagonal lattice for recording geographical data. The simpler grids are usually not related to geodetic coordinate systems.

VECTOR

A linear line segment, normally short, used to construct any line form on a plotter, drafting unit or display.

WINDOWING

A method of designating and separating out a particular area of map data for presentation on a display.

APPENDIX B

A REVIEW AND EVALUATION OF SELECTED CURRENT AND DEVELOPING STATE GEOGRAPHIC INFORMATION SYSTEMS

I. The New York Land Use and Natural Resources Inventory (LUNR)

1. General Description:

The Land Use Natural Resources Inventory is a computerized record of an aerial survey of New York's land resources. Supported by retrieval, analysis, and display computer programs it is the most well established state-wide land information system in the United States.

The LUNR system contains 130 land use categories and four categories of supplemental data for the entire state. This data was manually encoded (grid overlay), then converted to automated records and stored as a sequence of descriptors for cells one kilometer by one kilometer. Two sets of computer programs are available for retrieval and analysis of stored data. The first, DATALIST provides tubular summaries of raw data and limited statistical analysis. Plan MAP IV, the second outputting programs, has sophisticated analysis capabilities and produces computer graphic maps.

2. LUNR History

LUNR began from a formal pledge by Governor Rockefeller in 1966, to provide an inventory of the states natural resources. At this time the Center for Aerial Photographic Studies at Cornell University had just completed an inventory of the resources of the north-eastern section of the state using air photo interpretation. The data derived from this study had been coded and stored in a computer using the SYMAP program. Based on this experience Cornell was contracted to develop LUNR.

Prior to the development of LUNR, researchers at Cornell had gained considerable experience with gathering and analyzing land resource related information for large areas. Redevelopment of LUNR was reviewed as an opportunity to develop techniques applicable to similiar projects and problems regardless of particular geographic locations. However, final system operation was required to be independent of Cornell expertise and workable utilizing commonly available techniques and equipment. Its concept was not one of an ultimate geographic information system but one of operational applicability currently before more sophisticated techniques are available on a practical basis. As well as in New York, the system has been applied in Puerto Rico, El Salvador, and Colorado.

State Planning first required performance of a five county pilot study. The pilot study purpose was to test classification systems and do initial computer programming. Initial plans called for the application of SYMAP to store and retrieve the collected data base. The pilot study showed this program to be incapable of efficiently processing large volumes of data, so a new set of programs was devised.

An initial land use classification scheme of six basic land use grew to a list of 130 land use types after interviewing potential users from state agencies, planning organizations, and universities.

The LUNR Land Use Inventory was conducted between 1968 and 1970 for a cost of $500,000. This broke down to a cost of $10.00 per square mile. The federal government, dispersing Appalachian development funds, paid over 701 of this cost. Once the inventory was completed a statewide users service was established to serve the general public and local planning and governmental organizations. Two separate operations are involved, one at Cornell and the other at State Planning in Albany.

3. LUNR System Description

1. Data Acquisition Compenent

Aerial photography was the data source for of the data in the LUNR system. Panchromatic photography of a scale 1 in. = 2,000 feet was used to map the entire state. 17,000 9x9 prints were required to map the entire state. Those were combined into 129 larger photo mosaics.

About 151 of the land use data required some type of field checking. Supplemental data such as highway access were recorded in this way.

Four other supplemental data classes were developed for incorporation into the LUNR system. This section consists of four general parameters which were geographically referenced to the 1 km Z grid system. The data are:

1. General soils map of the states which was derived from a map prepared by USDA. This map contained 214 soil series. Resolution is limited to units of greater than 300 acres. A single LUNR inventory cell occupies 247.1 acres. A single value - the dominate soil mapping unit - was recorded for each cell except where less dominant occurred often in an area but were masked by the dominant soil of each cell. When this occurred a representative number of cells were also encoded with more than one soil type.
2. The general geologic map of New York which contained 152 bedrock categories was collapsed into a map containing only 11 functional mapping units. These units were interpreted as to their basic characteristics and capabilities. As with the soils maps, only dominant rock type for each cell is recorded.
3. The third map, land form and bedrock depth, indicated land form surficial geology types and over burden depths for only a portion of the total statewide coverage (based on development pressure). Data values were recorded as percentages of each kilometer grid cell.

The final data type, Agricultural Economic Viability was prepared on the basis of interpreted soil data, topography, climate and water resources. Parameters for locations markets, access and social cultural viability were also recorded. Three levels of viability were recorded and entered into the system.

3.2 Data Inputting

As previously outlined, the system uses 1 km Z cells related to USGS quads. This system resolution is 1 km ^{2}. This is recognized as overlay gross for urban planning but rural planning is the central focus of the entire system. At this level 140,000 cells were required for statewide coverage. Each has an X and Y coordinate which is referenced to the southwest corner of the state.

In terms of map preparation three overlays were required for land use.

(1) Area overlay-polygons of uses were outlined and 1 area of each cell was estimated. 1 acre is the smallest unit of mapping for source data inputting.

(2) Point overlay - consisted of small features which were noted as present or not and streams were valued by length per cell.

(3) Compilation overlay include all minor civil division and roadway classes and lengths per cell.

Coding sheets were used (one per cell) for each cell and land uses plus the supplemental data types were punched onto cards and then merged onto disc storage. Coding for all for geographic referencing was by coordinates, counties, minor civil divisions, and watersheds.

LUNR data is stored on two direct access disc packs. One disc contains the land use data and the over contains the four categories of supplemental data. There is storage presently available for 200 data items per cell and this capacity can be expanded to thousands per cell.

Data Retrieval, Analyzing and Output

Two types of retrieval and analysis programs are developed specifically to provide easy data access so unskilled users can retrieve desired data and perform sophisticated analyses without the assistance of trained programmers.

DATALIST provides tabular summaries of raw data or analyzed data. Output is by cell. Ten data categories can be examined per computer run. DATALIST is useful in statistical referencing providing an inexpensive means for matching cells and data characteristics. However, it cannot recognize patterns or identify adjacencies. Addition, substraction, multiplication, division, and corss tabulation are the primary mathematical functions provided.

PLANMAP programs produce line printer maps which show up to ten visual density displays produced by overprinting. Steps in PLANMAP analyses are study area delineation, assignment of weights to data values, and excluding unnecessary data elements from analysis. Once established data are divided into final weighted values with up to ten different data levels. In addition some more specific optional analysis steps are: conditional weights (if then statements), fixed weighting ranges, scale factoring, decimal setting, and a number of format commands for map and tabular display.

4. System Use

The overlay DATALIST and PLANMAP have all been used in the day to day activities of the State Planning Office in New York Regional agencies and communities have as well used the system extensively.

Some specific applications have been:

Micropatterns study which examined the impact of a range of proposed state facilities in central New York and then to develop recommendations concerning the role of the State Planning Office in facility siting decision making.

The Appalachian Area Resource Study utilized the system for providing statewide data for New York.

State Parks and Recreation uses the system to conduct surveys of potential recreation and openlands in conjunction with the preparation of statewide outdoor recreation plans.

Others have been reservoir siting, market studies, geologic feature identification, nuclear power plant siting and new towns siting.

5. Future Plans

Although early plans called for periodic updates it now seems likely that an entirely new data base and storage and retrieval system will be developed. An interagency task force investigating state spatial information needs recommended a system with greater resolution be tailored to fairly specific user needs and be regularly updated.

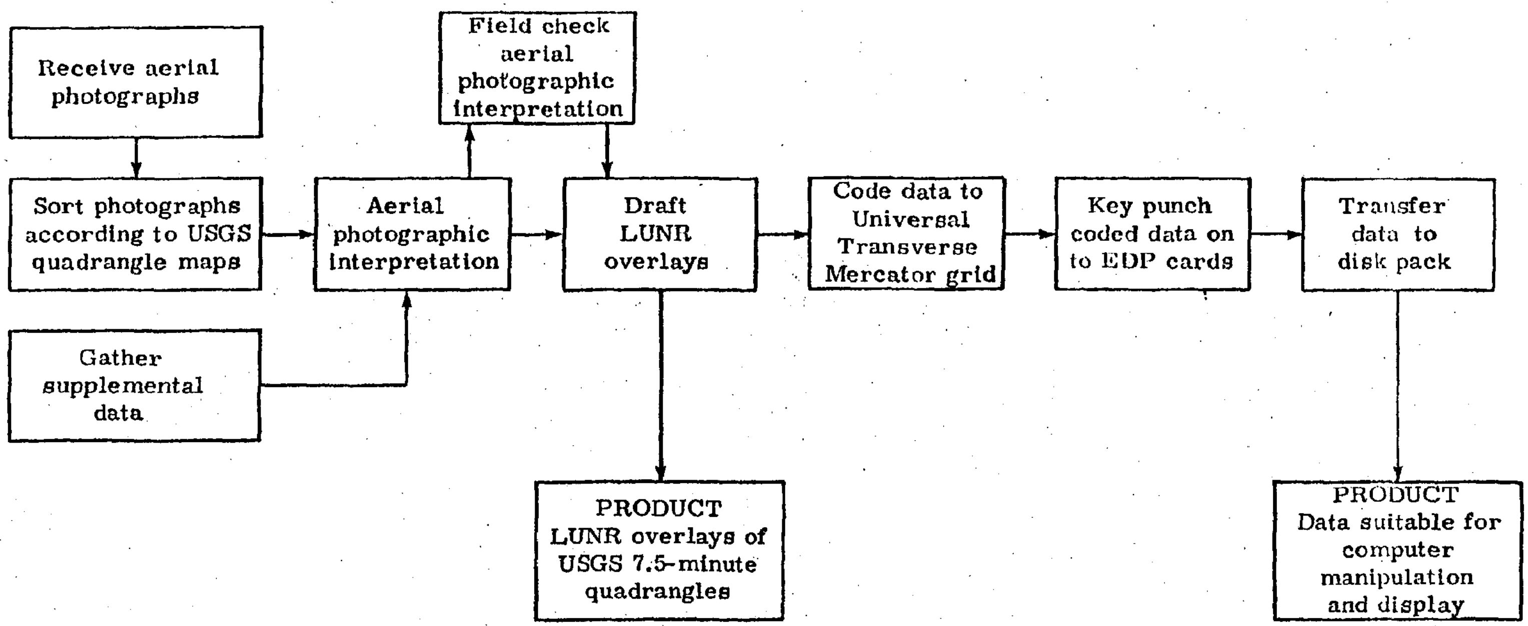


Figure 1. LUNR Procedure Summary

II. The Maryland Automated Geographic Information System (MAGI)

General Description

The MAGI System stores geographic data in a consistent and coordinated manner similiar to integrated base mapping systems. "The data bank and associated data handling system developed are utilized to generate computer maps displaying the capability and suitability of the land for various uses across the geographic context of the entire state of Maryland.

The input utilized to generate these composite maps are numerically weighted map variables collected and stored in a geographically referenced data base. The computer map outputs are used in a variety of ways which generally relate to the development of a Generalized State Land Use Plan. Maryland planners are currently utilizing the system in the interpretation of photographic and satellite data from ERTS I.

For the state the objectives of MAGI go beyond initial resource inventory, data base managment, and analysis of landscape potentials. The system is hoped to lead to the establishment of a basic structure for continued state wide integration and analysis of numerous types of geographically based statistics (i.e., census and socio-economic data as well). It is recognized that the system will never replace all forms of manual data handling but will continue to operate as a tool for increasing analysis capabilities and decreasing the costs of statewide data storage and retrieval.

MAGI History

The system was designed and implemented by the Environmental Systems Research Institute of Redlands, California under the joint sponsorship of the State of Maryland and the National Aeronautics and Space Administration. Work was begun in 1971 and the system was operational in 1974.

Impetus for the development of a geographic information system was legislation creating the Maryland Department of State Planning and assigning to it the responsibility for preparing and continually updating "plans for the development of the state, which plan or plans collectively shall be known as the State Development Plan."

Realizing the need for the systematic development of such plans and the requirements for extremely large scale data storage and analysis operations the Department of State Planning initiated six general efforts leading to a state wide planning information system. These efforts are as follows: l)the selection of the significant geographic indicators which influence intelligent decision making for land use planning; 2)the collection of existing base maps which described important geographic differences in these indicators; 3)the design of automated computer filing systems for this geographic data; 4)the design and development of computer programs and procedures for analysis, retrieval and updating of computer encoded data types; 5)the development of computer models for the analysis of data and; 6)the development of in house capabilities for operating the system; specifically training state personnel in complete system use.

MAGI System Description

I. Data Acquisition Component

At the time of the project's initiation the state had already maintained a substantial amount of high quality data for state level planning. This data was checked for its validity, level of applicability, timeliness, comprehensiveness and flexibility in potential use. Each source agency was asked to identify the idiosyncrasies associated with their data and present a complete analysis of their perceived data limitations and usefullness.

Data was collected from many sources. Eachof these sources maintained data of varying quality and age, with incompatible scales on different base maps. Because of budget constraints this data was not always remapped on uniform base maps but each series was referenced tothe State Coordinate Grid System and reformatting occurred subsequently during the digitizing and data processing stages of the system.

A total of fifteen landscape variables were selected and gathered for the entire state. They were selected according to their applicability to three categories:capability analysis;suitability analysis;and special studies analysis.

Utilizing the definition that capability analysis is the "ability of a resource to support various activities,and various levels of activities because of inherant physical activities, " soils, geology,aquifers and aquifer recharge areas,mineral resources,topography,natural features, vegetation and surface hydrology were selected as primary capability variables.

Suitability is defined by Maryland planners as an "entity,usually manmade,induced or influenced that allows one to indicate preference as to a resource supporting various levels of activity given the capability of alternative sites or areas. Publically owned property,sewer and water facilities,transportation systems,historic sites and existing land use were identified as primary suitability variables to be included into the system.

In addition to the thirteen capability and suitability variables several other data items were collected and digitized for special purposes. Watershed boundaries were included to be used in the mapping and analysis of statistical data generated for surface water in Maryland by U.S.G.S. and EPA. Electoral districts were included because census tract data breakdowns are nested by electoral boundaries and state tax and land use records are also maintained at this level of spatial aggregation.

The fifteen data variables came from over thirty data sources and were compressed into thirteen sets of maps covering the entire state. Mineral resources and aquifers and aquifer recharge areas were mapped together as were electoral districts and watershed boundaries.

All base maps were at one of three scales-1:62,500,1;63,630,and 1:126,720.With the exception of land use all maps variables were mapped and eventually digitized on a county basis.As statewide analysis required some type of coordinate referencing system for the eventual merging of separate county data files a localized version of the State Plane Coordinate System which utilizes Lambert Conformal Conic Projection or Universal TransverseMercator was adopted.Each of the thirteen variables on the county maps were referenced with this system.

2.Data Inputting

Two systems for digital encoding were employed-manual encoding and automatic digitizing. For the manual coding a grid was drawn at the appropriate scale for the different base maps but woold correspond to 10,000 ft. planar lines was drafted and overlayed for referencing.A smaller grid at a scale of 2,000 ft. per encoding cell was drawn on the overlay.The rows and columns for each cell was then numbered.For the data types mapped as polygons,the recording for each cell was determined by the polygon which occupied most of the area of the cell.Land use included encoding the secondary and tertiary polygons and soils the secondary polygon.Data mapped as lines were recorded by determining the most important qualities occupying a given cell.For example, if two roads of varying size filled the same cell then the larger road code was assigned.Map data expressed as points(i.e.,natural and critical areas) was encoded by row and column of the cell and number coded for individual retrieval.

All variables except for watersheds and electoral districts were encoded by manual techniques. These maps were digitized as polygon source maps.Once digitized a special program was utilized to convert this polygon data to grid data to maintain compatibility with the files which were manually encoded.Because of relatively large polygon size for these data variables this procedure saved a considerable amount of time in cell by cell coding and editing.

3.Data Storage

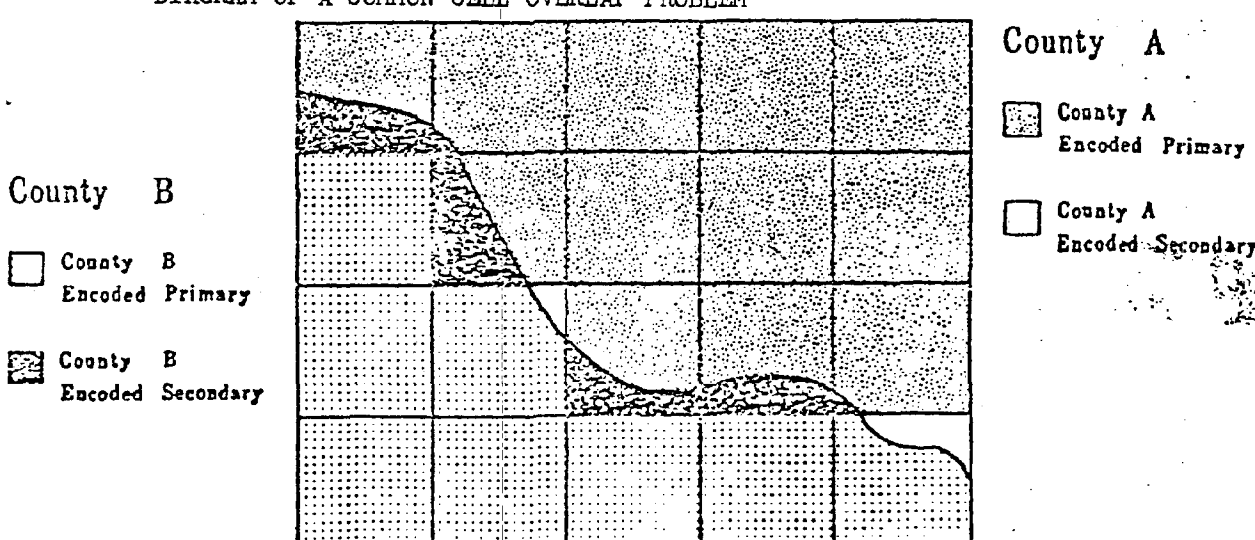
-Single variable county files were formed after the first edit was performed by producing maps for each file and systematically comparing it with base data. It generally required three complete edits to produce clean single variable data files.

-Multi variable county files were formed after the final edit of the single variable files. Single variable files were combined to form a multi-variable file for each county in the state. In order to check for a direct overlay a special mapping routine was applied which would delineate valid data mismatches, drafting errors and invalid mismatches of data variables. Such errors generally resulted from poor grid alignment on the maps which were encoded at different base scales. Such errors once identified were easily corrected.

-State-side files were created utilizing a different procedure because creation of the county files only required a simple combining of different data items the merging of counties required more sophisticated software to insure proper spatial placement of adjoining but still separated map boundaries. Decisions concerning coding were required when adjoining data which would reside in the same cell once the state -wide file was formed. For example an overlapping cell could be encoding one way on one county map because of the dominant polygon and encoding something completely different on an adjoining county map because of the presence of a different dominant polygon.

Figure 1

DIAGRAM OF A COMMON CELL OVERLAP PROBLEM



A computer program(MAPMERGE) was applied to permit the accurate adjoining of bordering data record. The running of this program for construction of the state file initially involved the assembly of county files into regional area files which corresponded to standard regional planning land use districts in the state. The regions were then merged into a final single state-wide file containing:

TABLE 1

Variable #	Variable Name	# of Digits
1	Grid cell row number	3
2	Grid cell column number	3
3	county number	2
4	surface water quality	2
5	engineering geology (primary)	2
6	engineering geology (secondary)	2
7	transportation facilities (primary)	2
8	transportation facilities (secondary)	2
9	ownership (primary)	5
10	ownership (secondary)	5
11	mineral resources	2
12	sewer and water districts	2
13	vegetation	2
14	soils (primary)	3
15	soils (secondary)	3
16	soils (tertiary)	3
17	natural features (primary)	3
18	natural features (secondary)	3
19	topographic slope (primary)	1
20	topographic slope (secondary)	1
21	watersheds and sub-watersheds	4
22	electoral districts	2
23	historic sites (first in cell)	6
24	historic sites (second in cell)	6
25	historic sites (third in cell)	6
26	land use (primary)	4
27	land use (secondary)	4
28	tertiary land use (1973)	4
29	primary county comp plans	6
30	secondary county comp plans	6
31	tertiary county comp plans	6
		105

Data Retrieval, Analysis and Outputting

The MAGI system was developed with the capability of producing computer maps for editing data as well as final display graphics. Because of the large number of data variable classes for certain variables(i.e.,soils)it was impossible to graphically display all variations on a single map. For final modelling and presentation it was decided that all maps would callapse sub-classes into ten or less final mapped variations. An example of this procedure is the aggregation of thirty soils series into ten new soils classes based on agricultural productivity for modelling any information which required productivity data in the procedure.

Most of the models developed as integral parts of the system were descriptions of the suitability and capability of the land to support various land uses. The outputs of the various models were generally series of grid maps graphically defining the capabilities of the state for various potential activities. Each cell can be mapped with a value which expresses weightings assigned to one or more characteristics of the cell. Weightings were established by estimating constraints to and amenities for each data variable to all other variables and finally relative to the specific activities being modelled.

System Use

To date the system has been primarily used within the state planning office. However a number of specific projects have been able to incorporate system analysis and output products into their overall frameworks. For example:

1. Mapping mining and extraction potentials for the entire state.
2. Identifying all agriculturally productive lands-both actual and potential.
3. Industrial location siting studies.
4. Mapping residential development suitability and capability for rapidly developing areas of the state.
5. Preparation of a state-wide comprehensive recreation plan

	DATA IDENTIFICATION & ACQUISITION	ANALYSIS & REFORMATTING	ENCODING & EDITING	LIBRARY REFERENCING SYSTEM	PREPARATION OF MATERIALS	DATA MANAGEMENT	COMPUTER SOFTWARE DEVELOPMENT & DOCUMENTATION	MERGING & CORRECTION OF DATA FILES	FINAL IMPLEMENTATION & TESTING OF SYSTEM	COMPUTER MODELS
Soils	$2,600.00	$11,300.00	$13,835.00	N/A	$6,000.00	$5,700.00		$1,485.00		
Geology	450.00	3,000.00	5,088.00	N/A	350.00	550.00		990.00		
Slope	500.00	2,150.00	3,727.00	N/A	350.00	650.00		990.00		
Mineral Resources	450.00	1,000.00	1,160.00	N/A	350.00	450.00		220.00		
Aquifers	400.00	1,500.00	1,000.00	N/A	350.00	400.00		220.00		
Surface Waters	200.00	1,100.00	1,150.00	N/A	300.00	300.00		990.00		
Natural Features	1,200.00	4,300.00	2,160.00	$ 1,750.00	8,000.00	2,350.00	500.00	495.00		
Vegetation	300.00	4,500.00	3,757.00	N/A	700.00	650.00				
Wildlife Habitat (Not Digitized)	350.00	2,000.00		N/A	700.00	650.00				
Sewer & Water Service	500.00	1,600.00	1,010.00	N/A	1,000.00	600.00		495.00		
Transportation	200.00	2,800.00	1,420.00	N/A	700.00	300.00		990.00		
Public Land	500.00	2,300.00	1,031.00	1,200.00	1,000.00	200.00	500.00	990.00		
Historic Sites	300.00	4,800.00	2,350.00	2,400.00	1,000.00	500.00	500.00	990.00		
Land Use (1970)	500.00	1,000.00	2,568.00	N/A	500.00	650.00		990.00		
Land Use (1973)	200.00	30,650.00	Planned	N/A	500.00	200.00		990.00		
Plans	300.00	2,700.00	Planned	N/A	500.00	500.00				
Boundaries (Electoral Districts & Watersheds)	300.00	2,250.00	3,750.00	N/A	N/A	500.00				
Non-variable Costs						2,250.00	6,440.00		3,200.00	2,500.00
	$ 9,250.00	$78,950.00	$44,006.00	$ 5,350.00	$22,300.00	$17,400.00	$7,940.00	$10,835.00	$3,200.00	$2,500.00

III. North Carolina
Land Resources Information System

General Description

The stated purpose of this developing system is to "provide local governments, state agencies and private citizens and businesses with the information capabilities necessary for making land use decisions." Thus because this system will be used at various levels of operation and by numerous users for a variety of tasks system flexibility was the primary desing objective.

A basic concept of the system is to include both manual and automated techniques for data storage, retrieval and manipulation. The manual aspects of the system involves, among other things a catalog of land related information. The automated aspect of the system will provide for the capability for collecting data in computer readible form, storing and manipulating it with the computer and retrieving the information in plotted, printed, or tabular form.

Existing computer facilities of the state Office of Management Systems, which consists of an IBM 370/158 will serve the bulk of the system. This facility is being used for large econometric modeling, processing of large quantities of data and when volumes demand, storage of data. Additional hardware was acquired to provide capabilities required for acquiring storing, manipulating, and outputting geographically referenced, spatial information.

Realizing the complexity of such a system and its extentive data base of land use and land related information the creation of the data base is proceeding incrementally. As land resource data is gathered by various agencies and groups for their particular problem solving tasks the data base development will continue. In this way the amount of data handled will accumulate gradually being entered into the system as it is captured. The collectors and users of the data will be the ones to enter it into the data base.

In order to construct a data base which contains accurate, useful and legally defensible data the data is collected and entered into the system in its most basic form. In this way, the accurracy and integrity of the data base can be maintained and flexible use is assurred. By storing spatial data as lines, points, and polygoms information will be provided to the widest range of users. The system will tie to the State Plane Coordinates as the overall system of reference. In this way, all data stored in the system is geo-coded in a similar manner for the entire state. A hirarchial grid structure is being utilized as a mechanism for ensuring compatibility and consistency for the collected data. One system will be using a compatible hierarchial structure of grid cell sizes, ranging from very small to very large.

LRIS History

Under the North Carolina Land Policy Act of 1974 a Land Policy Council was created and directed to carry out several mandates of which the primary one relates to the definition, preparation and maintenance of an information service for the land resources of the entire state. One Land Resources Information System is entented to provide a system into which the land related data collected by various agencies can be entered and stored in a consistent compatible way and be easily retrieved by various users.

LRIS Description

The following list presents important elements of the system which is currently in the design phase.

1) System data inputting.
   a. digitizing manually with interractive edit for capture of polygons.
   b. input data by coordinates
   c. input data in grid form
   d. all data converted to State Coordinates
2) Data Editing
   a. interreactive
   b. machine data compaction and reduction routines
3) Retrieval and Display
   a. print and plot source data
   b. retireval of digitized map

c. recall by map, attribute, combinations of attributes

d. recall by windowing

4) Hardware Capabilities

a. disc storage

b. drive digetizer and plotter

c. Fortran IV compilation

d. hardwired floating point

e. tape drivers

5) Software capabilities

a. mathematical and statistical calculations

- area, lengths, regression, cross tabulation, proportions

b. scale changes

c. overlay grid and polygon data

d. perform Boolian combinations

e. deal with island within polygons

f. input, output and convert data in

- UTM coordinates

- State Plane Coordinates

g. expandable file sizes

h. expandable analytical capabilities

Projected System Use

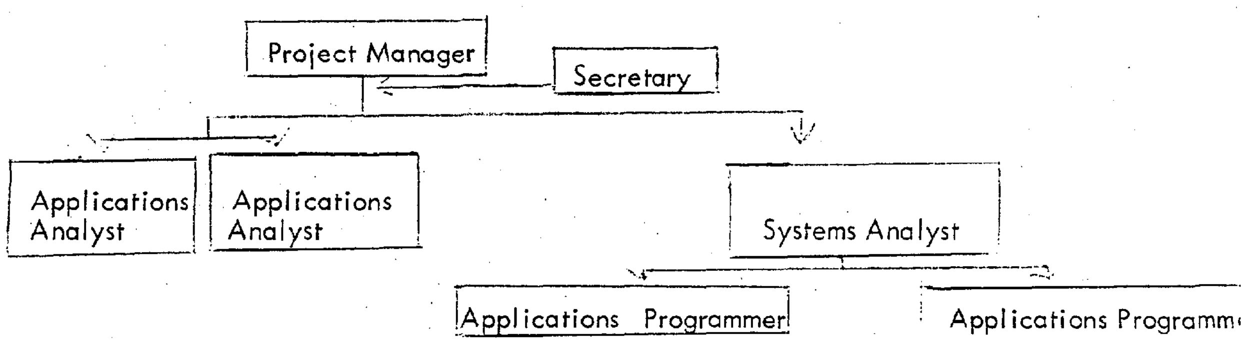
1. The State departments of Natural and Economic Resources, Environmental Management Division propose to use the system to digitize topography, climatalogic data and soils to plot this information for overlay with land cover and use activity data for non-point source regional wastewater management.
2. Interactive plotting and analysis of LUDA data which will be available on tape in a computer ready format.
3. Historic Preservation - Study of archaeologic sites, old roadway networks, and house

and barn types which will atlas and aid in design of predicture models relating historic site locations with known environmental and socio-economic conditions.

4. Resource inventory with interractive overlay capabilities for overlay with various socio-economic data types.
5. Economic Development - with data available, analysis of locational characteristics and suitability and capability analyses will be undertaken.
6. Forestry - the analyses and plotting capabilities of the interractive graphics system will be used to quantify the type and extent of forest operations being carried out and to correlate these factors with the location of forest operations and determined in earlier data gathering.
7. The operational interactive graphics system will permit the review of land , water and air dumping permits in terms of impacts on areas of environmental concern, as the Coastal Zone Management Act of 1972 calls for a review of development permitting based on utilization of all possible resource and socio-economic data types.

System Program and Costs

The system once implemented will operate with a support staff of seven persons. Figure A presents the organizational diagram for this operational support staff.



Represenatives Costs for system development are presented below. Direct purchase of 165 equipment alternative was selected to augment existing IBM 360 capabilities.

COSTS

Alternative 1: Direct Purchase of IGS Equipment

Calandar Year 1977

	Jan. - June 1977	July - Dec. 1977	Jan. - June 1978
Capital Costs:	$150,000	Fiscal Year 1978	
(Software & Hardware)			
Subtotal	$150,000		
Operational Costs:			
-Staff Salaries	$44,406	$44,406	$44,406
-Staff Training and Travel	6,025	6,025	
Office Miscellaneous:			
-Supplies	90	90	90
-Telephone	60	60	60
-Desks(6), Chairs(8)	1,200		
Maintenance	7,500	7,500	7,500
Disk Work Packs	3,500	3,500	
Plotter, Forms, Pens	150	150	150
Printer Forms	90	90	90
Physical Site Preparation	1,000		
Subtotal	$64,021	$61,821	$52,296
GRAND TOTAL	$214,021		

FUNDING

Alternative 1: Direct Purchase of IGS Equipment

Calendar Year 1977

	Jan. - June 1977	July - Dec. 1977	Jan. - June 1978
		Fiscal Year 1978	
Funds for Capital Costs:			
Dept. of Administration -Research and Development	$ 50,000		
Land Policy Council -Research and Development Land Classification	10,000		
DNER -208	50,000		
Dept. of Cultural Res.	5,000		
Subtotals	$115,000		
Funds for Operational Costs:			
Land Policy Council	13,000	25,000* 7,500	25,000* 7,500
DNER -LANDSAT Training	9,000 **		
Parks & Recreation	3,000	5,000*	5,000*
Subtotal	$ 25,000	$32,500	$32,500
GRAND TOTAL	$140,000		

*expansionary budget
**this funding is not assured at this time

IV The Montana Energy Resource Geographic Information System (ERGIS)

1. General Description

The general purpose of ERGIS development was the establishment of a geo-referenced data bank for planning land and resource utilization in Montana. The system, more specifically, was designed to increase work efficiency and to add the capability of handling complex environmental data for resource management with the Montana Department of Energy Resources. It was decided early in the system development that the focus would be on:

a. converting all resource information into digital form for data manipulations;

b. the conversion would be as automated as present technology would cost-effectively permit;

c. constant and instantaneous revision and updating of the data base would be required;

d. all types of errors would have to be greatly minimized; and

e. inventory maps should be produced at any scale, by an area and by any combination of data elements.

2. ERGIS History

The ERGIS system was developed in-house by systems analysists with the Energy Planning Division. Work was initiated in 1974 and parts of the system were operational by mid 1976. The system is currently utilized and supervised by the Energy Planning Division but through cooperative agreements, a number of federal and Montana State agencies have also used the system.

3. Specifics of the ERGIS

An overall review of the ERGIS system and the inter-relationship among its subsystems is show in Figure #1. The major subsystems are 1) input material analysis, 2) input device selection, 3) storage formats, 4) storage devices, 5) output formats and devices, and 6) manipulation specifications.

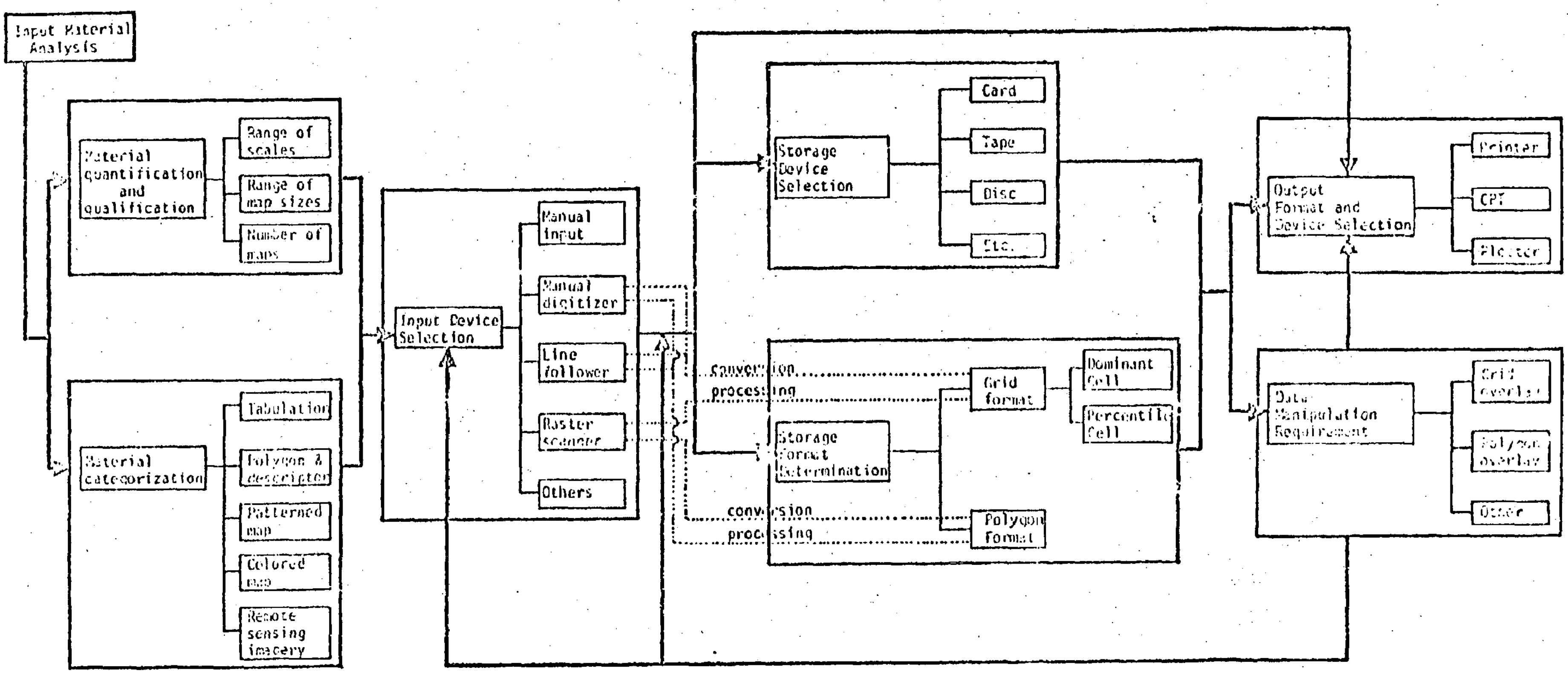

Input Material Analysis
Material quantification and qualification
Range of scales
Range of map sizes
Number of maps
Material categorization
Tabulation
Polygon & descriptor
Patterned map
Colored map
Remote sensing imagery
Input Device Selection
Manual input
Manual digitizer
Line follower
Raster scanner
Others
Storage Device Selection
Card
Tape
Disc
Etc.
Storage Format Determination
Grid format
Dominant Cell
Percentile Cell
Polygon Format
conversion
processing
conversion
processing
Output Format and Device Selection
Printer
CRT
Plotter
Data Manipulation Requirement
Grid overlay
Polygon overlay
Other

Figure I

Data Input Subsection:

For the system data bank, most input materials are in cartographic map formats. The system designers attempted a systematic quantification and qualification of all potential input materials in order to maximize eventual system efficiency. After an extensive study, similar to the methodology proposed in the main body of this report, input data and devices were selected. The digitizing equipment consists of a Broomall GP-100 vaster scanner. This piece of equipment can scan documents up to about $9\frac{1}{2}$ inches by 23 inches. The document is scanned into a cellular format at a resolution of 50, 100, 200 or 400 cells per inch. This scanner reads the document merely by shining a light on it and measuring the intensity of reflected light. Once scanned, the data is plotted and checked and then written to tape. This data can then be taken to the main computer for processing. (Fig. #2 presents the system hardware organization).

Based on an extensive review of current data storage and retrieval techniques the micro cell system was selected. ERGIS researchers felt this offered the greatest efficiency of random accessibility, greatest flexibility of arbitrary output formats and the greatest data compatibility.

Some of the data items which have been input into the system include Physiography, Vegetation, Sediment Risks, Tree Size, Forest Stocking, Specially Managed Areas, Existing Land Use Patterns, Range Vegetation Types, Range Conditions and Regional Comprehensive Plan Maps.

Analysis and Output:

The ERGIS system has a number of software packages for manipulation of digitized or scanned data. The packages are of two types: (1) software for cellular format processing, and (2) software for conversion to polygon format on which design has recently started. The software was written in FORTRAN with a minimum of machine-dependant instructions. Following is a brief explanation of the major programs.

(1) ACRE - calculates acres by counting cells.

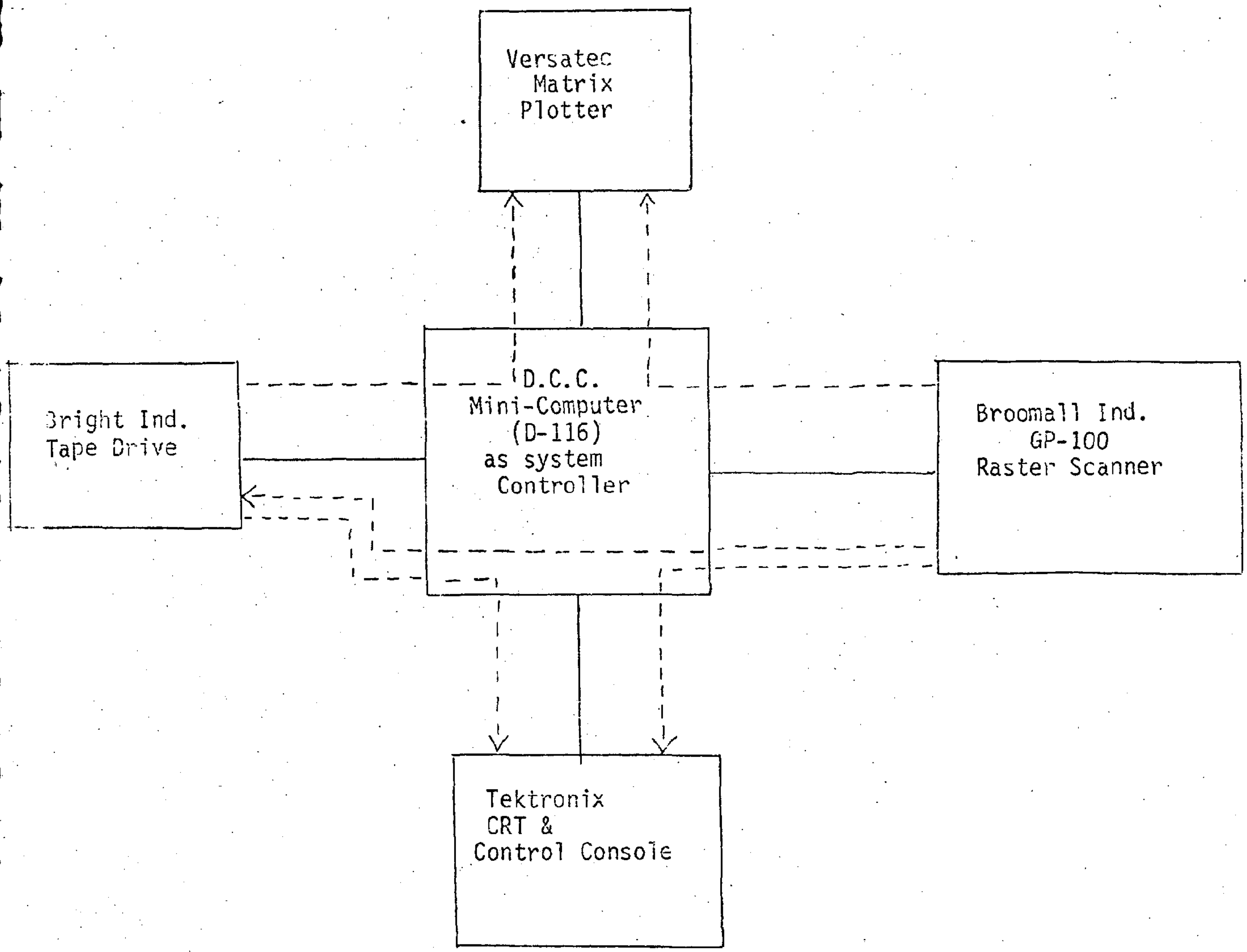

Versatec
Matrix
Plotter
Bright Ind.
Tape Drive
D.C.C.
Mini-Computer
(D-116)
as system
Controller
Broomall Ind.
GP-100
Raster Scanner
Tektronix
CRT &
Control Console

Figure 2.

(2) AGGREGATE - reduces the number of cells on a map by aggregating numbers of cells into larger single cells based on dominant types of the original cells.

(3) CORRELATE - can composite up to five maps in a single computer run. The descriptors inserted in a new map are determined by the descriptors of the old maps and the way the user has set the program to run.

(4) INPUT - can take data from other celluar mapping systems and since it uses 80 records it is very useful in accepting data punched on cards.

(5) JOIN - is used to join maps together. Since the EPD's raster scanner can scan maps at a maximum size of $9\frac{1}{2}$ inches by 23 inches, biger maps must be either photographically reduced or cut into smaller segments. Program JOIN can join scanned segments into one map.

(6) OVERLAY - is utilized in compositing if certain types of maps do not have descriptors or inventory patterns covering an entire study area. Thus, when one map is overlain on another, cells with a descriptor of O will not overwrite cells with non-zerodescriptors.

(7) PRINT - sends a map file to a line printer. It prints each cell in a map as a character on a line printer. Up to triple printing is utilized.

(8) RECOVER - will make line maps from printer maps.

(9) SREW - is used to extract a piece of a map or to extend the sides of a map.

Conclusions

ERGIS has been established for use in land and resource management which is data manipulation related. It has been their finding that for a regional or statewide data bank with large amounts of data that needs to be digitized, constantly maintained and updated with a minimum of errot, an automated digitizer was essential.

The important capabilities of ERGIS are:

1) file separation and file mergence in order to achieve random output ability, 2) vectorization of scanner data which can be printed or plotted, and 3) conversion of a polygon data base to a fine-cell grid data base until polygon manipulation is deemed feasible.

(No cost or time estimates were readily available for ERGIS).

Uses of the System

Some uses of the system have included·

1) The Montana Dept. of Community Affairs has been digitizing and manipulating data for a number of natural resource and socio-economic characteristics primarily for use by land use planners, expecially at the local level.

2) The Dept. of State Lands is using the system in mine siting studies.

3) The Dept. of Fish and Game is generating computer maps for acreage calculations according to species habitat and ownership.

4) The Energy Planning Division is using the mapping capabilities in routing transmission lines. As well, they are using the system for monitoring pollution from power plants and other industrial sources.

5) A number of Indian tribes are investigating using the system for managing tribal lands.

V. Texas Natural Resources Information System

Introduction to TNRIS

The goal of the Natural Resources Information System for Texas is to facilitate the fulfillment of specific statutory responsibilities and administrative needs of State agencies that must plan, developm operate, manage and conserve the natural resources of Texas. The system was designed to provide a flexible mechanism for maximum availability of natural resources data/information, consistent with cost and efficiency, to state, federal, regional, local and private entities for support of state resource programs such as water resource planning, coastal zone management, land use planning and energy conservation activities.

The TNRIS is not only a system in traditional engineering terms. It is defined as a service mechanism for (1) assembling data in both machine processable, and non-machine processable form; (2) processing raw data into meaningful data; (3) adjusting and organizing processed data into forms and formats suited for modern storage retrieval and manipulation procedures; (4) storing these data in a systemized information base; (5) disseminating data from this base of information; and (6) manipulating this data into maps, graphs, models and study plans, and simulation systems needed to manage natural resources as determined by user requirements.

The heart of the NRIS is an information base which was designed to meet the needs of the user agencies. Although comprehensive in nature and compatible with existing and planned federal and state systems, the data base is not completely centralized. Each agency has continued to be responsible for maintaining its own information files.

NRIS organization closely follows the design of the United States Geologic Survey National Water Data Exchange (NAWDEX) System. NRIS and NAWDEX are actually a combination of the linked network organization approach and a hierarchal approach, where all user agencies make up the linked network and a 'Systems Central' provides a coordinating point of contact and gives a required hierarchy. The basic advantage of this type of a system for Texas has been that user agencies continue to maintain internal data storage and retrieval systems for the data that they collect and use. Systems Central provides an interface between users so that data transfers, when required, are easy. The system also provides a central index available to all users, and is the processing site for external requests for use of the system.

TNRIS History

A TNRIS was established in 1972 but a number of important events occurred previous to this. Primary, was the creation of a "centralized data bank incorporating hydrologic data gathered in Texas" by the Texas Water Development Board in 1967. Then in 1971 an Inter-agency Council decided that this Texas Water Oriented Data Bank could serve as a foundation for a complete natural resources data system. 1971 marked the completion of the natural resources data identification and categorization project. At this time there was also developed a NRIS conceptual design with recommendations for implementation. Subsequently, the system actually became active with the compilation of NRIS file descriptions and a survey of NRIS data/information needs. Work on geographic data manipulation and analysis capabilities was initiated in 1974 and was operational by 1975.

Specifics of the Texas NRIS.

Basic to this system as to any large information system are three elements: (1) the users of the system, (2) the data base, and (3) the information services which are provided.

As the TNRIS was designed to serve a wide range of users (decision maker, planner, technician, research, and the general public) the data base reflects this diversity. Though totally comprehensive and systematically organized along defined lines, the data base is generally not centralized. Each agency maintains data bases which exist in both machine processable form. The following categories and subcategories form the present TNRIS data base:

I. Base Data (i.e., base maps, geodetic control).

II. Meteorology.

  A. Climate

  B. Air Quality

  C. Man's Activities (i.e., permits, etc.)

III. Biologic Resources.

  A. Animals

  B. Plants

  C. Microrganisms

  D. Man's Activities

IV. Water Resources

  A. Surface

  B. Subsurface

  C. Man's Activities

V. Geologic and Land Resources

  A. Surface

  B. Subsurface

  C. Man's Activities

VI. Socio-economic Resources

  A. Social

B. Economic

C. Commerce

D. Government

E. Archeologic

The information base has been implemented using standard codes and procedures designed to ease use of the system and for system compatibility.

TNRIS services include providing information on data abailability, providing basic data and information retrieval, and providing a number of data analysis capabilities. Specifically the range of TNRIS services include:

(1) Data Inventories

(2) Data Reports

(3) Technical Studies

(4) Environmental Impact Assessment Procedures

(5) Preparation of planning and resource management models

(6) Project management

(7) Assistance in project design, operation and control.

The system, for geographically specific data such as soils and land use, uses a polygon mapping system. The system enters data based on the digitization of line segments for polygons, lines for linear data, and points for point data. The system is based on techniques proposed for the USGS Digital Cartographic Data Base. The system has the capability to transform a file between any two map projections defined to the syste, as well as, to geodetic coordinates. Presently available projections are Mercator, UTM, State Planar, Lamberts Conformal Conic, Albers Equal Area and Orthographic.

Within the system data processing involves the chaining of three data files. The largest is the chain file which lists X, Y coordinates and contains feature item coders for the left and right sides of all line segments. The next files, the definition file, contains the fourt major data types: points, lines, area, and networks. The third file is an easily updateable index for all definitions and boundaries.

Standard color cartographic plotting and interactive analysis and display are system output modes. Analyses include overlay of up to eight data sets, cross tabulation, area and distance calculations and a number of specific procedures such as variety analysis. Work is currently underway for incorporating LANDSAT Digital imagery files and pattern recognition/ classification software into the system.

TNRIS Summary

It appears that the system has been able to accomplish three of its basic objectives:

(1) makes data available to users, consistent with cost and efficiency, rather than just to source agencies;

(2) it has minimized redundant data collection programs; and

(3) provides data analysis capabilities which would be beyond the resources of a number of user agencies.

System implementation in Texas did require a close working relationship between the people who would be operating the system and system users. The system central is currently staffed by six persons who are a mixture of computer analysts and natural resource planners and managers. (Exhibits 1 through 4 illustrated numbers aspects of system operation and design).

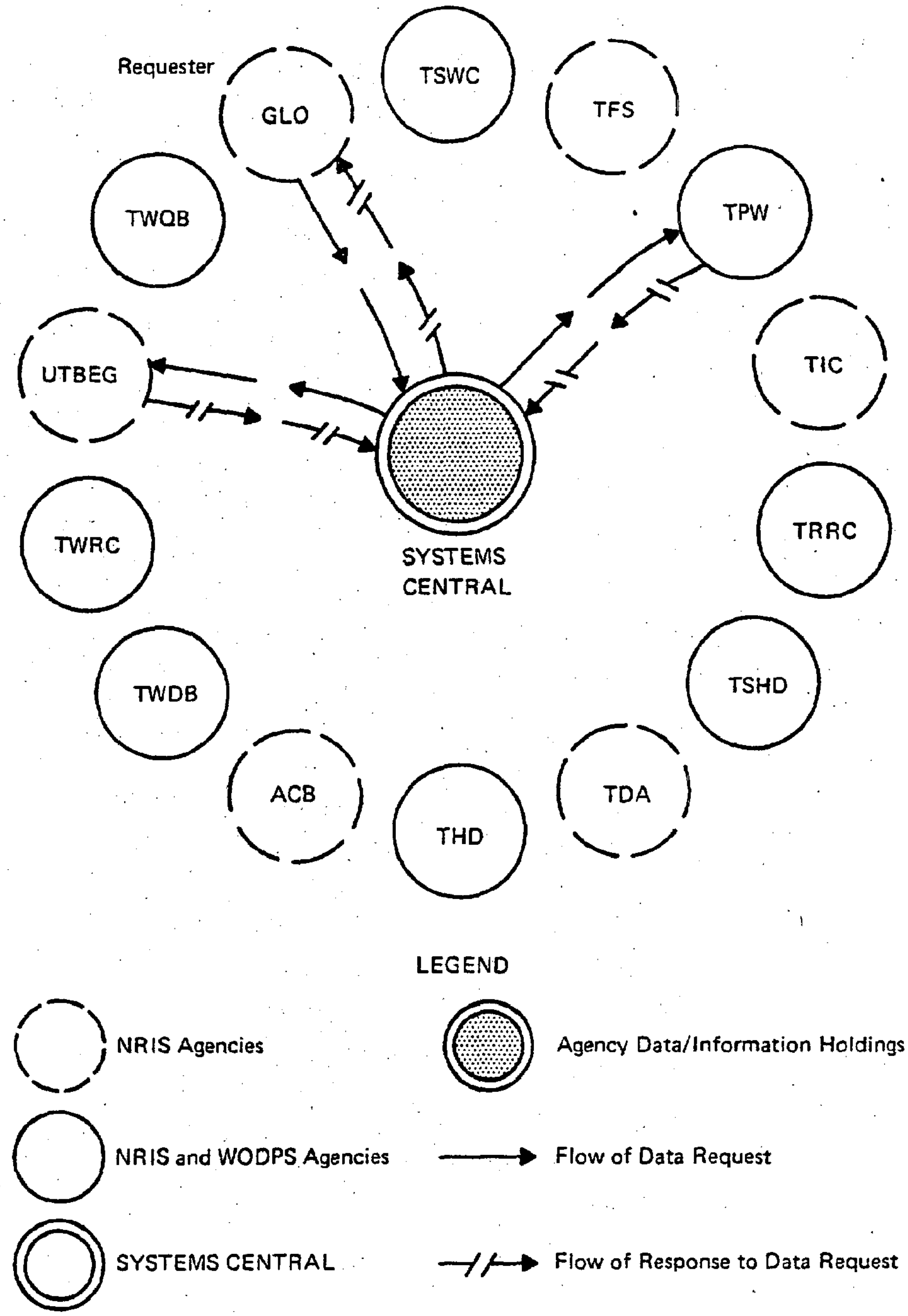


Exhibit 1
NRIS Data and Information Handling Situations
Request Situation 1

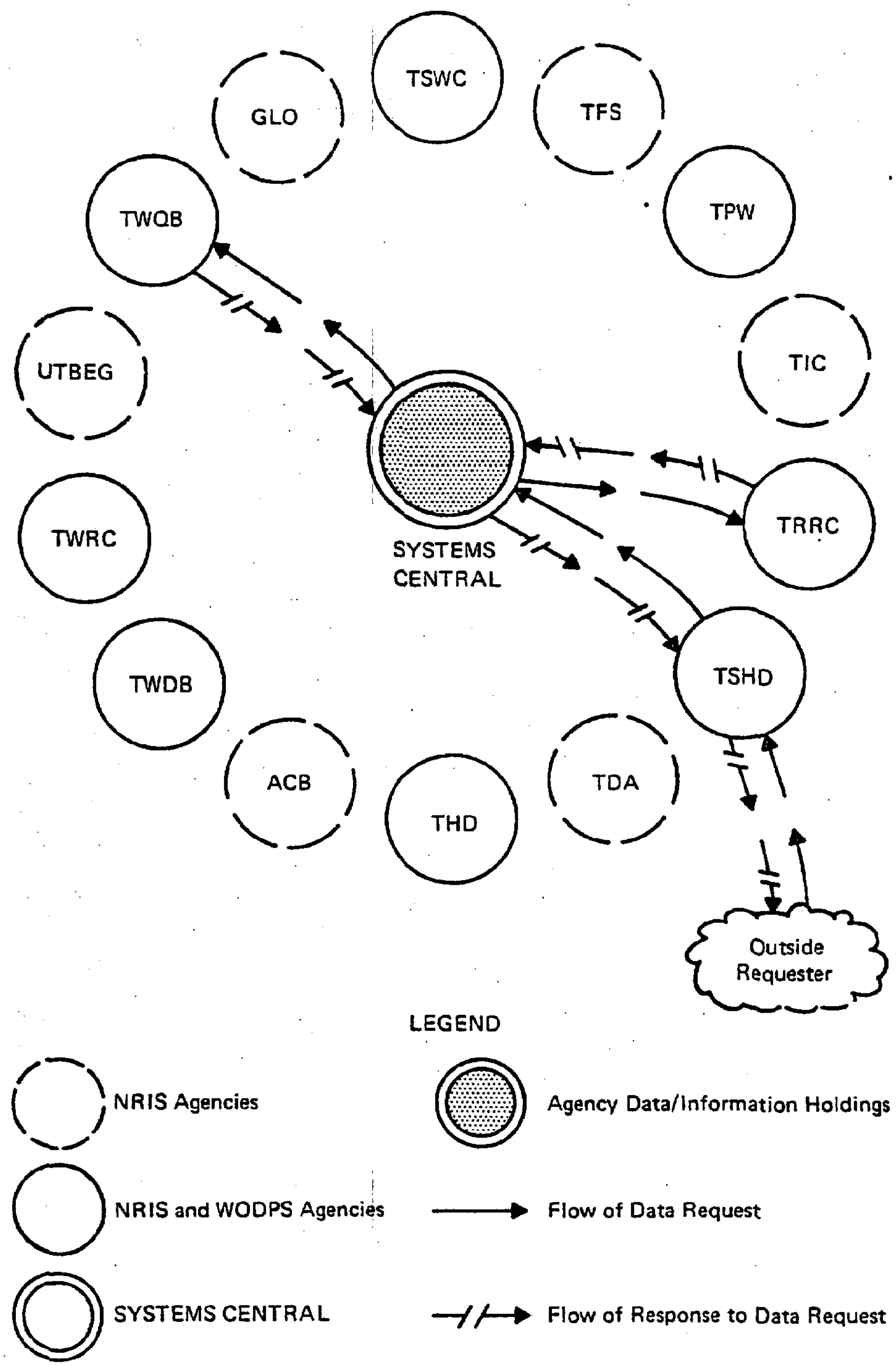


Exhibit 2
NRIS Data and Information Handling Situations
Request Situation 2

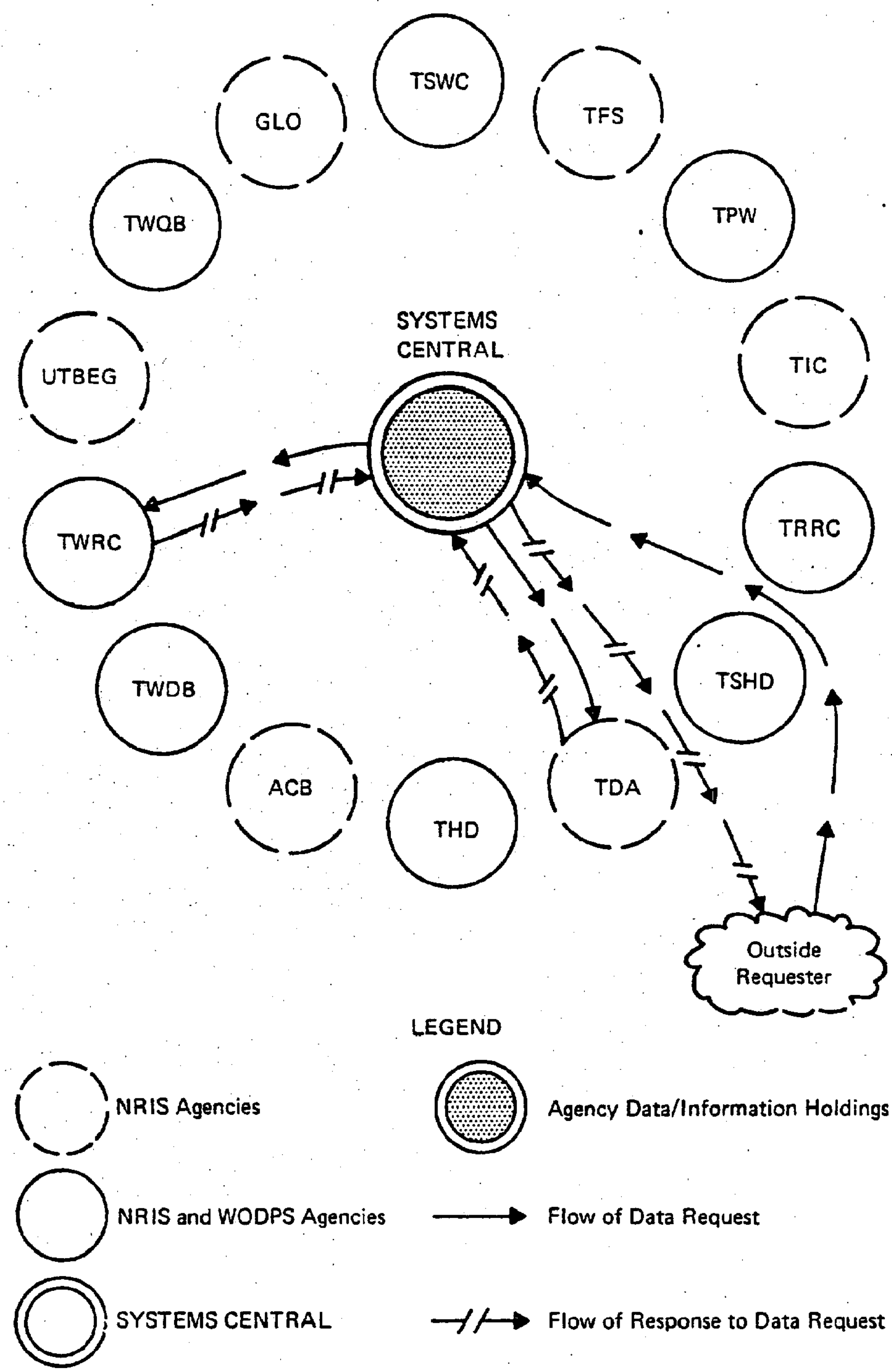


Exhibit 3

NRIS Data and Information Handling Situations

Request Situation 3

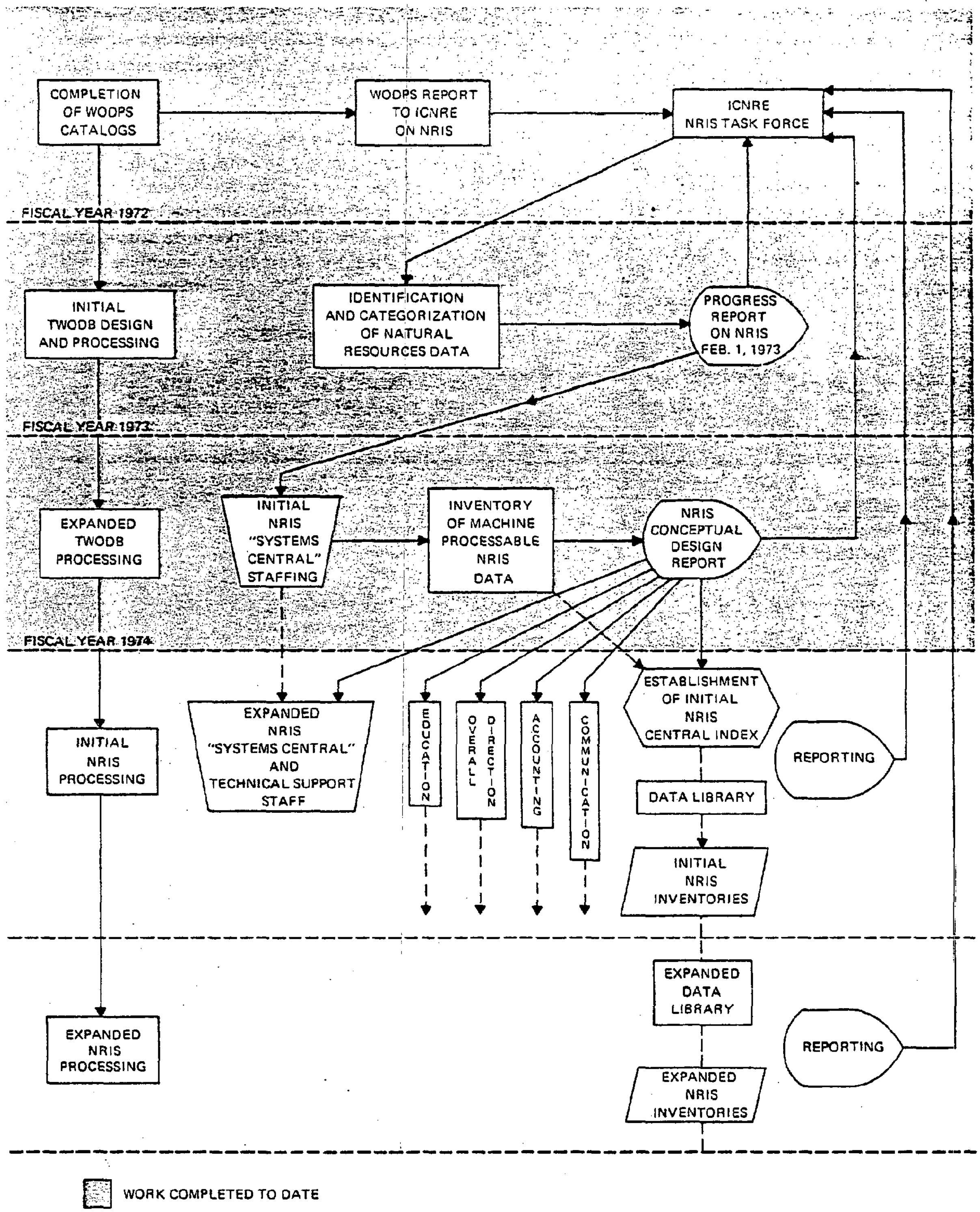


Exhibit 4
Schedule of NRIS Implementation

VI Statewide Geographic Information Systems - -

Summary and Evaluation

A comparison of a number of existing and developing systems permits a number of observations concerning the State-of-the-art of such systems.

Data

A. Digitization is technically the major hinge in the information flows of all systems served. The process of creating an error-free file is critical because current hardware and software cannot intelligently ignore or correct non-logical errors. The process found in all systems is essentially:

a. pre-edit of graphic data

b. digitizing

c. correction of digitizing errors

d. structuring the data file

B. The development of interactive data inputting capabilities is of considerable aid in correcting digitizing errors. Also such capabilities permit browsing a data file (i.e., Montana, ERGIS).

C. It has proven expensive to interface line and point data with cell or polygon data once a system has been designed, so file structure should be designed with the possibility of this interface in mind (MAGI, ERGIS).

Informational Outputs

A. User expectations of automated systems have been generally too high if users know nothing or little about specifics of the system being utilized or developed (LUNR, MAGI).

B. Standard statistical packages such as SPSS and BIOMED have not proved adequate to meet the needs of spatial data analysis. There are statistical methods to handle spatial data, but in general, they have not been adequately developed within the context of existing systems. Such statistical procedures recognize the potential spatial dependance of one observation on another (LUNR).

C. It is critical that all users understand the manipulative techniques being utilized so system outputs can be easily used in user decision making. This is often not the case and many systems systems manipulative capabilities are not used or are used incorrectly. Many state systems have utilized go-between processes as a part of user education programs. In this process, system developers and sponsors act as intermediaries between various users and the system, translating both general and specific user problems into terms to which the system can respond. Currently there are no explicit systematic mechanisms by which to determine the amount and types of increases in user performance by utilizing information that comes from the system. Often if there is no perceivable change in decision making it can be fatal to the continuity and survival of the information system. The systems reviewed have exhibited that there is a clear need for either the sponsors or designers of the system to take the initiative and actively advocate use of the system. This can only be accomplished if the systems staff includes personnel capbable of understanding user problems (all systems).

System Management Institutional Framework

A. With all systems there have been significant time delays in project approval, the purchase of hardware and software, and hiring and training staff to operate the systems. Generally, past system programs have not been structured to reduce the adverse effects of such delays (LUNR, MAGI).

B. Systems initially must win the support of high level decision makers. The main problem appears to be overcoming negative attitudes and misconceptions which exist concerning the use and utility of computers (NORTH CAROLINA).

To accomplish this, the following groups must be included in system design.

a. the data gatherers;

b. the systems designers; and

c. the anticipated user group.

These groups must collectively identify the data supplies, evaluate data supplies with respect to the needs of the user groups, and solve the problems of data flow from the supplies to the system.

C. Documentation of every aspect of system development and operation or significant problems arise with respect to:

1) inevitable staff changes;

2) the transfer of the system or of particular aspects of the system from outside contractors to system sponsors.

Also, staffing must recognize the following activities and stages in system development:

1) system conceptualization;

2) system debugging; and

3) operation, maintenance and updating.

D. It is not proved satisfactory to rely heavily on universities or outside organizations such as consulting companies as places to build information systems even though the growing organization temporarily avoids staffing problems.

E. There have been significant problems in maintaining fiscal continuity for the development of information systems, given that such development generally takes anywhere from 3 to 10 years to complete.

F. System publicity has proven a problem area because if development takes a number of years,it is necessary to report progress on a regular basis. Successful mechanisms for accomplishing this have been series of deliverable products integrated with user education programs. The user education process must start at the current level of capability of the user rather than at the systems projected level. Education is not how to use the system but rather how the users problems can be handled in quantitative terms using data. It probably should be assumed that the user must be shown approaches to problem solving which are entirely new. It is impractical to assume user feedback unless it is part of user education for a user cannot provide valid feedback if his capabilities are not equal to the systems manipulative powers.

Other systems reviewed but not detailed here include:

The Minnesota Land Management Information System

The Canadian Geographic Information System

The Alabama Resource Information System

The Hawaii Information System

The Composite Mapping System of the Federation of Rocky Mountain States

A PROPOSED NATURAL RESOURCE INFORMATION PLANNING PROGRAM FOR MAINE

The basic purpose of this program is to provide an organized framework for coordinating Natural Resource Information planning activities of State agencies and to coordinate State activities with those at other levels of government. Implementation of this program should provide for efficient use of available funding and substantially increase the usefulness of existing natural resource information.

The Program would consist of six functional areas:

I. Indexing
II. Collection Planning
III. Collection
IV. Storage and Retrieval
V. Interpretation and Analysis
VI. Distribution

I. INDEXING

Essential to a Natural Resource Information Planning Program is the ability to quickly and easily determine what natural resource information exists that is relavent to a given planning or management activity.

To provide such a determination, a central automated index of all existing natural resource information pertaining to Maine needs to be established and regularly updated. Once such an index was established, a number of special purpose tools to meet specific planning and management needs could be developed. These might include:

a. Printouts containing all index entries for a specific geographic area (for example Regional Planning Commissions or Minor Civil Divisions)

b. A document listing all available digital information files, or

c. A document listing all spatially distributed resource information (similar to the Wisconsin Inventory of Land Resources Data). Such a document is currently being prepared in preliminary form by the State Planning Office.

The index would be available for special purpose searches as well. For example when an agency or contractor needed to prepare an Environmental Impact Assessment for a project, a special search could be run that would identify all existing, relevant natural resource information.

Substantial progress towards establishing such an index has been made by TRIGOM (The Research Institute for the Gulf of Maine). Recently TRIGOM was funded by the State Planning Office to conduct a study of how the transfer of natural resources information from sources to users could be expedited.* This study recommended that:

1. A natural resource information center be established within the Maine State Library.
2. The center prepare and administer an index of Maine natural resource information. This index would consist of:
   a. Affiliation with the National Cartographic Information Center (NCIC) and use of the NCIC remote-sensing imagery index; and
   b. The Maine Index (as proposed by TRIGOM) for non-imagery information.

II. COLLECTION PLANNING

A natural resource data collection planning process would be established and an initial Natural Resource Data Collection Plan prepared for Land and Water Resources Council approval. The planning process would provide for periodic updating of the Data Collection Plan and certain of its component parts.

Activities/Reports to be accomplished as part of the Natural Resource Data Collection Plan would include:

A. Inventory of Resource Information

The existing natural resource data base would be inventoried and major items would be published in loose-bound form as a handbook. The handbook would be updated annually by the State Planning Office and would be modeled upon the "Inventory of Wisconsin Land Resources Data". The handbook could be derived from information contained in the MAINE Index and updated through the Index.

---

* Natural Resource Information Transfer Study, TRIGOM, January, 1978. Available upon request from the State Planning Office.

B. Data Needs Assessment

Existing and potential natural resource data users at state, regional, and local levels would be interviewed as part of a process to assess user needs. Survey results would be reviewed and a final data needs report prepared.

The user needs assessment would be updated on a periodic basis (every 3 - 5 years).

C. When possible, planning reports would be prepared on major categories of natural resource data being collected or under consideration for collection. These reports would investigate the relationship between the resource type and planning needs, and should present recommendations concerning whether or not data should be collected for planning purposes in that category, how the data should be collected (for example, at what scales and level of detail, and what kinds of units should be mapped, etc.), and geographic priorities for collection. Ideally, these reports would be prepared prior to initiating new collection efforts.

Categories of data to be considered for report preparation include:

1. Bedrock Geology (completed in draft form)
2. Surficial Geology
3. Soils
4. Vegetation
5. Wildlife Resources
6. Estuarine Resources
7. Marine Resources
8. Water Data (report in progress)

D. Technical Reports

A number of special-subject technical planning reports should be conducted as a prelude to preparation of the Data Collection Plan. Subjects of these reports would include:

1. Standardized Base Mapping
2. Projected Data Outputs from Federal Programs through 1985
3. A Review of Selected State Resource Information Planning Activities (recently completed by the Council of State Governments).

E. <u>Plan Formulation</u>

All reports would be reviewed by the Advisory Group prior to their completion. Report findings and recommendations would be reviewed and used in preparing a first draft of the Data Collection Plan. The Draft Plan would be expected to state the following:

1. Unmet Data Needs
2. Unmet Data Needs requiring State action for timely collection (including standards for collection of such data)
3. Priority for state collection efforts
4. Implementation Program (5 Year)
    a. Collection
    b. Coordination
5. Research Needs

Once priorities for State collection efforts had been set, special studies would be conducted to develop and review detailed specifications for high-priority data categories. These specifications would include: coding, mapping unit composition and size, accuracy, methodology for collection and presentation, etc.

The initial Draft Collection Plan would receive wide review. The Advisory Committee would forward a final draft to the Land and Water Resources Council with recommendations for implementation and other action as appropriate. Questions that cannot be resolved within the Advisory Committee would be forwarded to the Council for resolution. The Data Collection Plan would be partially updated on an annual basis, and fully reviewed and updated every five years.

III. COLLECTION

The Data Collection Plan would be implemented through collection activities at all levels of government. Collection activities by each level would be coordinated in the following ways:

A. Federal Collection Activity

Federal collection activity would be monitored. The State would submit a statement of its priorities to the various federal collection agencies to provide maximum consistency with the Data Collection Plan. Formal liaison would be established between the Land and Water Resources Council and the federal collection programs via the Advisory Committee and the data planning coordinator at the State Planning Office.

B. State Collection Activity

State agency collection activities would be coordinated through the collection planning process to provide consistency with the Collection Plan and to avoid duplication of effort. A data collection clearinghouse would be established in the State Planning Office. All state and regional agency proposed data collection activities would be submitted to the clearinghouse and reviewed by the Council's Advisory Group. Comment would be made concerning consistency with the Collection Plan. Agencies would not be required to be consistent with the Collection Plan, but the Advisory Committee could recommend that the Council review proposals where it appeared that consistency could be achieved without extra cost and the agency involved had not adequately made a case for lack of consistency.

C. Regional Collection Activity

The Land and Water Resources Council would circulate the Collection Plan and would request that regional agency collection activities be consistent with the plan where possible. Where regional collection activities were funded from State or Federal programs, collection activities would be subject to the same formal clearinghouse procedures as state agencies.

D. Local Collection Activities

Where appropriate, Classification Codes, Standard Base Maps, etc. would be provided to communities engaged in local data collection. They would be encouraged to be consistent with provisions of the Collection Plan.

IV. INTERPRETATION AND ANALYSIS

A. Interpretation

Reports would be published in handbook form to assist planners at all levels in using resource data. These handbooks would provide an indepth understanding of the data collected, how it was collected, andhow it can be used for resource planning and management.

Subjects for handbooks might include:

1. Surficial Geology **(in progress)**
2. Wildlife Resources
3. Groundwater **(in progress)**
4. Surface Water
5. Soils
6. Vegetation
7. Intertidal areas **(in progress)**
8. Marine Resources
9. Estuarine Resources

B. Analysis

Included under this activity would be the Council's program for identification of areas suitable for development and general resource analysis activities of state agencies. It would not normally include specialized analysis, conducted by a single agency to meet inhouse needs that is not broadly useful to other agencies (for example, specialized wildlife population analysis conducted by the Department of Inland Fisheries and Wildlife).

1. Identification of areas suitable for development

Resource characteristics, including those of geology, soils, and slopes would be analyzed and rated to guide selection of favorable sites. Where possible, standard analysis methodologies would be developed and adopted by the Council so that individual agencies conducted such analysis for specific program needs could contribute to the larger, statewide program with their efforts.

2. Other Analysis Activities

As the State's resource data base is developed, substantially increased analysis activity can be anticipated. Where possible, methodologies for such analysis would be standardized, or at a minimum, reviewed and commented upon by the Advisory Group. Such activities might include special siting studies, resource analysis to support policy studies, etc.

V. STORAGE AND RETRIEVAL

As long-term natural resource information needs begin to be classified through the process of developing a Natural Resource Data Collection Plan, an organized system of storage and retrieval of such data will need to be developed.

Substantial progress in this category will likely have to wait until work in earlier categories is well underway.

A number of special reports would be conducted to support development of an organized system for storage and retrieval. These would include:

1. A study of the feasibility of establishing a Geographic Information System at the State level for the storage, retrieval and analysis of specifically distributed natural resource data. (complete in draft form)

2. A study to prepare recommendations concerning how to improve the distribution of existing natural resource information to existing and potential users. (completed)

VI. DISTRIBUTION

The final function of a complete information system is distribution of information to users. This function should include both the simple distribution, as well as outreach activities to improve awareness of data, to point out how available data can be used to solve problems, etc.

A study has been conducted under the Coastal Program that is related to this function. This was the Natural Resource Information Transfer, which was completed by TRIGOM in January.

REFERENCES CITED

Adams, V.W. Earth Science Data in Urban and Regional Information Systems - A Review. U. S. Geological Survey, U.S.D.I. Circular 712. 1975

Baxter, F.P. and T.L. Cox. A Computer Assisted Land Management Decision System. Tennessee Valley Authority. Dec. 1975

Calkins, Hugh. "An Information System and Accountability Theory in Planning. P.H.D. Dessertation, University of Washington, 1972.

Crowder, R. "L.U.N.R.: Land Use and Natural Resource Inventory of New York: What it is and How it is to be used." Albany, Office of State Planning, 1972.

Dueker, K.J. and R. Talcott. Statewide Land Use Analysis and Information Requirements. Working Paper 13, Institute of Urban and Regional Research, University of Iowa, 1974

Durfee, R.C. ORRMIS: Oak Ridge Regional Modeling Information System. Oak Ridge, Tennessee. ORNL. 1974

International Geophysical Union. Second Interim Report on Digital Spacial Data Handling in the U.S. Geological Survey. IGU Commission on Geographical Data Sensing and Processing. Ottowa, Canada. 1976

Tomlinson, R.F. Geographic Data Handling: The Proceedings of the UNESCO/IGU Second Symposium of Geographic Information Systems. Ottowa, Canada: IGU Committee on Geographic Data Sensing and Processing. 1972

U.S. Department of the Interier. Information/Data Handling: A Guidebook for Development of State Programs. U.S.D.I. R.A.L.I. Program, Washington, D.C. 1975.

Wright, Linda. Maine Land Use Planners: A Survey of Resource Information Needs (draft). Orono, University of Maine, 1977.

Murray, Timothy. Honey Hill: A Systems Analysis for Planning the Multiple Use of Controlled Water Areas. Alexandria, Virginia: U.S. Army Engineer Institute For Water Resources. 1971.

Pencker, T.K., P.M. Mark, and J.L.P. falty Geographic Data Structures Cambridge. Laboratory for Computer Graphics and Spatial Analysis. Howard. 1973.

Schweitzer, R.H. Mapping Urban America with Automated Cartography. Washington, D.C. Geography Division, Bureau of the Census. 1973.

Steinitz Rogers Associates, Inc. The Santa Ana River Basin: An Example of the use of Computer Graphics in Regional Plan Evaluation. Campbridge, Massachusetts. 1975.

Talcott, Richard and Kenneth Dueker. Government Policy and Decision Making Implications for Land Resource Information System Design. The Institute of Urban and Regional Research. University of Iowa. 1974.

Texas Natural Resource Information System File Description, Jan. 1974.