Biological Services Program

FWS/OBS-77/22
JUNE 1977

A GENERAL DESIGN SCHEMA FOR AN OPERATIONAL GEOGRAPHIC INFORMATION SYSTEM

Interagency Energy-Environment Research and Development Program
Office of Research and Development • U. S. Environmental Protection Agency

Fish and Wildlife Service

U.S. Department of the Interior

FWS/OBS-77/22
June 1977

A GENERAL DESIGN SCHEMA FOR AN OPERATIONAL GEOGRAPHIC INFORMATION SYSTEM

By

Larry Salmen, George Nez, James Gropper
John Hamill and Carl Reed

Information Systems Technical Laboratory
Federation of Rocky Mountain States, Inc.
One Drake Park, Suite 142
333 Drake Road
Fort Collins, Colorado 80521

U.S. Fish and Wildlife Service
Contract No. 14-16-0008-2155

Dr. Kenneth P. Burnham, Project Officer
Western Energy and Land Use Team
Drake Creekside Building
2625 Redwing Road
Fort Collins, Colorado 80521

Interagency Agreement Number
EPA-IAG-D5-E685

Mr. Clinton Hall, Director
Energy Coordination Staff
Office of Energy, Minerals
and Industry
Environmental Protection
Agency
Washington, D.C. 20460

This study was conducted
as part of the Federal
Interagency Energy/Environmental
Research and Development Program
U.S. Environmental Protection Agency

Performed for

Western Energy and Land Use Team
Office of Biological Services
Fish and Wildlife Service
U.S. DEPARTMENT OF THE INTERIOR

DISCLAIMER

The opinions, findings, conclusions, or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the Office of Biological Services, Fish and Wildlife Service, U.S. Department of the Interior, nor does mention of trade names or commercial products constitute endorsement or recommendation for use by the federal government.

This report is meant for use within the Fish and Wildlife Service only. Please do not circulate outside the Fish and Wildlife Service.

PLEASE DISREGARD

Library of Congress Catalog Card Number: 77-82532

Report 1.2

Abstract

This report is the second in a four part report titled User Needs Assessment. The other three volumes discuss the actual user needs in terms of both analysis and required data. They also discuss possible areas in Wyoming and Montana that could be used as test areas. (Appendix B gives the abstracts from 3 other reports.)

Report 1.2 of the Task I Final Report User Needs Assessment presents a general system design and outlines a basic set of system capabilities for the U.S. Fish and Wildlife Service's Geographic Information System. The system is to operate at two levels. The first level is the manual system. The manual system is for cataloging maps, relevant articles, and other paper documents. The second level is the automated level. At the automated level, a computer based information system will allow for data input, data analysis, and information output based on user specified criteria. The discussion of the system is non-technical in format and is geared toward those individuals with little or no computer-information system experience.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

The goal of this project is to develop a coordinated manual and automated Geographic Information System (GIS) for use by the Area and Regional Office of the U.S. Fish and Wildlife Service, Region Six. The function of this report is to outline the general conceptual design of a GIS for FWS. A broad set of system capabilities are presented. This represents the first step in the design process. Later design steps will firm up and outline in detail all the functional characteristics of the system. To begin with, the system must have both a manual component and an automated component.

The proposed system configuration will allow many users who have simple spatial problems to use a manual information system, while more complex spatial problems would utilize the more sophisticated power of the automated information system. The system design is necessarily based on the findings of a User Needs Assessment (Report 1.1).[1] The requirements of the users serve to define the operational boundaries of the information system. It is necessary only to fulfill the needs of FWS - not the entire world. The user needs also serve to define what are the relevant internal input analysis and display capabilities. These capabilities in turn are dependent on the data required by the users. Summarizing from Report 1.1, the FWS users group requires data of the following general characteristics:

---

[1]User Needs Assessment for an Operational Geographic Information System Within the U.S. Fish & Wildlife Service

(1) different data resolutions or geographical accuracy ranging from a 2 1/2 acre resolution (1:10,000 scale map) necessary for site specific planning and facility location to 640 acre resolution (1:500,000 scale map) suitable for region wide planning of an entire river basin or state.

(2) different data currency or updating frequency ranging from the requirement for 3-6 month currency for corporate mineral development plans to 10 year currency for soils and geologic map data.

(3) different geographic coverage or extensiveness of the study area ranging, for example, from a 200 square mile study area around a small water development project to an 85,000 square mile study of the entire state of Wyoming.

(4) differing data formats ranging from summary tabular reports to detailed 1:24,000 scale maps.

The general system design presented in this report is based on data required (both existing and hoped for) by FWS personnel to fulfill the mandated objectives of the Fish & Wildlife Service.

The discussion of the system is presented in three main sections. The first section (2.0) gives an analogy of a GIS in a very broad, non-technical format. The second section (3.0) discusses the manual portions of the GIS. For those who are interested, the third section presents automated system description in finer detail both technically and conceptually (4.0).

2.0 AN INTRODUCTION AND NON-TECHNICAL OVERVIEW OF GEOGRAPHIC INFORMATION SYSTEMS

Whether a system is completely manual, completely automated, or a combination of both, it will maintain five primary functions: 1) acquisition of data and its insertion into the data base, 2) maintenance of and retrieval from the data base, 3) analysis of data retrieved from the data base, 4) monitoring of the system, and 5) interaction between the user and the data base.

Operating an information system is in many ways like operating a restaurant. If you think of data as food, the five functions are quite readily understood. To begin, food must be acquired (data acquisition). Usually it is easiest to purchase food from commercial sources (secondary data acquisition) provided that the needed items are available, the quality is acceptable, and the supply is reliable. If the food is not available, then the more expensive, time consuming process of growing food must be considered (primary data collection). Generally, it is best to use readily available items, or at least have any necessary growing done under contract.

Once acquired, the food has to be stored (data input). Before the food is stored, it must be checked for quality and freshness (input data editing). Items that are spoiled or imperfect must be discarded and replaced.

Once the food is stored, the chef must be able to retrieve the food items as required by the demands (orders) of the customers (data users). At the same time, the chef keeps a close check that the food is not going bad, and replenishes his supply to meet the demands of his customers (data update).

After the food is retrieved from storage, the chef must prepare and cook the items to meet the specific demands of his customer (data analysis and data output). In a restaurant, the degree of preparation and presentation may be minimal. A smorgasbord, for instance, may require only retrieval, cleaning, and layout of a selection of food. The degree of preparation, such as in a French cuisine restaurant, may be much more complex. In such instances, the patron may wish to prepare his own meal. The point is that some customers wish only to get a quick peanut butter and jelly sandwich, while others may want a four course meal with fine wines, (served by a wine steward), after dinner liquers, and the service required of nobility.

Critical to the existence and profitability of a high calibre restaurant is the existence of demanding clients (users of the information system). In the first stages of establishing a restaurant, it is unwise to provide dishes that are too complex and too expensive for the client's taste. The restaurateur must decide whether to cater only to existing demands or whether gradually to introduce new dishes with which the client can become familiar and eventually request. The restaurant management can do this by monitoring reactions of his clients to the dishes and listening to their suggestions and complaints.

One point where the client and restaurateur interact is the menu (user interface). The client makes his selection from the menu. If the menu is lacking, the client will say so. He may possibly get up and leave the restaurant, never to return.

The analogy between a restaurant and an information system should not be drawn too far, for the processes that make up an information system are complex, segmented and closely interrelated. First, the manual and computer aided interface of the system will now be discussed in greater detail. Then, a more detailed conceptual and technical description of the FWS GIS is presented.

2.1 The AUTOMATED-MANUAL System Interface

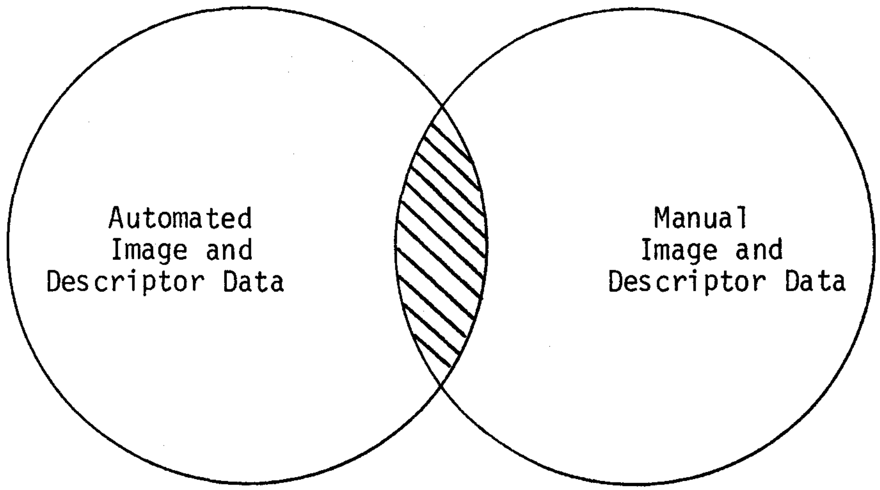
A computer based geographic information system must necessarily consist of two operational components: an automated portion and a manual portion. The automated portion consists of those system elements that are 1) amenable to computerization, 2) frequently used, and/or 3) require complex calculations for use by the user. The manual portion of the system consists of those system elements (such as map files, index card files, and field notes) which can be efficiently handled using manual techniques. The manual approach may be more desirable to implement than the automated system. The manual approach may be desired due to infrequent use of data, to the effectiveness of the present format, to the difficulty and cost of automation, or to the fact that the present manual system effectively serves the needs of its users.

Figure 2.1 presents a graphic representation of the automated-manual configuration of the system. Both the automated and manual portions of the system consist of two basic types of data: image and descriptor. Image data, also known as graphic or map data, range from uninterpreted aerial photographs to printed base maps. Image data may be tied to some geographic reference system, such as Latitude/Longitude, State Plane Coordinates, or Universal Transverse Mercator. Descriptor data, usually in the form of tables, text, or labels, may be maintained in the form of field notes, completed questionnaires, or card files.

Descriptor data may be tied to image data through a geographic reference. For example, as shown in Figure 2.1 soil interpretation reports can be tied geographically to soil survey maps.

Image or descriptor may be either in manual form or automated (computer) form. Image data may be entirely devoid of descriptive information, such as a map of soils, polygons or uninterpreted aerial photographs. On the other hand, descriptor data may be entirely devoid of any geographic referencing scheme, such as soil interpretation guides or bibliographies. However, the majority of

FIGURE 2.1 THE SYSTEM INTERFACE PROBLEM



The "Interface"
(automated index to manual files, for example)

the data of interest for the use in the FWS GIS contains both descriptor and image data in some logical relationship.

Figure 2.1 shows two important relationships:

(1) the automated-manual interface relationship, and

(2) the degree of overlap between the automated and manual interfaces.

The interface between the automated and manual portions of an information system has often been ignored, especially in the design stages of the system. This is a serious shortcoming, especially when an automated system is going to be introducted into a previously manual environment.

If the automated system is to maintain viability, the capability must exist to automate manual data when it is cost efficient to do so, and to provide the user with the ability to simply index manual data such as field notes, which are not cost efficient to automate. This amount of interface is depicted by the degree of overlap between the manual circles and the automated circles in Figure 2.1. As the system becomes more automated, the degree of overlap increases. A point of complete overlap comprises a totally automated system.

3.0 THE MANUAL SYSTEM

The manual map file system will store and index paper copy maps, tables, and charts. This will be accomplished through an organized set of map file cabinets and a series of indexes. The indexes would exist in both an automated and manual form.

3.1 Methods of Indexing Data

All data (maps, charts, tables, surveys) that cover a test area will be stored and indexed in the manual system (see Report 1.4 for case study discussion). These spatially related data will be indexed by the following data characteristics:

1) Location, such as state, county, and USGS 1:250,000 quadrangle. This provides an immediate tie to an index map.
2) Subject (soils, vegetative cover, land use)
3) Format (maps, aerial photos, tables, charts)
4) Physical location. Since this manual index will include data stored at the Billings Area Office, WELUT offices, Region Six offices, and other FWS offices, its physical location will be indicated.

3.2 Construction of the Manual System

Five components make up the structure of this manual system. They include:

1) A set of large map file cabinets to be maintained at the WELUT offices, at the Billings Area Office, and in the Denver Regional Offices. Maps are stored in these cabinets in an organized manner. Special arrangements are also required for aerial photographs.
2) A set of file cabinets for storage of descriptive data. These descriptive data would include such items as reports, U.S. Census data, wildlife survey data, and water quality data. Where applicable, the descriptive data will be cross-referenced with the image data (maps and aerial photographs).
3) One card catalog index of all data in all three offices (BAO, WELUT and Denver). This card catalog, much like a library card index system would index each map, table or chart, by the 4 characteristics listed above in 3.1. An identical copy of this card index could be placed in every biologist's office.
4) An automated file which contains the card catalog index. This automated file would be updated when new maps are located and stored in any office (Denver, WELUT, or BAO).
5) A uniform control and monitor process in which an established methodology is used to evaluate new data sources and determine whether or not they should be included in the manual system. The control and monitor process would also be used

to determine if entities should be changed or deleted and who has access to what data.

This automated file could be printed out to produce new card index sets. The automated file can be sorted or subsetted to deliver a specialized card index for selected topics or geographic regions.

The above components represent the general conceptual structure of a manual geographic information system. However, before the manual system can be implemented, much more detailed specifications are required. These specifications would be centered upon: (1) the basic storage equipment, (2) document imput forms and document editing, sorting and output routing, (3) protection from people damage, molds, and rodents, (4) back-up storage capabilities, (5) reproduction facilities, (6) updating, and (7) staffing and budgeting.

3.3 Organization of the Manual System

The manual geographic information system would have the following organizational characteristics:

1) The map file maintenance and computer support for this system would be centralized in a small (2-4 man) WELUT or Region 6 technical staff office. Copies of card index files, and remote terminal access to the automated files would be available to Area Offices in remote locations. Large computer printouts could be delivered by mail.

2) As the system became larger, more frequently used, and better understood, Area offices and other remote users would perform more of the data entry and retrieval tasks.

4.0 THE AUTOMATED SYSTEM

In order to effectively meet the geographic data needs of BAO, the Regional Offices, and OBS Special Projects, an automated Geographic Information System is recommended to operate in parallel with the manual system. This Geographic Information System (GIS) must be able to serve the common needs of the user groups in a timely, accurate and cost effective manner while remaining compatible with existing manual systems of handling data.

Geographic Information Systems is not a field by itself, but rather a common ground between information processing, geography, geology, biology, land use planning and the many other fields utilizing spatial analysis techniques. The boundary condition that separates GIS from other types of information systems is the spatial nature of the data.

4.1 Overall System Design

Given this broad framework for Geographic Information Systems, the next phase in the system design process is a conceptual design of the proposed automated system. Such a design says nothing about exact processes, physical characteristics of hardware or software, or data characteristics; it is more the initial expression of the general systems capabilities, logical structure, and boundary conditions of the proposed system based on the present and future needs of the perceived users.

The capabilities and structure of any GIS can be described in terms of six major subsystems (Figure 4.1): 1) a data input subsystem, 2) a data base management subsystem, 3) a data display and analysis subsystem, 4) an information monitoring subsystem, 5) a user interface language processor subsystem, and 6) an implied administrative management subsystem. Given the needs of FWS users, five of the six subsystems were outlined in Figure 4.1. Except for the management subsystem, each of the five GIS subsystems are discussed below in terms of general characteristics. No statements are made regarding implementation methods or operational procedures.

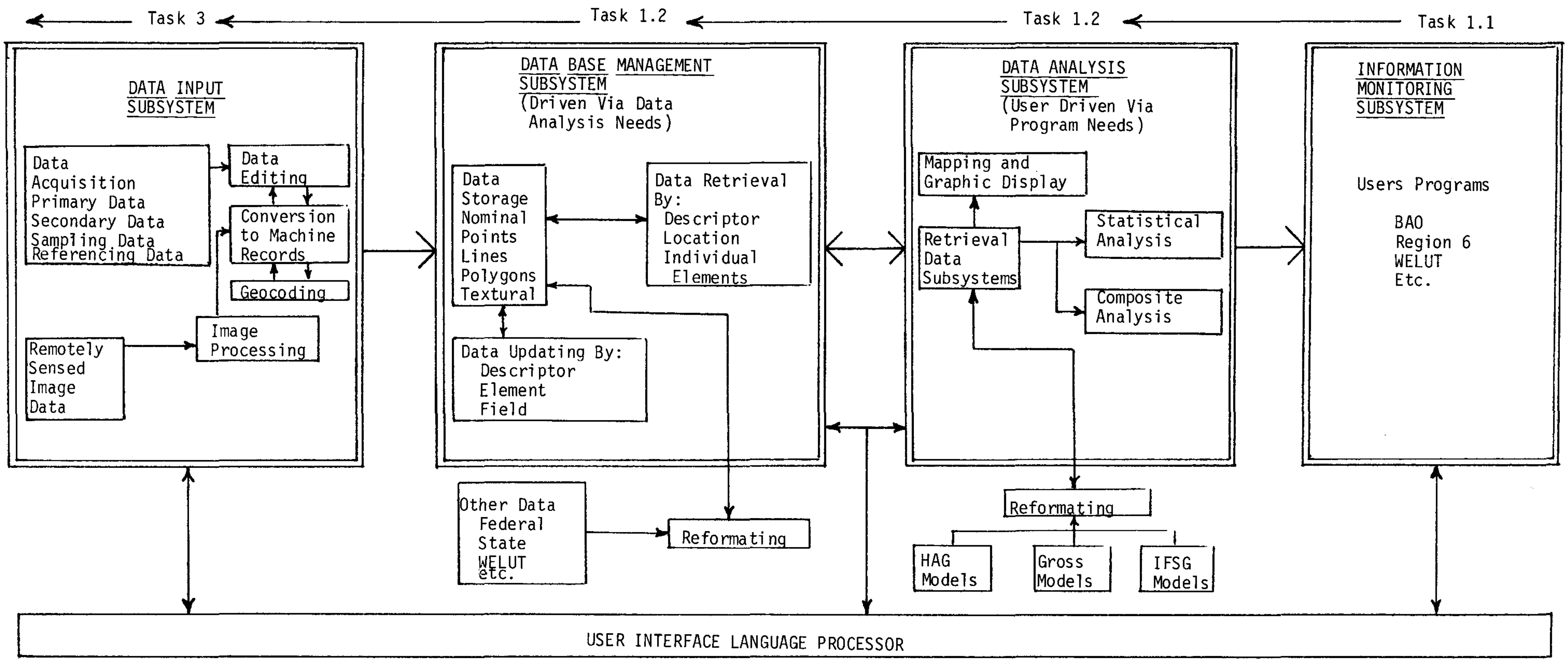
4.2 Information Monitoring and Use Subsystem

The Information Monitoring and Use Subsystem is the point at which information produced by the systems is delivered to the user and is brought to bear on a problem of concern. The key to this subsystem is the monitor process which allows for users to express their satisfaction or dissatisfaction with the products and operation of the system. Feedback, sometimes referred to as the theory of the "screaming user," begins during the initial design process and continues during the life of the system. The feedback is used to correct weak points, add new capabilities, and delete unused capabilities.

The second monitor function is more internal to the system but again relates back to the users of the system. The internal monitor function acts as an accounting system and keeps track of

Figure 4.1

PRELIMINARY GEOGRAPHIC INFORMATION SYSTEM DESIGN



several operational aspects of the system. Some of the aspects monitored might be:

(1) how often a particular data base is accessed;

(2) number of maps generated;

(3) time required to perform a particular analysis;

(4) an update file containing information on when data bases were updated; and

(5) accounting by user i.d. codes.

Internal monitoring is often referred to as tuning. The information provided by a tuning procedure can show where the system is too expensive to run, where bottlenecks occur, or what are the basic types of questions addressed to the system. It allows for the continual optimization and evolution of the system based on what the users are doing.

4.3 Data Analysis Subsystem

The Data Analysis Subsystem serves two primary functions. First, it provides a basic set of techniques for doing spatial display & analysis of geographically referenced data. Second, it allows the GIS to interface with existing and planned FWS modeling efforts. The specific spatial display and analysis techniques implemented depend on the types of analysis & display most commonly required by the different branches & projects within FWS (refer to report 1.1). They can be grouped under three major headings: 1) mapping & graphic display; 2) statistical analysis & report generation; 3) map overlay analysis, and 4) modeling efforts. These three analysis & display groups operate on the data retrieved from the data base given a set of user criteria. The analysis techniques themselves may also generate new data bases that can be later analyzed by the same or a different user. For example, correlation analysis generates a correlation matrix, which could be stored as a data base & then retrieved for later analysis. The three major analysis & display groups are now discussed.

4.3.1 Mapping and Graphic Display

Mapping & graphic display is the process of converting the automated image & descriptor data into graphic format. Mapping is the cartographic representation of the image data. Examples of map display would be contour maps, choropleth maps, or dot maps. Maps can be produced on a variety of output devices. The usual output devices are the line printer, digital plotter, cathode ray tube, & microfilm plotter. Figures 4.2 through 4.5 are examples of some common map displays.

Graphics are pictorial representations of the data such as bar graphs, charts, and line graphs.

4.3.2 Statistical Analysis and Report Generation

Statistical analysis and report generation is the process of summarizing data into some standardized format that is usable by the GIS user. Statistical analysis can range from the simple calculations of means to sophisticated multi-variate techniques, such as discriminate analysis and factor analysis. While the GIS should be able to generate simple summary statistics, the majority of statistical analysis will probably be done external to the GIS. The GIS will construct a data matrix that could be input into such statistical packages

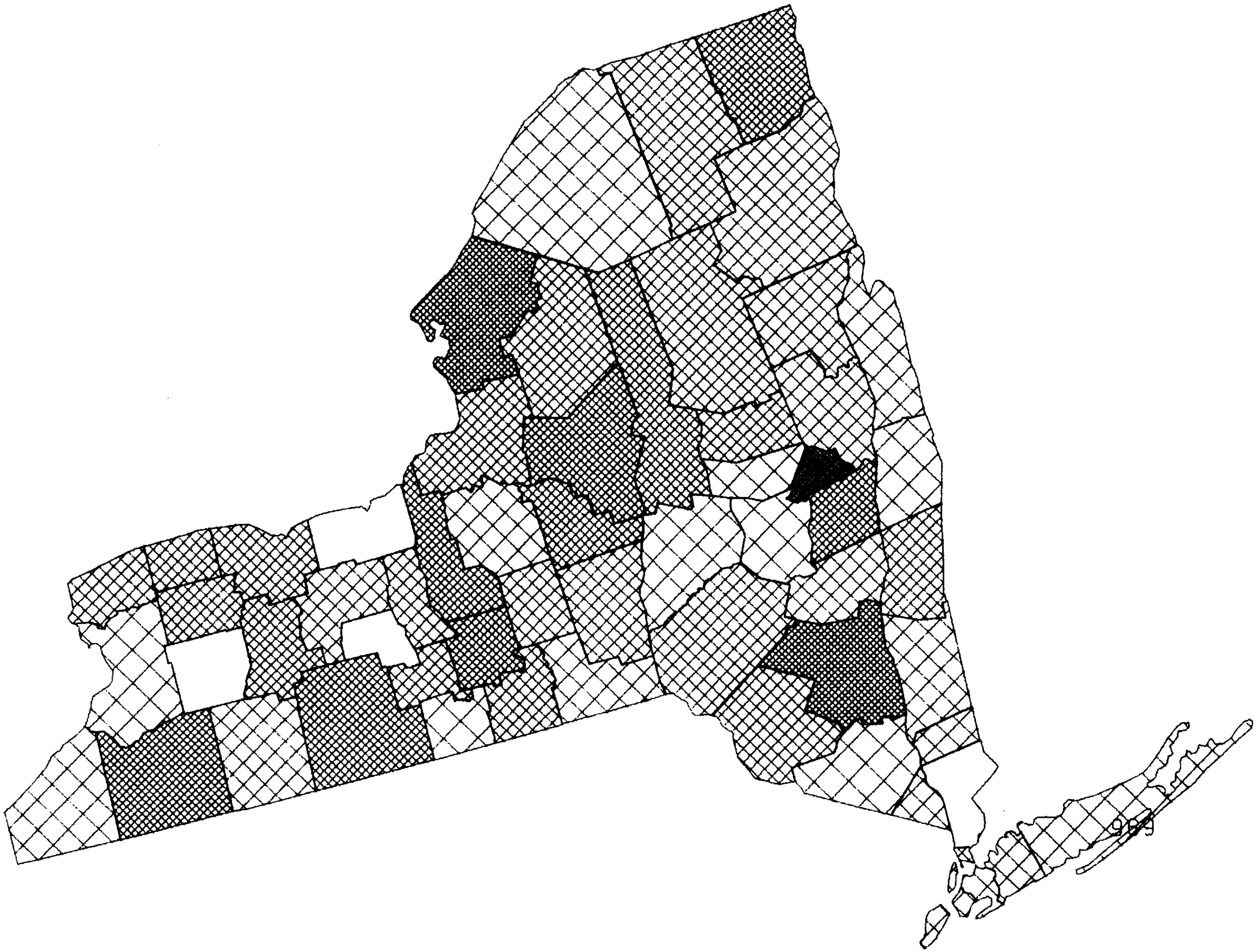
FIGURE 4.2

LINE PRINTER CONTOUR MAP

Program CONTOURING:

Photographically reduced to 50% of printout size.
Pictorial results of a gravimetric survey.

FIGURE 4.3 PLOTTER CHOROPLETH MAP



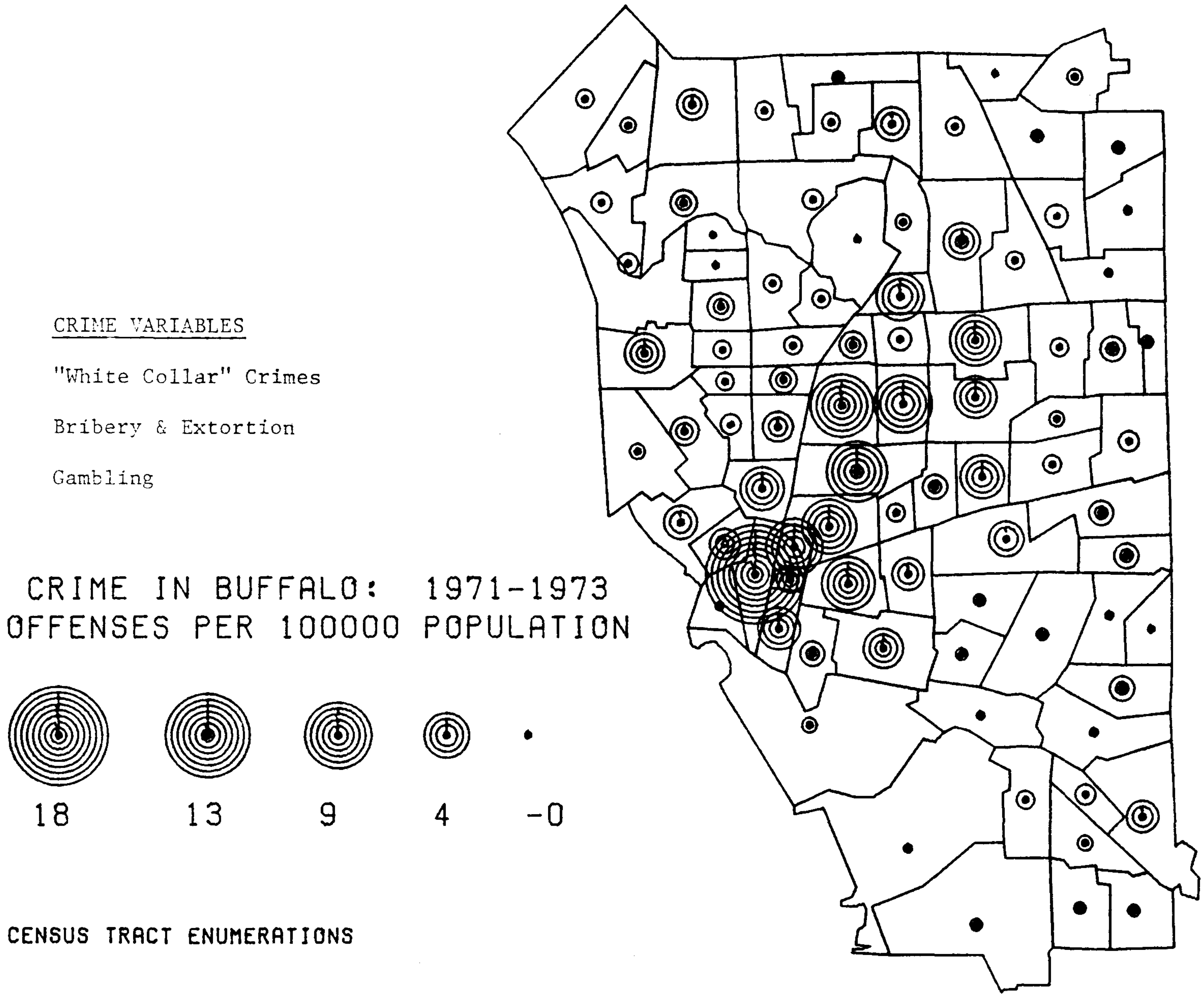
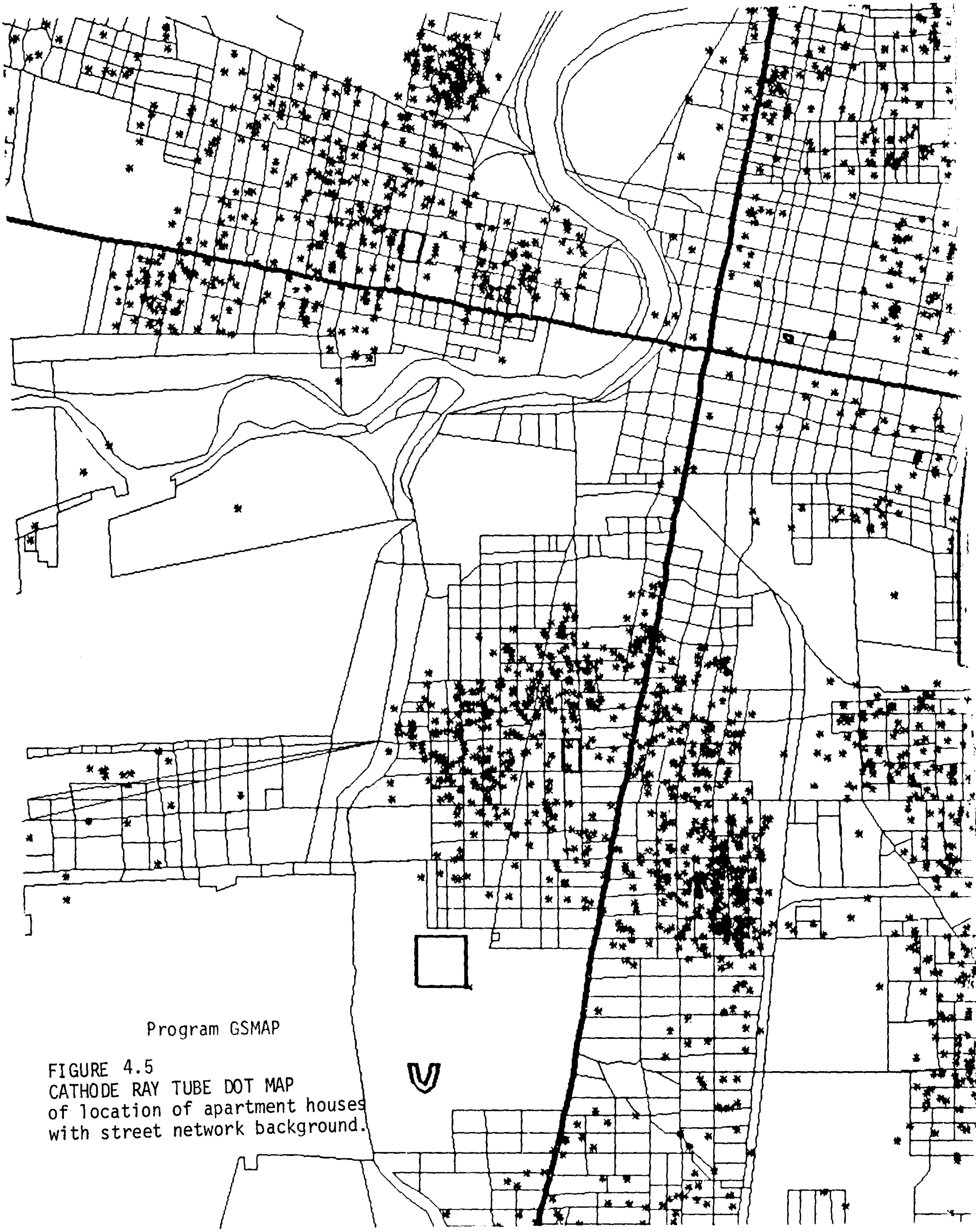


FIGURE 4.4
DIGITAL PLOTTER
GRADUATED CIRCLE MAP



Program GSMAP

FIGURE 4.5
CATHODE RAY TUBE DOT MAP
of location of apartment houses
with street network background.

as SPSS (Statistical Package for the Social Sciences) or BMD II (Biomedical Statistical Package). Systems such as SPSS provide extensive capabilities for statistical analysis, data recoding transformation and scaling. Report generation also provides summary information. Reports are usually in tabular form. Figures 4.6 and 4.7 are examples of Report Generation.

4.3.3 Map Overlay Analysis:

Map overlay (compositing) is the process of overlaying different maps for a given area. For example, to produce a habitat map for a given species, it might be necessary to overlay a vegetation map, a topographic map, and a species map. The composite or overlay process can be applied to cellular, polygon, or point data sets. Figures 4.8 and 4.9 show the overlay process for a cellular data set.

4.3.4 Modeling:

Since this project is not oriented toward the construction of new specialized models, the ability must exist to reformat existing data subsets and feed these data into existing and planned biological or economic models. For example, many of the OBS special projects are specifically geared toward model development and implementation. The GIS being designed must be able to interface with these models for both input and output.

4.4 Data Base Management Subsystem

The Data Base Management Subsystem (DBMS) maintains the geographically referenced data base. A data base may be defined as a collection of inter-related data stored together without harmful or unnecessary redundancy.

AGRICULTURE CAPABILITY OF LAND CONVERTED TO URBAN BETWEEN

1964 and 1968

****** SUMMARY OF AREAS SELECTED ******

PLU64 - PRESENT LAND USE(1964)

USE	AREA	PER CENT
B	0 0	0.0
8	0.0	0.0

PLU68 - PRESENT LAND USE(1968)

USE	AREA	PER CENT
B	4,851.1	100.0

PLU73 - PRESENT LAND USE(1973)

USE	AREA	PER CENT
B	4,851.1	100.0

AGRCL - AGRICULTURE CLASS

CLASS	AREA	PER CENT
1	666.9	13.7
2	519.0	10.7 ←
3	2,049.0	42.2
4	727.2	14.9
5	230.8	4.7
6	261.9	5.4
7	152.7	3.1
0	163.8	3.3
8	79.4	1.6

TOTAL AREA SELECTED IS 4,851

Fig. 4.6

Computer-produced list of measurements performed on maps in digital form.

The quality of agricultural land urbanized between 1964 and 19[illegible] is shown.

(CGIS System)

AGRICULTURE CAPABILITY OF LAND CONVERTED TO URBAN

BETWEEN 1968 and 1973

****** SUMMARY OF AREAS SELECTED ******

PLU64 - PRESENT LAND USE(1964)

USE	AREA	PER CENT
B	0.0	0.0
8	0 0	0.0

PLU68 - PRESENT LAND USE(1968)

USE	AREA	PER CENT
B	0.0	0.0
8	0.0	0.0

PLU73 - PRESENT LAND USE(1973)

USE	AREA	PER CENT
B	12,685.1	100.0

AGRCL - AGRICULTURE CLASS

CLASS	AREA	PER CENT
1	581 5	4.5
2	2,547.9	20.0 ←
3	3,526.0	27.7
4	1,299.8	10.2
5	1,586.3	12.5
6	1,310.5	10.3
7	1,011.5	7.9
0	259.5	2.0
8	561 8	4.4

TOTAL AREA SELECTED IS 12,685

Fig. 4.7

Computer-produced list of measurements performed on maps in digital form.

The quality of agricultural land urbanized between 1968 and 1[illegible] is shown.

(CGIS System)

Fig. 4.8 LOGICAL SELECTION AMONG CELLS OF SEVERAL MAPS

Typical Problem: to find all cells having certain common attributes.

Topic 1
ENVIR. SUITABILITY
FOR RESIDENCE
ordinal scale
1 = best
.
.
.
9 = worst

```
11122255588844444
112223333388444
6666677778844
66667777788
344666666
4445555     IF.TOPIC1 LT 6
444444
3333
```

Topic 2
ZONING
nominal scale
A = R1, R2, R3
B = Commercial
C = Industrial
D = Flood Zone
E = Airport Approach
F = Not Yet Zoned

```
AAAAAADDEEEEEEEEE
AAAAAADEEEEECCC
AABBDDDCCCCC
AABBDDDCCFF
AAAADDCCF
AAAABBC   AND.TOPIC2 EQ A
FFFB
FFC
```

THEN .TEMP-1 EQ M

Topic 3
LAND PRICE
interval scale
1 = below $500/acre
2 = 501 - 999
3 = 1000 - 1999
4 = 2000 - 3999
5 = 4000 - 7999
6 = 8000 and above

.TEMP-1

```
MMMMMM--------
MMMMMM------
-----------
---------
MMM----
MMMM--  IF.TEMP-1 EQ M
-----
---
```

```
66442111333222222
666521113332222
6666411144555
456631114455
444443115
3333666     AND.TOPIC3 LT
33344
333
```

PHYSICAL - ECONOMIC SUITABILITY

```
--XXXX----------
----XX--------
-------------
------------
XXX-------
XXXX-----
--------   THEN.NEWTOPIC EQ X
------
===
```

Fig. 4.9 ARITHMETIC OVERLAY OF CELLULAR MAPS

Typical problem: To convert different topic maps into comparable legends (ordinal, i.e. rank ordered), then to assign the maps different relative weights, and finally to sum up these cell values into a composite map.

FOUNDATION SOILS

Val.	Code
excellent	1
medium	2
inferior	3

Map Weight, mult. ea. cell by 1.0
(importance of this map)

UTILITY AVAILABILITY

Val.	Code
excellent	1
medium	2
poor	3

Map Weight, mult. ea. cell by 2.0
(importance of this map)

TRAFFIC IMPACT

Val.	Code
heavy	3
medium	2
light	1

Map Weight, negative .... −3.0
(importance of this map)

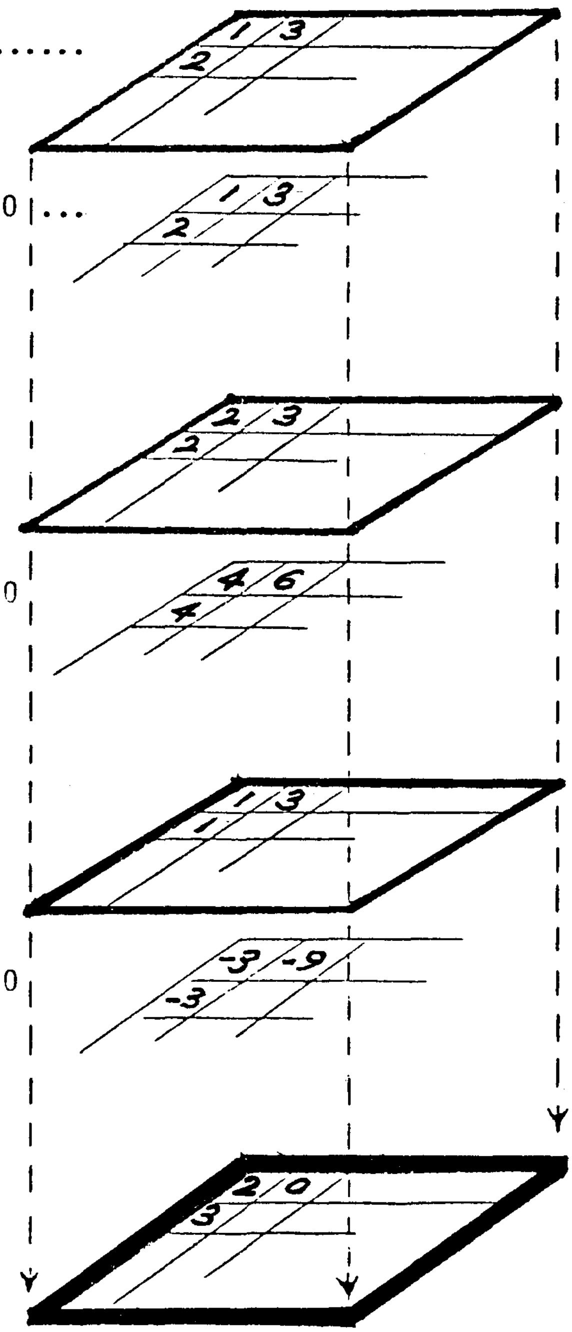
COMPOSITE MAP OF
RESIDENTIAL SITE SUITABILITY

superior = 1

. .

. .

inferior = 9



The data stored are as independent of the programs that use the data as possible. A common or controlled approach is used to add new data and modify and retrieve existing data within the data base. Figure 4.10 presents a basic set of objectives for a data base organization. These objectives are explicitly defined by the operating structure and analysis requirements of the potential users.

A data base management system designed with these objectives in mind, provides three basic forms of activity:

(1) - data retrieval

(2) - data updating

(3) - data interface with other agency files.

4.4.1 Data Retrieval

Data retrieval is the process of extracting data from the data base given a set of criteria. The criteria set may be descriptor criteria, locational criteria, individual element criteria, or a combination of these criteria. Some examples of types of data extraction required of a GIS are:

- Locate a defined item - image and descriptor data (single data set)

  example: Locate the Fort Peck Reservoir in Montana

- Locate a set of defined items - image data and descriptor data

  example: Locate all reservoirs on the Three Forks, Wyoming, 1:250,000 scale map.

FIGURE 4.10 Objectives of Data Base Organization

PRIMARY OBJECTIVES OF DATA BASE ORGANIZATION

- The Data Can Have Multiple Uses
  Different users who perceive the same data differently can employ them in different ways.

- Intellectual Investment is Protected
  Existing programs and logical data structures (representing many man-years) will not have to be re-done when changes are made to the data base.

- Low Cost
  Low cost of storing and using data, and minimization of the high cost of making changes.

- Less Data Proliferation
  New application needs may be met with existing data rather than creating new files, thus avoiding the excessive data proliferation in todays tape libraries.

- Performance
  Data requests can be satisfied with speed suitable to the usages of the data.

- Clarity
  Users can easily know and understand what data is available to them.

- Ease of Use
  Users can gain access to data in a simple fashion. Complexity is hidden from the users by the data base management system.

- Flexible Usage
  The data can be used or searched in flexible ways with different access paths.

- Unanticipated Requests for Data can Be Handled Quickly
  Spontaneous requests for data can be handled without application programs having to be written (a time consuming bottleneck), by means of high level query or report generation languages.

- Change is Easy
  The data base can grow and change without interferring with established ways of using the data.

- Accuracy and Consistency
  Accuracy controls will be used. The system will avoid having multiple versions of the same data items available to users in different stages of updating.

- Privacy
  Unauthorized access to the data will be prevented. The same data may be restricted in different ways from different users.

(Figure 4.10 - continued)

- Protection from Loss or Damages
  Data will be protected from failures and catastrophes, and from criminals, vandals, incompetents, and persons who might falsely update them.

- Availability
  Data are quickly available to users at almost all times when they need them.

[1] Martin, J. 1976 Computer Data Base Organization. John Wiley.

- Locate an item or set of items with criteria undefined - image and descriptor data.

  example: Locate the largest reservoir in the state of Montana in terms of average water acreage

- Locate items based on defined relationships - image data and descriptor data.

  example: Locate all reservoirs within Larimer County within 1 mile of a dairy farm.

- Locate a set of items where criteria are another data set-image and descriptor data.

  example: Locate all reservoir sites in Larimer County that have non-porous soils.

4.4.2 Data Updating

Data updating allows the user to add, delete, and change data in the data base. For example, if a user discovers that the descriptive data describing a vegetation cover polygon is inaccurate, he should be able to delete the inaccurate information and add corrections. He should also be able, if required, to delete the entire polygon and associated descriptive information. Data updating allows for the integrity of the data bases to be maintained. Before an update procedure may be initiated, it will be necessary to specify a set of regulations governing the update process. These regulations provide a standardized approach to the update process, protect the data bases from unauthorized change, and in general protect the data, as much as possible, from human damage.

4.4.3 The Data Interface

The data interface with other agencies will allow the FWS system to exchange data between FWS files and other state and federal agencies files.

An example of this process might occur when FWS requires water quality data. The Environmental Protection Agency (EPA) maintains a large data base on water quality (STORET). The FWS GIS should be able to integrate EPA data into its own structure.

4.4.4 Data Structure

These three basic forms of data management activity; data retrieval, data update, & data transfer, are tied to a common data structure. A common data structure is the way the data are logically & physically stored in the computer. The data structure design is based on the type, volume, & retrieval characteristics required by FWS. Presently, FWS geographic data requirements suggest that the data structures will have to maintain:

- nominal data (water shed number)
- points (well locations)
- lines (streams)
- polygons (soil types)
- descriptive (data describing points, lines and polygons)
- textual (indexes, bibliographies, etc.)

The actual logical and physical storage structures have not been designed yet.

The actual decisions on the physical storage structure will depend on the types of data to be handled, the desired response speed to a user query, and the volume of data to be handled. The volume of data to be processed is one of the more important factors. One of the characteristics of geographic data bases is very large data volumes. Volumes of 1,000,000 characters are not uncommon. Figure 4.11 provides some preliminary calculations of data volumes for selected programs in BAO and the Regional Offices. These numbers represent the number of records for one variable using a cellular or grid coverage (cell volumes are easy to calculate).

Data that are in the format of polygons could have smaller or larger data volumes for the same coverage depending on how many points are used to specify the boundary of the polygon. For most purposes, the volume of data for polygons will be less than for cells. Volume is also an important consideration during the data input process.

4.5 Data Input Subsystem

The data input subsystem allows the user to transform a manual source document, such as a map, into machine readable form for entry into the data base. The data input subsystem is comprised of five basic manipulations:

1) Data acquisition
2) Image processing of remote sensor data and aerial photography

Figure 4.11

DATA VOLUME ESTIMATES FOR FWS*

PROGRAM	SCALE-AREA	No. of RECORDS
BAO**		
Land Management	1:10,000	107,000
Habitat Species	1:50,000	21,400
Stream Survey***	1: 1,000	500,000
	1:24,000	20,833
Utilization of Fish & Wildlife	1:25,000	161,280
Geographic Referencing	1:10,000	200,000
	1:24,000	83,300
Regional Mineral	1:50,000	83,300
	1:50,000	40,000
Stream & Structural Physical & Biological Data Survey**	2.5 acres	20,833
Endangered Species	1:24,000	533,300
	1:25,000	51,200
Ownership & Management Plans		51,200
Coal Coordination****		
Habitat Species	1:50,000	806,400
Regional Mineral	1:50,000	806,400
Ownership & Management	1:50,000	806,400
Regional		
YAMPA		
Habitat Species	30 acre	143,300
Regional Mineral	30 acre	143,300
Stream Structural	30 acre	143,300

* The formula used was: Area/(Scale/1000)

** BAO estimates for CMR only, which is @ 1.1 million acres

***Streams data were considered to require half the coverage of other data; such as vegetation and soils.

****Entire State of Wyoming

3) Conversion to machine records

4) Geocoding

5) Data editing

Data acquisition is the process of obtaining data in manual or automated form that fulfills the requirements for a particular project. The data acquisition function must deal with the data, their characteristics, the people and institutions that supply them, and their intended uses. If possible, the following characteristics should be determined for each data source before acquisition:

Data Characteristics

- location, ownership, and availability
- data format and volume
- documentation availability
- data classification systems

Institutional Consideration

- procedures for acquisition
- continuity of data flow
- agreements to supply data
- committee of data suppliers
- opportunities to change procedures
- proprietary nature of data

Evaluation of Data

- applicability to problem
- assessments of accuracy and reliability

Many forms of source data are available as maps, tables, aerial photos and multi-spectral remote sensor imagery. These data may be primary data, which are unprocessed raw data, such as a field geologist's field maps, secondary data, which are organized raw data such as the U.S. Census data, sample data, such as animal population surveys, or referencing data, such as the USGS graticules.

In addition to acquisition, some data will require processing before conversion to computer format. Remote sensed data requires additional processing. This is called "image processing" and usually includes rectification and interpretation of air photos and rectification and classification of any satellite data.

Conversion to machine records is the actual process of converting the source documents into digital form and inserting the resultant data into the data base. Running in parallel with the conversion is the process called geocoding. In the case of the FWS GIS, geocoding is the assignment of a geographic reference to each data item. This may involve a geographic transformation process. Geographic transformation allows for the conversion from various map projections, such as Transverse Mercator or Albers Conformal, into a common projection. This greatly simplifies the structuring of the data base and facilitates retrieval.

Input data editing is the process of correcting errors resulting from the conversion of source documents into machine readable form. Errors may arise due to human error, software problems, source document errors, or hardware errors. These errors could be incorrectly spelled names, misassignments of codes to polygons, inaccurate air photo interpretation, or inaccurate digitizing of cartographic documents.

Some means must exist to correct as many errors as possible at the point of data entry. Corrections of cartographic data after they are in the data base, can be very costly and time consuming.

There are two primary considerations in the design of the input subsystem: 1) type of source data to be encoded; and 2) the volume of data to be encoded. Considering FWS needs, the GIS will be handling point, line, polygon, cellular (including remote sensor), and textual data. Below are listed the costs of converting these types of data from source format to computer format. These are only averages. Costs may be higher or lower depending on the particular characteristics of the source document.

These costs are only for the actual encoding and data editing. Data preparation costs are not included.*

1) Points - between .07 and .15 cents per point.
2) Lines (arcs) - 30 to 70 cents per inch.
3) Grid - .03 to .10 cents per cell.
4) Textual - .005 to .01 cents per character.

As can be seen, the cost range is great. Since these are only "ball park" figures, costs could be greater or lesser than those shown above. Data preparation could double these costs.

* cost and volume from actual tests, International Geographical Union, 1976.

By way of illustration, given an "average" 7.5 minute U.S.G.S. quadrangle, suppose we wished to digitize all the highways. There are approximately 250 inches of highway lines on a quadrangle. At an average cost of .56 cents per line inch (digitizing and edit) this is a cost of $140 for one quadrangle.

4.6 User Interface/Language Processor Subsystem

The user interface/language processor subsystem allows the prospective user to interface with the GIS. It also protects him from the internal complexities of the system. While the interface allows the user to interact with any portion of the GIS, it is primarily geared toward "Data Base Management" and the "Data Analysis Subsystems."

The interface provides for 3 externally apparent services and one internal hidden service. The internal service is part of the tuning process discussed in 1.3.2 and will not be discussed here. The external services are:

- Tutorial capabilities
- Query capabilities
- System security

A Tutorial capability allows users unfamiliar with the system to be "lead by the hand" through the process. It represents a hierarchical structure: it allows the facile users to suppress the tutorial, the moderate users to refresh their memory, and keeps inexperienced users from having to run to a manual each time there is a problem. The query capability allows the user to phrase a question in an easily understood format, called a query language. The user can thus easily manipulate data, retrieve data, and display and analyze data. The tutorial function is important for maintaining a high technical standard in the system. Yet tutorials have usually been overlooked in other GIS development.

System security prevents unauthorized users from entering the system, prevents unauthorized use of certain sections of the data base, and prevents a user from performing a function that could destroy data, cost too much, or degradate system integrity.

5.0 SUMMARY

The function of this paper has been to outline the general conceptual design of a GIS for FWS. A broad set of system capabilities was presented. This represents the first major step in the design process.

The next steps will serve to state the system capabilities in much finer detail, especially at the technical level. The detailed system capabilities can be stated once a better grasp of data sources, data volumes, specific types of analysis, and sources of available (usable) computer software have been specified. These are the primary functions of Task 2 and Task 3 of this project. Basically, these functions will involve 1) detailed analysis of the characteristics of the required data, 2) analysis of the output requirements for the system, and 3) reflecting these first two analyses, the development of a very detailed system description. The detailed system description can then be used as the basis for software selection.

Appendix A

GLOSSARY OF TERMS USED IN THIS REPORT

Geography:	Analysis of the patterns, structures, arrangements, and relationships of man's work on the surface of the earth (human geography) and of the patterns, structures and relationships of the earth itself (physical geography).
Spatial:	A key word used by geographers to indicate that they are interested in the relationships of phenomenon located in two and three dimensional space either in the abstract (the mathematics of point distributions) or on the ground (analysis of the habitats for various wildlife species).
Geographic Reference:	Explicit two and three dimensional locators of a phenomenon, such as the Latitude-Longitude coordinates of a water well site.
Geocode:	The process of attaching a geographic reference to a phenomenon, such as assigning a Latitude-Longitude to a water well site.
Spatial Analysis:	Techniques of analyzing geographically referenced data.
Spatial Display:	Methods of visually depicting geographically referenced data (contour maps or choropleth maps).
Data:	Data are raw numbers that convey little or no information in themselves.
Information:	Organization of data so that they are in a form useable for decision making.
System:	A system is a collection of entities and activities meaningfully connected and satisfactorily bounded which interact for a common purpose or purposes.

(Appendix A - continued)

Information Systems:	When information or data resources are joined together in both a formal and informal manner, they are considered to have some organized structure. They become an information system to support the decision making processes of an organization.
Geographic Information System:	An information system that can input, manipulate, and analyze geographically referenced data in order to support the decision making processes of an organization.
Data Field:	A data field is a collection of data items within a record. For example, a data field called DATE may be composed of data items MONTH, DAY, YEAR.
Record:	A record is a named collection of data fields. For example, the Joseph Smith may be composed of the data fields EDUCATION, DATE and OCCUPATION.
File:	A file is a named collection of occurrences of a given type of record. For example, a named file called EMPLOYEES may be composed of records for Joseph Smith, Peter Strong, and Harry Jones.
Data Base:	A data base is a collection of the occurrences of multiple record types, containing relationships between records, data fields and data items. For example, personnel files are a data base.
Data Base System:	An organized collection of data bases.
Hardware:	The actual physical equipment of a computer system (line printers, cabinets, teletypes, card readers, etc.).
Software:	A set of instructions, either man generated or machine generated, that tell the computer what to do.
Data Base Management System:	A set of software that allows the user to access and change data in a data base.
Data Input:	The process of transforming data from a manual form to a form that the computer can understand.

Appendix B

<u>ABSTRACT</u> Report 1.1

This report documents the results of a five month survey of user needs for a geographic information system witin Region Six of the U.S. Fish and Wildlife Service (FWS). This report culminates the first of seven project tasks being performed under contract to the U.S. Department of the Interior, U.S. Fish and Wildlife Service, Office of Biological Services, Western Energy and Land Use Team (WELUT).

This project is defined and sponsored by the Western Energy and Land Use Team to promote more effective consideration of fish and wildlife values and resources in the numerous state and federal decisions concerning the rapid development of western land, energy, mineral and water resources.

The goal of this two year project is to develop an operational capability within the Fish and Wildlife Service to accept, store, manipulate and display spatially related data for use in a variety of FWS wildlife resource impact projects.

This task focused on the assessment of three groups of users: (1) the Denver Region Six Offices of FWS, (2) the Billings Area Office within Region Six, and (3) the Office of Biological Services Special Projects. Some twenty individual user offices were interviewed with structured questionnaires and follow-up sessions to identify their legislative mandates, programs, tasks, work elements, and specific spatial data needs.

Seven logical groups of data emerge as those which are required by multiple users. These generic data bundles or data compilation modules include:

1) Habitat-Species Distribution

2) Stream Structural, Physical & Biological Survey Data

3) Economic-Demographic-Land Use Change

4) Seasonal Demand For and Utilization of Fish and Wildlife Resources

5) Geographically Referenced FWS Land Ownership & Lease Data

6) Regional Mineral & Energy Related Activity Pattern Data

7) Ownership & Management Plans of Other Agencies

Individual user groups will require the results of each of these compilation modules to differ by 1) spatial resolution (from 2 1/2 acre polygons for site specific planning to 640 acre polygons for macro regional analysis), 2) data currency or updating frequency (from 3 month currency to 10 year currency), and 3) geographical coverage (from a 200 square mile site specific problem to an 85,000 square mile statewide analysis).

Representative users and cases can be selected from this broad user community. Working on selected test cases for these few users will demonstrate the solution of specific problems which can be applied to the more general group of users.

ABSTRACT Report 1.3

This appendix includes the forms used for the user needs assessment performed under contract to the U.S. Department of the Interior, U.S. Fish and Wildlife Service, Office of Biological Services.

The forms in this report are the results of interviews with twenty groups of users in a series of introductory meetings. The groups interviewed consisted of the following:

1) Ten Billings Area Office Groups
2) Six Denver Region Office Groups
3) Four OBS Special Project Groups

Form #1 identifies the mandates, general programs, and staffing for each organizational element, Form #2 describes the work elements of each program, and Form #3 describes the detailed spatial data needs, processing, and outputs required for each work element. These forms start with general programs and successively move toward greater detail.

ABSTRACT Report 1.4

The "User Needs Assessment for an Operational Geographic Information System" (Report 1.1) identified many FWS programs which would be benefitted by a GIS. To demonstrate the relevance of a GIS to these programs, seven candidate test areas are proposed (Figure 1.1). These test areas are primarily meant to serve as prototype demonstrations of a GIS. However, to make the demonstrations relevant and meaningful to FWS users, the test areas were delineated to address real FWS problems whenever possible.

BIBLIOGRAPHIC DATA SHEET	1. Report No. FWS/OBS-77/22	2.	3. Recipient's Accession No. PB 269 275/AS

4. Title and Subtitle A General Design Schema for an Operational Geographic Information System	5. Report Date June 1977
	6.
7. Author(s) Larry Salmen, George Nez, James Gropper, John Hamill and Carl Reed	8. Performing Organization Rept. No.
9. Performing Organization Name and Address Information Systems Technical Laboratory Federation of Rocky Mountain States, Inc. One Drkae Park, Suite 142, 333 W. Drake Road Fort Collins, Colorado 80521	10. Project/Task/Work Unit No. WELUT No. 002-76
	11. Contract/Grant No. 14-16-0008-2155
12. Sponsoring Organization Name and Address Western Energy and Land Use Team, Office of Biological Services, U.S. Fish and Wildlife Service, Drake Creekside Building, 2625 Redwing Road Fort Collins, Colorado 80521	13. Type of Report & Period Covered
	14.

15. Supplementary Notes

Coal Fund; EPA-IAG-D5-E685; L.C. Number 77-82532.

16. Abstracts

This publication, the second in a four-part report, presents a general system design and outlines a basic set of system capabilities for the U.S. Fish & Wildlife Service's Geographic Information System. The system is to operate at two levels: a manual system for cataloging maps, relevant articles, and other paper documents; and an automated system for data input, data analysis, and information output based on user-specified criteria. The publication is non-technical in format and is geared towards those with little or no computerized-information system experience

17. Key Words and Document Analysis. 17a. Descriptors

mapping
software
data base
test areas
system capabilities
geographic information systems

17b. Identifiers/Open-Ended Terms

U.S. Fish & Wildlife Service, Region 6
Western United States
natural resource decision making

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Release unlimited	UNCLASSIFIED	54
	20. Security Class (This Page) UNCLASSIFIED	22. Price PC:A03/MF:A01