

GIRAS: A Geographic Information
Retrieval and Analysis System for
Handling Land Use and Land Cover Data

GEOLOGICAL SURVEY PROFESSIONAL PAPER 1059



GIRAS: A Geographic Information Retrieval and Analysis System for Handling Land Use and Land Cover Data

By WILLIAM B. MITCHELL, STEPHEN C. GUPTILL, K. ERIC ANDERSON,
ROBIN G. FEGEAS, *and* CHERYL A. HALLAM

GEOLOGICAL SURVEY PROFESSIONAL PAPER 1059



UNITED STATES DEPARTMENT OF THE INTERIOR

CECIL D. ANDRUS, *Secretary*

GEOLOGICAL SURVEY

H. William Menard, *Director*

Library of Congress catalog-card No. 77-600047

**First printing 1977
Second printing 1978**

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D. C. 20402
Stock Number 024-001-03007-1

CONTENTS

	Page
Abstract	1
General system description	1
Source data	2
Land Use and Land Cover Map	2
Political Unit Map	3
Census County Subdivision Map	3
Hydrologic Unit Map	3
Federal Land Ownership Map	3
State Land Ownership Map	3
Data volumes	4
Data structure	5
Data capture	7
Data retrieval	11
Data manipulation, analysis, and output	11
Data manipulation	11
Data analysis	12
Data output	12
GIRAS II, current system design	13
References	16

ILLUSTRATIONS

	Page
FIGURE 1. General system flow of GIRAS	3
2. Topological elements of a polygon map	5
3. GIRAS file structure	6
4. Map header control points	7
5. Graphic input procedures for GIRAS	10
6. Arc-to-polygon program sequence	11
7. Island routines within the arc-to-polygon program	11
8. Final arc-to-polygon error-checking sequence	12
9. Design features and contrasting capabilities of GIRAS I and GIRAS II	14
10. System flow diagram of GIRAS I and GIRAS II	15

TABLES

	Page
TABLE 1. U.S. Geological Survey land use and land cover classification system for use with remote sensor data	2
2. Federal land ownership	4
3. Data volumes of a sample of digitized map overlays	4
4. Estimated data volumes, 1975-1981	5

GIRAS: A GEOGRAPHIC INFORMATION RETRIEVAL AND ANALYSIS SYSTEM FOR HANDLING LAND USE AND LAND COVER DATA

By **WILLIAM B. MITCHELL, STEPHEN C. GUPTILL, K. ERIC ANDERSON, ROBIN G. FEGEAS, and CHERYL A. HALLAM**

ABSTRACT

The U.S. Geological Survey is currently producing land use and land cover maps and associated overlays (e.g., political units) for the United States. These maps are being digitized, edited, and incorporated into a digital data base. The data will be available to the public in both graphic and digital form, and statistics derived from the data will be published. To accomplish these tasks the Geographic Information Retrieval and Analysis System (GIRAS) has been designed and developed. GIRAS is designed to accept digitizer input, provide comprehensive editing facilities, produce cartographic and statistical output, permit retrieval and analysis of data, and exercise data base management tasks. The editing and output procedures are currently operational and utilize an arc segment, polygon approach in a production mode. The system incorporates facilities for gridding the polygon data to make it compatible with grid-based data sources. Current system development is focused upon an interactive data base to enable immediate retrieval and display of map information. Users will be able to search for either locations or attributes and display results in a graphic or tabular form.

GENERAL SYSTEM DESCRIPTION

The accelerating need to facilitate the rational planning, management, and utilization of natural resources and the environment in the United States has been linked, in recent years, with an accompanying need for improved techniques and methods for the storage, analysis, and display of large quantities of spatial environmental data. Concomitant with these needs, rapid developments in the past decade in computer technology and in the applications of computers to environmental and natural resources data handling have promoted the creation of increasingly sophisticated spatially-oriented information systems. Understandably, those agencies of the Federal Government, planning bodies in the regions and States of the United States, and the private sector which are concerned with the planning and management of the nation's environmental and natural resources are likely to be in the vanguard of those system developments. Early in 1976, for example, more than 50 system activities in the U.S. Geological Survey were specifically concerned with gathering and handling spatial information in the

fields of geology, geography, topography, and water resources.

The Geographic Information Retrieval and Analysis System (GIRAS) is one such system. It has been designed to input, manipulate, analyze, and output digital spatial data developed for the land use and land cover mapping and data compilation program of the U.S. Geological Survey. The program is designed to provide systematic and comprehensive collection and analysis of land use and land cover data on a nationwide basis. In addition to the land use and land cover maps at scales of 1:250,000 and 1:100,000, associated maps are compiled showing hydrologic units, political units, census county subdivisions, Federal land ownership, and in some instances, State land ownership. GIRAS comprises three basic subsystems to handle that data: (1) Data input, (2) data retrieval and manipulation, and (3) data output. It incorporates a data base structure designed specifically for handling spatial data and provides comprehensive facilities for data editing, data manipulation, and graphic and statistical output. The initial development and current production mode of GIRAS I uses standardized batch computer routines for input and output processing. The general system flow of GIRAS I is shown in figure 1.

GIRAS I is a batch-oriented sequential system on which development began in 1973. Initial emphasis was on the editing and correction of digitized land use and associated data bases, the assemblage of simple archival data files based upon individual 1:250,000 topographic quadrangles, and the application of rudimentary data retrieval routines to produce "standardized" statistical listings and plotter graphics. The driving force of that initial emphasis was the clear need to provide basic standard statistics to the cooperating agencies. These included the Louisiana State Planning Office, the Florida State Planning Office and ten regional planning councils in the State of Florida, the Ozarks Regional Commission, and various Kansas State agencies. The recognition of the relevance and ap-

plication of up-to-date land use, political, hydrologic, census, and Federal and State land ownership data which could be stored, manipulated, and retrieved through modern computer-aided techniques have developed in the direct relation to the sophistication of planning organizations, their exposure to geographic information systems, and the availability of new planning data. The initial reaction — and the initial requirement — of cooperative users was for statistics and graphics that can be produced by the most basic of geographic information systems. Areal data and associated attributes are transformed to a fixed grid system which can produce simple composites of graphic data and simple tabular listings.

This report describes the facilities and procedures of GIRAS I, a batch-oriented operational system. Sections of this report that follow describe in more detail the source and characteristics of the data being processed in GIRAS I, the data structure being used, the procedures for data capture and editing, data retrieval, data manipulation, and data output. The final section of the report describes the system design features that have been established for the interactive GIRAS II which has been under development in parallel with GIRAS I since 1975.

SOURCE DATA

The basic set of data presently produced in the program includes:

1. Land Use and Land Cover Map,
2. Political Unit Map,
3. Hydrologic Unit Map,
4. Census County Subdivision Map,
5. Federal Land Ownership Map, and
6. State Land Ownership Map (optional).

LAND USE AND LAND COVER MAP

The basic purpose of this map is to provide land use and land cover data to be used as a data source in itself or in combination with the other data sets produced in the program. One of the basic sources for land use compilation is the NASA high-altitude U-2/RB-57 aerial photocoverage, usually at scales smaller than 1:60,000. The 1:250,000-scale topographic map series is used as the base map for the compilation of the land use and land cover and the associated overlays, with the exception that the 1:100,000-scale topographic map base is used if that base map is available at the time the data set is released to the open file by the Geological Survey. Although compilation of land use and land cover is performed on a film-positive base enlarged to approximately 1:125,000, digitizing is performed at 1:250,000. The associated overlays are both compiled and digitized at 1:250,000.

TABLE 1. — U.S. Geological Survey Land Use and Land Cover Classification System for use with remote sensor data

Level I	Level II
1 Urban or built-up land	11 Residential
	12 Commercial and services
	13 Industrial
	14 Transportation, communications, and services
	15 Industrial and commercial complexes
	16 Mixed urban or built-up land
	17 Other urban or built-up land
2 Agricultural land	21 Cropland and pasture
	22 Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
	23 Confined feeding operations
	24 Other agricultural land
3 Rangeland	31 Herbaceous rangeland
	32 Shrub and brush rangeland
	33 Mixed rangeland
4 Forest land	41 Deciduous forest land
	42 Evergreen forest land
	43 Mixed forest land
5 Water	51 Streams and canals
	52 Lakes
	53 Reservoirs
	54 Bays and estuaries
6 Wetland	61 Forested wetland
	62 Nonforested wetland
7 Barren land	71 Dry salt flats
	72 Beaches
	73 Sandy areas other than beaches
	74 Bare exposed rocks
	75 Strip mines, quarries, and gravel pits
	76 Transitional areas
	77 Mixed barren land
8 Tundra	81 Shrub and brush tundra
	82 Herbaceous tundra
	83 Bare ground tundra
	84 Mixed tundra
9 Perennial snow ice	91 Perennial snowfields
	92 Glaciers

Land use and land cover compilation is based upon the classification system and definitions of Level II land use and cover as shown in table 1.

All features are delineated by curved or straight lines which depict the actual boundaries of the areas (polygons) being described. The minimum size of polygons depicting all urban and built-up land (categories 11–17), water (51–54), confined feeding (23), other agricultural land (24), and strip mine, quarry, and gravel pit (75) categories is four hectares. All other categories of land use and land cover have a minimum polygon size of 16 hectares. Those sizes are also considered the minimum sizes to which polygons are digitized. In urban and water categories, the minimum width of a feature to be shown is 200 meters; that is, if a square with sides 200 meters in length is delineated, the area will be 4 hectares. Although this minimum-width consideration precludes the delineation of very

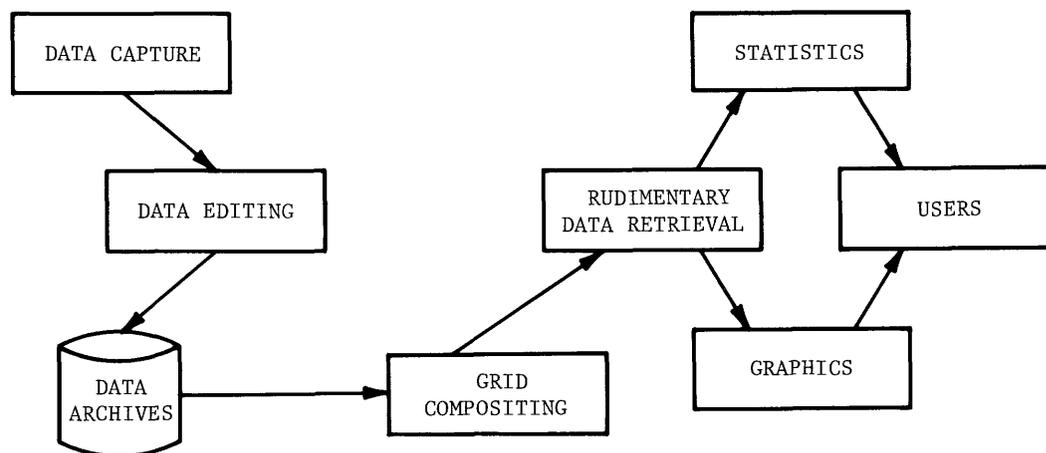


FIGURE 1. — General system flow of GIRAS.

narrow and very long 4-hectare polygons, triangles or other polygons are acceptable if the base of the triangle or minimum width of the polygon is 200 meters in length, and the area of the polygon is 4 hectares. For categories other than urban and water, the 16-hectare minimum size for delineation requires a minimum width polygon of 400 meters. Line weight for delineating land use and land cover polygons and for neat lines is 0.10 mm at the production scales of 1:250,000 or 1:100,000.

POLITICAL UNIT MAP

The Political Unit Map provides a graphic portrayal of the county and State boundaries and is compiled using base maps at either 1:250,000 or 1:100,000. Source material for the Political Unit Maps is from Bureau of the Census unpublished maps entitled "County Subdivisions — Townships and Places" and from the "Geographic Identification Code Scheme" and the "County and City Data Book," both Bureau of the Census publications (U.S. Bureau of the Census, 1972b and c). The "County Subdivisions — Townships and Places" maps are also used to separate Census County Subdivisions into Census Tracts. State and county political subdivisions are encoded with a five-digit number in accordance with the Geographic Identification Code Scheme with the exception that noncontracted "independent cities" in Maryland, Missouri, Nevada, and Virginia are given an eight-digit code reflecting the State and city codes.

CENSUS COUNTY SUBDIVISION MAP

This map provides a graphic depiction of Census Tracts in Standard Metropolitan Statistical counties and Minor Civil Division or equivalent boundaries in non-Standard Metropolitan Statistical Area (SMSA) counties. The overlay is digitized in the same manner as the other overlays and is based upon a Bureau of the

Census publication entitled "Census Tracts" for the SMSA covered by the sheet being compiled (Bureau of the Census, 1972a) and a Bureau of the Census publication entitled "Geographic Identification Code Scheme" (Bureau of the Census, 1972c). Census tracts are encoded with a four-digit number which is unique for each SMSA and a six-digit number which is unique for each census tract. Minor civil divisions are encoded with an eight-digit number: the first two digits are the State code, the next three are the county code, and the last three are the minor civil division identifier.

HYDROLOGIC UNIT MAP

The Hydrologic Unit Map provides a reference for statistics published by the U.S. Geological Survey and is based upon Hydrologic Unit Maps together with the list "Boundary descriptions and name of region, subregion, accounting units, and cataloging unit." The hydrologic units are encoded with an eight-digit number which indicates the hydrologic region (first two digits), hydrologic subregion (second two digits), accounting unit (third two digits), and cataloging unit (fourth two digits).

FEDERAL LAND OWNERSHIP MAP

The U.S. Geological Survey has the responsibility to research, obtain, and format maps, plots, and other descriptive data relating to Federal Land Ownership. Minimum size for the delineation is 16 hectares and ownership is encoded according to the agencies shown in table 2.

STATE LAND OWNERSHIP MAP

In some instances in which the U.S. Geological Survey has a cost-sharing cooperative agreement with a specific State, a map overlay showing an inventory of State-owned land is produced from data furnished by

TABLE 2. — *Federal land ownership*

Code	Agency
Department of Agriculture	
11	Agricultural Research Service
12	Forest Service (National Forest)
13	Forest Service (National Grassland)
Department of Commerce	
21	National Oceanic and Atmospheric Administration
Department of Defense	
31	Air Force
32	Army
33	Army (Corps of Engineers—Civil Works)
34	Navy
Department of the Interior	
41	Bonneville Power Administration
42	Bureau of Indian Affairs (does not include Indian lands held in trust)
43	Bureau of Land Management
44	Bureau of Mines
45	Bureau of Reclamation
46	Fish and Wildlife Service (National Wildlife Refuge)
47	National Park Service (National Monument, Seashore, and Recreational Areas)
48	National Park Service (National Park)
Department of Justice	
51	Bureau of Prisons
Department of State	
61	International Boundary and Water Commission, U.S. and Mexico
Department of Transportation	
71	Federal Aviation Administration
72	Federal Railroad Administration
73	U.S. Coast Guard
Other Agencies	
81	Energy Research and Development Administration
82	General Services Administration
83	National Aeronautics and Space Administration
84	Tennessee Valley Authority
85	Veteran's Administration

the State. Although this overlay is compiled to the same base map as used for the other overlays, the polygons are encoded according to the ownership codes used by the particular State.

DATA VOLUMES

The digitizing operation of the GIRAS has been functional since mid-1975. The records of the data-editing procedures have been used to estimate the data volumes that will be involved in future operations. Table 3 shows some measures of data volumes for the categories of Land Use and Land Cover, Political Units, Census County Subdivisions, Hydrologic Units, Federal Land Ownership, and State Land Ownership for

selected samples of 1:250,000 map quadrangles. The statistics show the numbers of arcs and digitizer coordinates input to the system as a result of digitizing as well as the numbers of arcs and coordinates output after a data reduction process in the editing stage. In addition, the number of polygons and the arc line lengths for the quadrangles are shown. The high, low, and average counts and measurements are also shown for each.

The most dramatic variations indicated are the digitizer coordinates reduced in the editing process while not basically affecting the number of arcs. The greater complexity of the land use and land cover maps is demonstrated by the average number of polygons recorded for them as compared to those recorded for other overlays. Equally instructive is the wide high and low variations in the number of polygons and the arc line length for quadrangles sampled.

Data volumes for the initial 9-year period of the land use and land cover mapping program have been estimated by using the average length of arc lines digitized for land use and land cover and the associated overlays in conjunction with the mapping schedules presently planned for the program. Another estimate of the data volume is provided by the numbers of input coordinate

TABLE 3. — *Data volumes of a sample of digitized map overlay categories*

	Sample size	Average	High
Land Use and Land Cover			
Input arcs	54	7,765	27,686
Output arcs	54	7,340	27,339
Input coordinates	54	877,003	2,170,480
Output coordinates	54	165,384	435,540
Polygons	51	2,986	9,865
Arc length (m)	50	97.91	214.95
Political Units			
Input arcs	49	42	101
Output arcs	47	42	101
Input coordinates	48	43,512	89,124
Output coordinates	47	3,526	7,834
Polygons	54	16	37
Arc length (m)	34	13.79	142.60
Census County Subdivisions			
Input arcs	48	509	2,259
Output arcs	43	419	2,259
Input coordinates	48	106,812	258,456
Output coordinates	45	10,283	29,088
Polygons	49	156	642
Arc length (m)			
Hydrologic Units			
Input arcs	48	34	67
Output arcs	43	36	72
Input coordinates	50	45,205	82,480
Output coordinates	45	4,850	10,966
Polygons	54	12	23
Arc length (m)	34	8.17	12.28
Federal Land Ownership			
Input arcs	33	47	604
Output arcs	32	59	604
Input coordinates	35	36,536	181,250
Output coordinates	33	5,492	13,896
Polygons	37	31	222
Arc length (m)	24	5.65	25.19

points by millions of points as they are received from the digitizing contractor. The two data volumes are shown in table 4 as they are distributed by category over the 9-year period. Since the beginning of the digitizing contract in early 1975, over 10 million output coordinates and over 5,000 meters of arc lines have been processed in the editing stage. That volume is the equivalent of over 80 million bytes of data.

TABLE 4. — Estimated data volumes, 1975 - 1981

Category (1975-1981)	Arc line length (meters)	Output coordinates
Land Use and Land Cover	37,408	37,000,000
Political Units	3,002	2,000,000
Census County Subdivisions . .	7,708	6,250,000
Hydrologic Units	3,592	3,250,000
Federal Land Ownership	2,698	2,750,000
State Land Ownership	2,665	3,000,000
Total	57,073	104,250,000

DATA STRUCTURE

The handling of spatial data, particularly in large quantities, is an area which has not been explored very thoroughly. Researchers have typically dealt with either small quantities of complex data or large quantities of simple data. Under these circumstances, questions of data organization have not been critical, but as large sets of spatial data are compiled, the necessity of having an efficient, comprehensive data structure becomes apparent.

The problems of data organization that appear minor to a researcher working with what he views as a manageable data set become large obstacles to someone who is operating in a production mode. The economics of the system are such that they are very sensitive to forms of data organization. A well-thought-out and optimized logical and physical structuring of the data may mean the difference between the success and failure of a large project.

The GIRAS data structure is the latest in a series of evolving structures used to represent digital land use data. As such it reflects our (and our users') biases concerning the information to be presented and its format.

Since all maps compiled in the land use and land cover mapping and data program are polygon maps, the data structure was designed to expedite the handling of this data type. The topological elements associated with polygon maps are shown in figure 2. In the GIRAS structure, the common boundaries, or arcs, are digitized only once. The arcs are then linked together to form polygons. Arcs are line segments defined by a series of x, y coordinate pairs. Topologically the structure is similar to the DIME format used by the Bureau of the Census, (Cooke and Maxfield, 1967) and to the structure described by Puecker and Chrisman (1975). In

dealing with spatial data, it is often more desirable to store certain data rather than to recompute it. Thus, a large amount of ancillary data (e.g., polygon area, perimeter, etc.) is carried along in the data structure. The storage of such information not only eliminates redundant computations, but also facilitates data retrieval and analysis tasks. The general structure of a map file is shown in figure 3.

The map header contains a substantial amount of in-

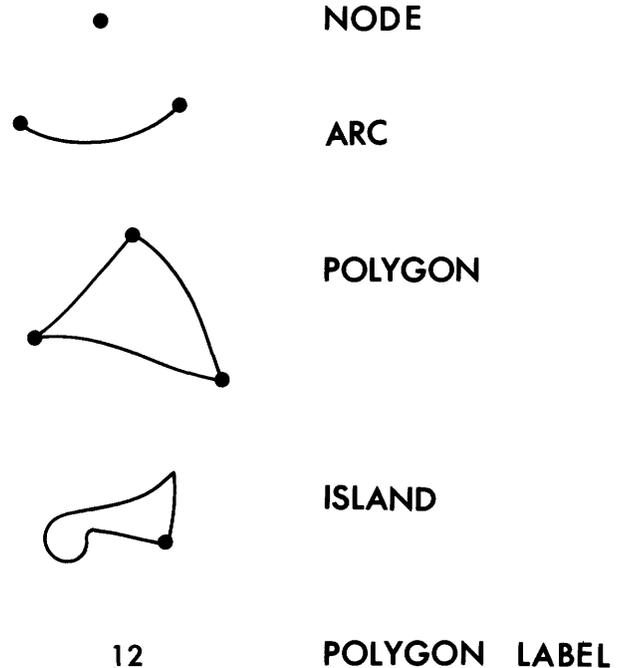
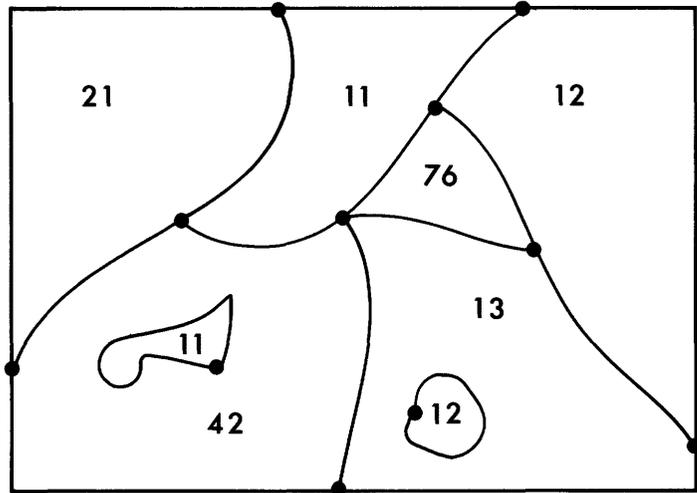


FIGURE 2. — Topological elements of a polygon map.

formation, including the amount of data in the map, the date of the source material, title, and control point information. The details of the contents of the map header are shown in figure 3. A map may be segmented into sections. This sectioning allows for more economical processing by constraining the amount of data processed at one time. It also facilitates geographic retrieval of subsets of the data. The section header is described in figure 3.

Within each map section there are four subfiles: arc records, coordinate data, polygon records, and a subfile assigning arcs to polygons (called FAP). Figure 3 shows the contents of the arc and polygon records. The subfiles are highly interrelated to facilitate analysis. The polygon record contains a pointer to the FAP subfile (PLA), which, in turn, points to those arcs comprising the polygon. The arc records point to coordinate strings describing that arc. Both the arc and polygon records contain information to further simplify analysis. For example, an arc record contains the attribute codes left and right of the arc. If an analysis requires the extraction of boundaries between two types of land use, then that information can be retrieved directly from the arc records. In addition, if the areas of the polygons adjacent to the boundary are desired, the arc record points to the appropriate polygon records where the areas are recorded.

The TEXT file relates to the entire map and contains definitions of attribute codes which can be used in manipulation procedures. The associated data file is provided to enable users to carry with the map additional information for their own purposes.

The design of the data structure was guided to some extent by the fact that the program is in a production mode for land use and land cover and associated maps. At present, map editing operates in a batch mode on sequential files. The editing process necessarily operates upon all the data, often employing insertions and deletions to the data. This operation is amenable to sequential processing. However, the data structure has been designed so that random access methods could be implemented. This would probably be the desired method when the maps are placed in a data base and the manipulations are focused upon retrieval and analysis rather than modification.

DATA CAPTURE

Once a particular data structure has been determined, a procedure must be established to capture data to place in that structure. The data capture procedure involves the conversion of line maps into a digital format. As defined here the digitization process includes not only the initial conversion to digital form, but also the editing process by which "clean" or logically correct data files are produced.

A number of alternatives are available in both hardware and methodology to accomplish the initial conversion to digital form. Our experience has highlighted two major considerations in the design of a data capture system. These are hardware independence and the minimization of digitizer-operator induced errors. An effective method must not rely heavily on either the data capture device or operator judgment. The input processing procedures, under development since 1973, attempt to accomplish this by limiting the amount of information required from the digitizer and allowing transportable software to perform automatic editing, error checking, and file creation.

The digital data from any given digitizer must be identified by map title, date, and other textual information. The digitizer coordinate frame of reference is defined by the map projection and the digitization of selected control points with known geographic coordinates. Figure 4 shows map header control points. In addition, the processing procedures require two files; the first contains unlabeled lines defined by a series of x, y coordinate pairs, and the second, a file of polygon labels, each tied to an arbitrary point within the polygon.

In digitizing, lines are not tagged in any way, and all that is required while digitizing lines is the ability to recognize line intersections (nodes). The lines of the map are completely defined when each arc, including the arcs which are the boundaries of the map, has been digitized at least once. The only points which need be digitized more than once are the nodes.

The arcs of the maps can be linked together at their

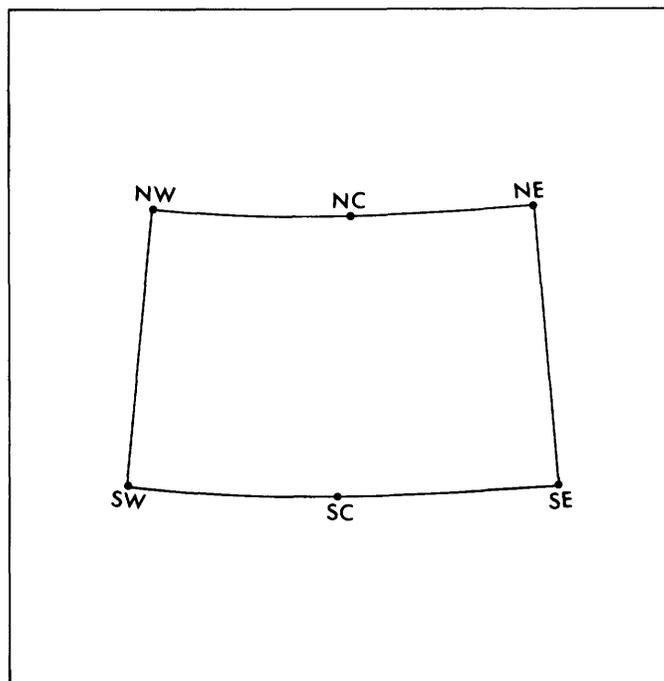


FIGURE 4. — Map header control points.

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

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

The types and sophistication levels of equipment which have been used to supply data to the processing procedures cover a wide range from simple manual encoding and card punch operations to an automatic line-following laser scanner. While any device which can translate lines into a series of x, y coordinate pairs may be used, until recently the device used has been some variety of digitizing table with a manually operated cursor.

Experimental projects involving map digitization were conducted in 1973 and involved the use of Bendix Datagrid¹ and Wang tables. Production digitizing of U.S. Geological Survey land use and land cover maps began in the fall of 1974 using a Computer Equipment Corporation digitizing station operated by the Johns Hopkins Applied Physics Laboratory. To support the editing, update, and limited revision of production land use and land cover data, and to support continuing research projects requiring map digitization, the program utilizes INTERMAP, an interactive digitizing and editing system. This system has allowed a large amount of flexibility in digitizing and editing tasks.

Our experience in processing data digitized on tables with manually operated cursors has shown us that the number of problems, and therefore the amount of editing necessary, has varied with the machine or digitizing station used and with the skill of the operator. Some machines, designed for automatic drafting work, have proved unsuitable due to the high inertia of the cursor or arm when attempting to depart from a straight line.

¹The use of brand names in this report is for identification purposes only and does not imply endorsement by the Geological Survey.

Many tables create problems merely because they do not permit the operator to view his work until after he has completed the digitizing operation, if at all. The problems caused by using such "blind" tables include missing data (either lines or polygon labels), lines digitized more than once, incorrect labels, and loss of origin resulting in different frames of reference for various parts of the same map. Digitizing systems which allow viewing of the data while digitizing and which include some degree of back-tracking capability produce better results.

The experience and skill of the digitizer operator play even more important roles in the quality of the output from manually operated digitizer tables. Map topology must be understood and care must be taken so that control points are digitized correctly and lines are followed exactly. Constant attention must be paid so that line intersections (nodes) are recognized and that two arcs are not digitized as one. Labels must be correct, and good bookkeeping kept so that all labels and lines are digitized.

In 1975, the next phase of production digitizing of U.S. Geological Survey land use and land cover maps was undertaken by contract to I/O Metrics Corporation, who used a SWEEPNIK laser scanner. That scanner is used in a line-following mode rather than a raster-scan mode so that no raster-to-vector conversion is necessary. The use of the laser scanner has solved many of the problems inherent in manual digitizing, but has brought up new problems to replace them. The most serious problems have come from the poor quality of the line data after scale reduction and photo processing of the original inked compilation maps. Weak lines on the original drop out and strong lines bleed into neighboring lines and labels. Graphics of this quality cause the operator to repeatedly reposition the laser probe manually during the course of a digitization run. However, our experience has shown that if the source maps are scribed, with digitizing by automatic scanner in mind, considerable savings may be realized. If the high quality graphics input standards are met, the digitized data returned by the scanner are far more accurate and error free than the data from a manual digitizing operation.

After the necessary data have been captured by the digitizing device, the data go through the following steps to produce a final clean file:

1. Conversion of data files to standard GIRAS format;
2. Data reduction, i.e., elimination of points not needed to define lines within a given spatial tolerance, (data reduction program);
3. Splitting of data sets into sections, if necessary;
4. Limited automatic edit and error detection for the arc data;

5. Manual batch or interactive edit of line data, with returns to step 4 until data are error free;
6. Merging of labeled polygon points with the arc data resulting in either further error detection or clean files (ATP program);
7. Manual batch or interactive edit of polygon label data with returns to step 5 or 6 if necessary; and
8. Edge match of each map section with neighboring map sections.

This process is diagramed in figure 5.

All new digitizer data are first converted to standard GIRAS format. Each digitizer requires a special format conversion program. This format conversion process can be combined with the data reduction step to reduce input/output charges.

The data reduction step is important since all further editing, manipulation, and retrieval costs are dependent upon the amount of data to be processed. Our consolidation procedure has proven to be computationally efficient, and provides an 80–90 percent reduction in the amount of data delivered by the SWEEPNIK scanner.

In the next step the digital data may be split into sections, depending on the amount of data involved. As part of the splitting process, arcs are added along split lines to form a continuous outside boundary for each section.

Next, an automatic editing procedure is applied to the arc data. The types of editing and error checking are:

A. Editing:

1. Deletion of one-point arcs,
2. Deletion of arcs shorter than allowed tolerance,
3. Deletion of duplicate arcs,
4. Adjustment of arc end points to meet exactly at nodes, and
5. Deletion of points within duplicate-point tolerance of nodes;

B. Error detection:

1. End point of an arc matches no other end points and
2. Less than three arcs meet at node
(Exception: one-arc islands)
(Inclusion: "apple in a window").

The map processors use the error listing generated by this procedure, a graphic display of the digital data (either a hard copy plot, or CRT display), and a copy of the compilation manuscript to check the data and to decide the necessary edit corrections. The digitizing errors which must be corrected are:

A. Line errors:

1. Two arcs represented as one (joined),
2. One arc represented as two,
3. Lines digitized more than once, and
4. Lines not digitized at all;

B. Polygon label errors:

1. Digitized point on an arc,
2. Digitized point outside the polygon,
3. Polygon mislabeled, and
4. Polygon left unlabeled.

The compilation errors which must be corrected are:

A. Line errors:

1. Line or portion of a line omitted,
2. Line drawn to an island, separating a polygon from itself ("apple in a window"), and
3. False boundary between polygons of the same characteristic;

B. Polygon label errors:

1. Polygon mislabeled and
2. Polygon left unlabeled.

The possible commands to correct these errors are:

A. Unlabeled arcs:

1. Join two arcs into one,
2. Divide one arc into two,
3. Delete an arc,
4. Delete a segment of an arc,
5. Add a segment of an arc,
6. Add an arc,
7. Translate an arc, and
8. Rotate and (or) stretch/compress an arc;

B. Labeled polygon points:

1. Change a label,
2. Add a labeled point,
3. Delete a labeled point, and
4. Move a point.

This editing process continues until no further node errors are found.

A major computer program of the input processing is the arc-to-polygon (ATP) program. ATP ties the labeled polygon interior points to the arc data, creating the cross referenced polygon-arc files described in the previous section. The program also serves as the final stage of topological checking and verification. The program sequence is outlined in figures 6–8. Most of the Polygon Label Errors are flagged in ATP as well as false arc crossings and duplicate arcs. The errors listed may indicate more arc editing and (or) editing of the labeled interior points. Since the program deletes duplicate and misplaced labels, the editing of labeled polygon points consists of either changing a label or adding a labeled point.

The final step in building the GIRAS file involves the edge matching of the outside boundary arcs of each map section with neighboring map sections either from the same map sheet, or other maps previously digitized and filed. This edge match process entails not only the completion of the map presently processed, but the update of the edge information of maps already stored.

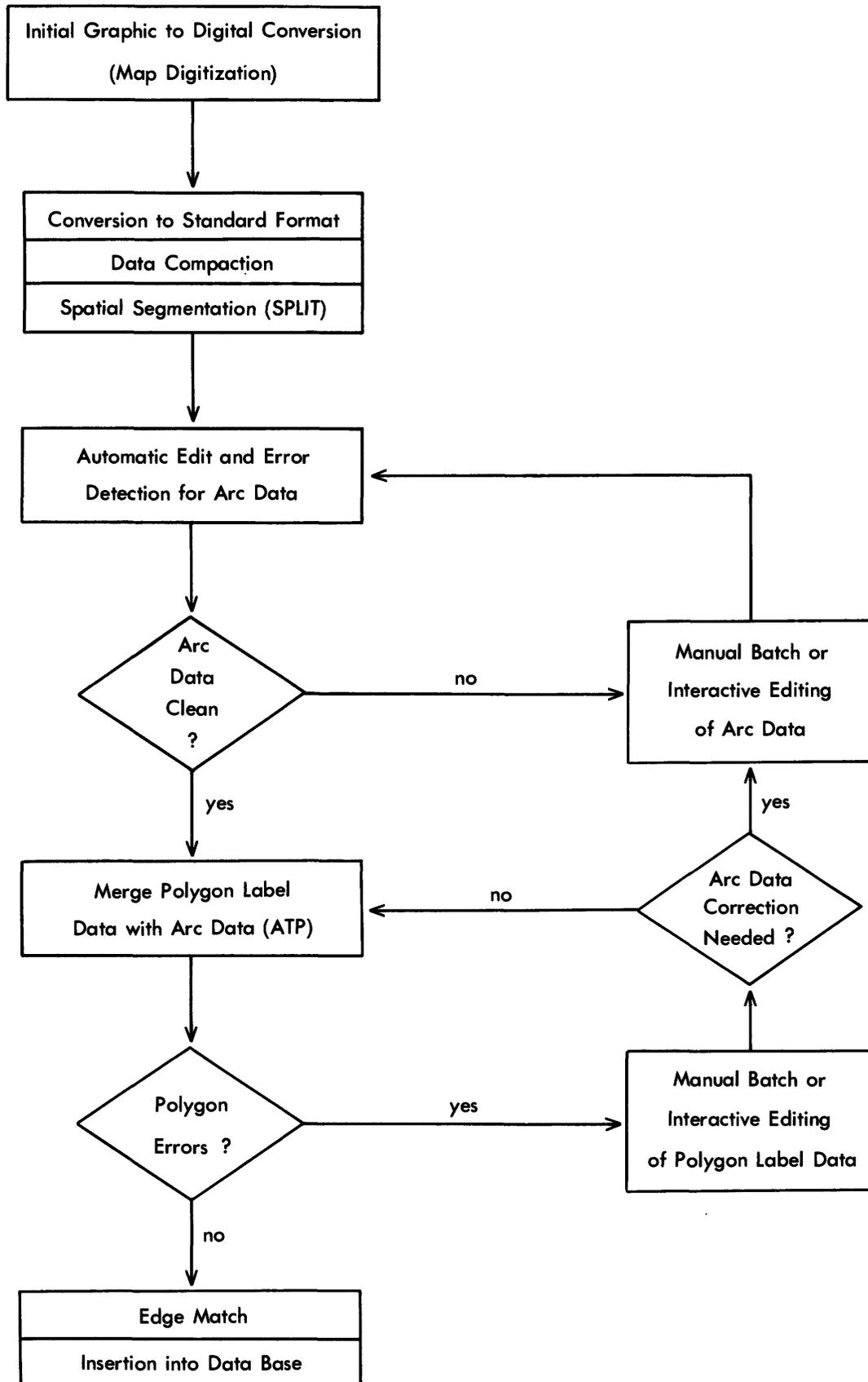


FIGURE 5. — Graphic input procedures for GIRAS.

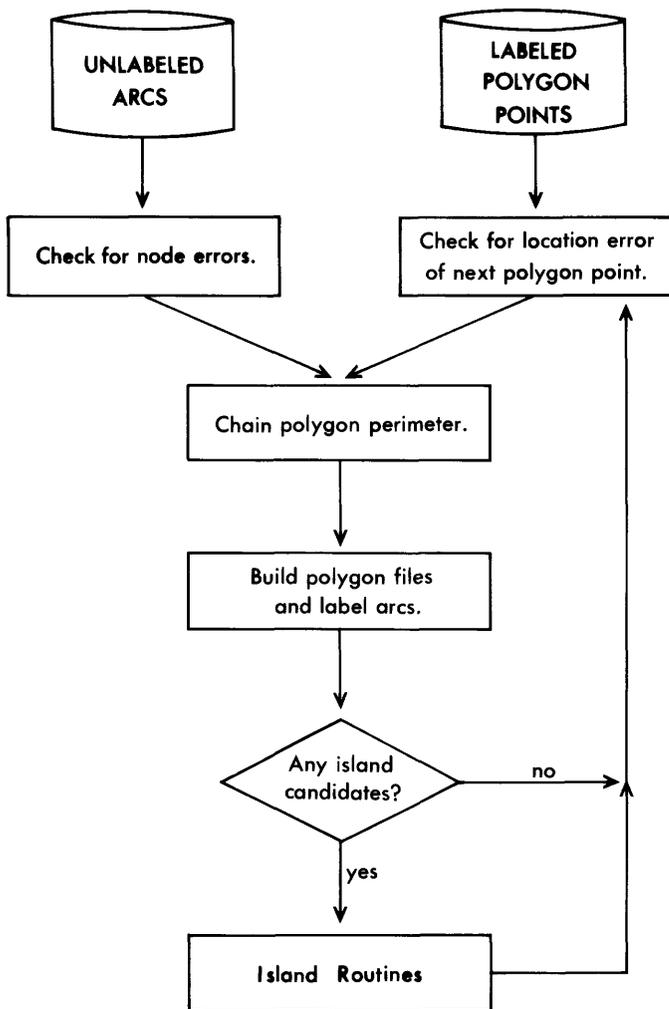


FIGURE 6. — Arc-to-polygon program sequence.

DATA RETRIEVAL

Once the source materials have been placed in GIRAS form, the data can then be processed by various retrieval and manipulation procedures. The manipulation procedures can be applied to the archived data or to subsets of that data after retrieval by a data base management system.

Until recently, complex, automated retrieval of digital spatial data has not been needed. Most manipulations have been conducted on entire map sheets or on sections of maps. The output software contains facilities for the selection of subsets of the data for manipulation. For example, a program which draws a color shaded map of polygons manipulates only the polygon with the desired attributes.

DATA MANIPULATION, ANALYSIS, AND OUTPUT

Spatial data in the GIRAS structure can be processed by a number of manipulation and analysis procedures such as:

1. Rotation, translation, and scaling of coordinates;
2. Conversion from geographic coordinates to specified map projections;
3. Restoration of original digitized map sheet from rectangular coordinate projection;
4. Conversion from polygon structure to grid cells of any specified size;
5. Production of area summary statistics from polygon or gridded data;
6. Geographic interpolation;
7. Filtering of nominal spatial data;
8. Feature generalization; and
9. Accuracy estimation of nominal maps.

DATA MANIPULATION

The first three procedures in the list deal with coordinate conversion. In order to accommodate the needs of various users, it is necessary to be able to rotate and translate the coordinate system and scale it to desired size. Similarly, a facility to change the data to a given map projection is also desirable, particularly when a supplemental base map exists on a different projection.

A number of existing spatial data systems can utilize only data stored by grid cells. Thus a program to convert the polygon structure to a grid cell format has been

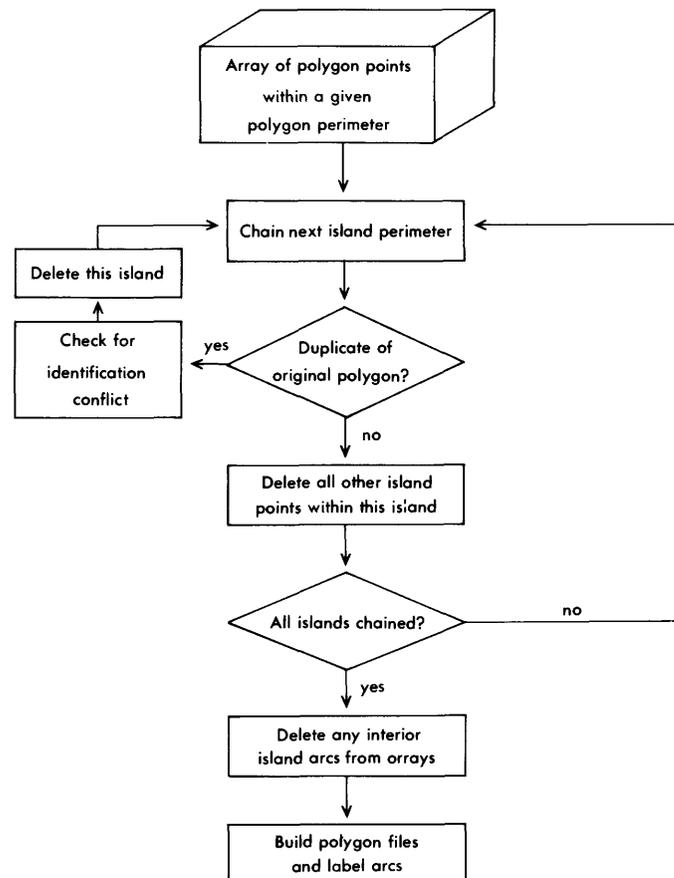


FIGURE 7. — Island routines within the arc-to-polygon program.

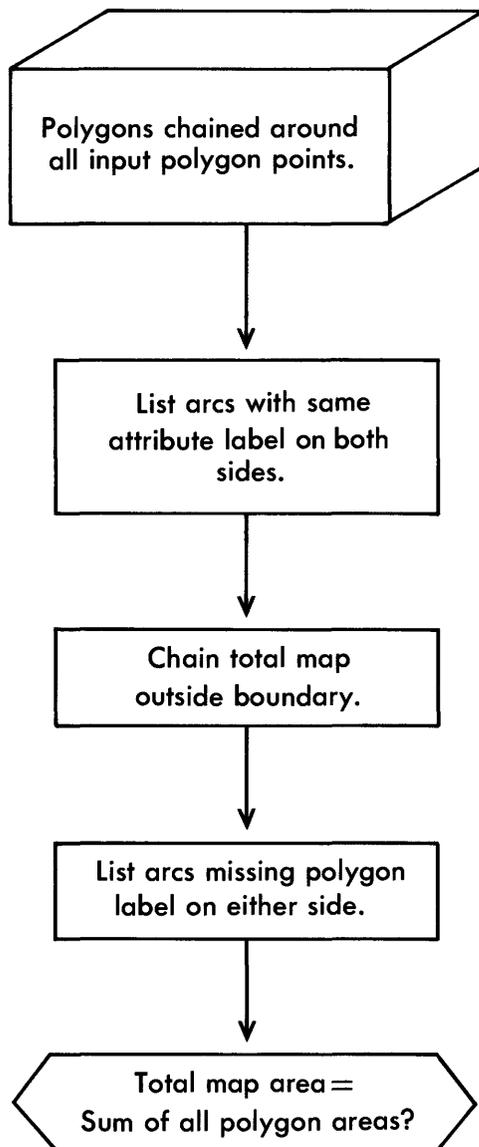


FIGURE 8. — Final arc-to-polygon error-checking sequence.

devised. The latest version of the program takes advantage of the information in the GIRAS data structure and greatly reduces computational time as compared to earlier techniques. Previous methods were based on a point-in-polygon search routine to see if the center of a grid cell was within a specified polygon. The technique used by the present program polygon-to-grid (PTG), uses only the data associated with the arcs, and not the polygons. The intersection points between each horizontal scan line of the grid system and the arc are calculated. The attribute data carried with each arc is then used to decide the categories of the cells along that scan line. This causes the relationship between grid size and the computation time to be linear rather than exponential. This is of importance since a small grid size (specifically one half of the smallest feature) is required to retain the detail of the map.

For gridded data sets, a method has been developed to evaluate that data set on the basis of the probability of occurrence of a certain cell state, given the states of the neighboring cells. This technique is described in more detail by Guptill (1977).

Programs exist which provide summary areal tabulations of the data, whether stored in polygon or grid form. State and local planners have expressed a great desire and need for such tabulations.

DATA ANALYSIS

When gaps occur in what one assumes to be a continuous field of scalar data, a geographic interpolation program is used to predict values for the missing data. In the most common case, a weighted average of the six nearest data points is used to supply an interpolated value. The GIRAS interpolation program uses this scheme, but it can accommodate other interpolation methods. The program is also capable of handling binary valued nominal data.

Filtering techniques for nominal data are a part of the GIRAS software. Several different methods have been developed for use with nominal data. These include such techniques as neighbor counting, modified neighbor counting, and various propagation processes (see Guptill, 1975). Additionally, an "optimal" spatial filter should also prove promising in the development of land use and land use change maps from Landsat Computer Compatible Tapes data. While data on a lattice are usually used in this program, polygonal data can be handled by using the results in this program, and restoring the polygons from the processed data.

Several procedures exist for feature generalization. These include the optimal filter algorithm mentioned previously as well as a modified version of the WEED program. These procedures indicate that a more suitable technique for feature generalization as required by scale changes in mapped data is needed.

DATA OUTPUT

The various display capabilities of the GIRAS software are:

1. Color/pattern shading of polygon;
2. Boundary and attribute plotting for polygons and arcs;
3. Pattern symbolization of grid cells;
4. Choropleth mapping of polygons;
5. Choropleth mapping of gridded polygons (line printer);
6. Isometric bivariate histogram plotting;
7. Perspective view block diagram;
8. Perspective view contour mapping; and
9. Perspective view pins diagram.

Some of the plotting criteria and options are:

Selective boundary plotting criteria:

1. Attributes on either side (by polygon),
2. Attributes not on either side,
3. Attributes on one side and another set of attributes on the other side (by boundary type),
4. Attributes on one side and another set of attributes not on the other side, and
5. Attributes not on one side and attributes not on the other side.

Polygon plotting options:

1. Boundaries only of given polygon attribute(s),
2. Polygon labels and (or) polygon sequence numbers,
3. Polygons selectively shaded, each selected attribute shaded with a unique user-specified color-pattern combination, and
4. Any scale.

Hard copy graphics output is presently (1977) generated by a Calcomp 763 drum plotter, although other output devices include the Gerber plotter and line printer. Most of the graphics procedures feature either a simultaneous generation of the drawing on the CRT screen, or allow for previewing of the drawing before actual plotting.

A procedure also exists to create open window press plate negatives using scribe coats generated on the Gerber 32 plotter from the GIRAS digital data. Solid colors are used to portray Level I land use and land cover and two digits carry the Level II land use and land cover data.

Eight separate plot files are generated on tape for the Gerber plotter, one file for each of seven Level I categories, and a file of all Level II land use boundaries. For each Level I category, only external boundaries (those separating that Level I category polygons from other Level I uses) and not internal boundaries (those separating different Level II sub-categories of the same Level I use) were selected for plotting. A 250 μm scribing tool is used to cut the set of scribe coat material.

Photosensitive peel coats are exposed from each Level I category plate for each set, and open window negatives are made for each color. The Level II land use boundary scribe coats are used to print the black land use boundaries. A screened standard 1:250,000 quadrangle black line base is added for reference.

GIRAS II, CURRENT SYSTEM DESIGN

The initial system of GIRAS I, now operating in a production mode, was based upon a batch sequential input processing subsystem leading to the storage of data in simple digital archives. Further, the manipulation of those archival files is accomplished by a grid compositing system to permit the production of rudimentary,

standardized statistical and graphic outputs for the early users of the program. Thus far, that system has been successfully utilized to input a large amount of land use data and associated environmental data. In addition, the system has been supporting an accelerating production of tabular and graphic data sets in response to requirements of State cooperative users.

The GIRAS I production system, however, was not designed to support the capture, editing, and correction of very large and complex data sets such as presently scheduled for the land use and land cover data and analysis program, nor to fulfill the extended and advanced manipulative or analytic desires of the cooperative users. So far, expectations of State cooperative and Federal agency users have rapidly exceeded the relatively modest capabilities that were originally planned for a batch-oriented production system, and the need for a more advanced and sophisticated geographic information system was recognized early in GIRAS operations.

In mid-1975, the past experiences in the operation of GIRAS I were examined, the requirements of State cooperative agencies and Federal agencies were considered, and studies by outside consultants on the computer handling of geographical data (Tomlinson, Calkins, and Marble, 1976) were reviewed. This reexamination resulted in the development of a radically different design philosophy and criteria which have guided the development of GIRAS II through 1977.

GIRAS II is characterized by its use of an interactive, on-line, time-sharing, random access input processing system, a complex data base to be managed by a data base management system, interactive manipulation and analysis facilities which will permit user intervention through remote terminals, and advanced statistical and graphic outputs which include interactive graphic displays for remote terminal users.

The major design features of GIRAS I and II for the five essential elements of a geographic information system are summarized in figure 9 to contrast the capabilities of the two systems in terms of the ways in which a user of the system might operate it. The five elements are: (1) Data capture, (2) input processing, (3) data base management, (4) manipulative and analytic operation, and (5) statistical and graphic output. The two columns in the center of the diagram indicate the increased flexibility and value of the system under development as compared with the system now in operation.

The general system structure of GIRAS II and the transition linkage of GIRAS I with GIRAS II is shown in figure 10. Although the two systems differ drastically in complexity, explanation of a few differences in their design will serve to contrast the capabilities for applica-

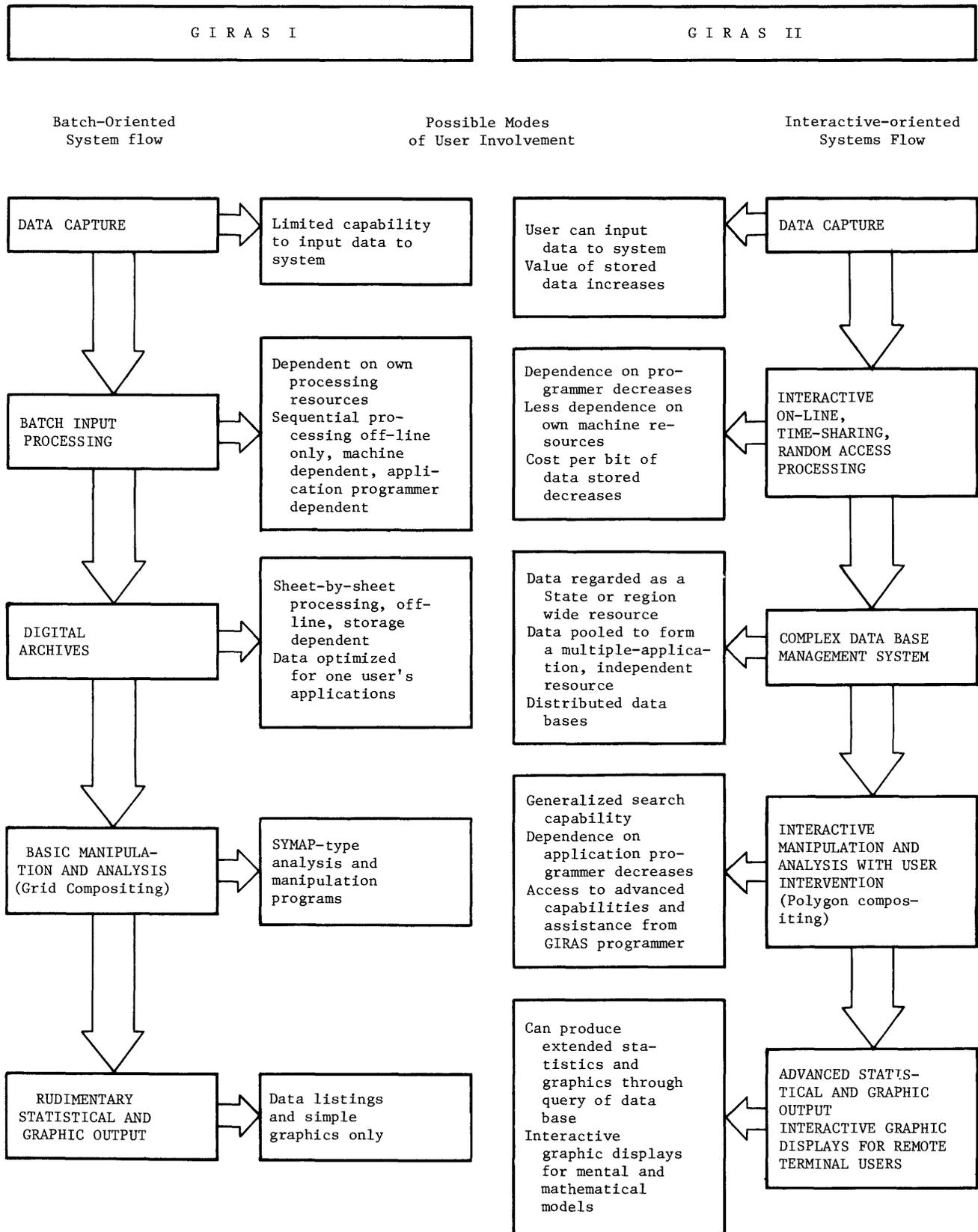


FIGURE 9. — Design features and contrasting capabilities of GIRAS I and GIRAS II.

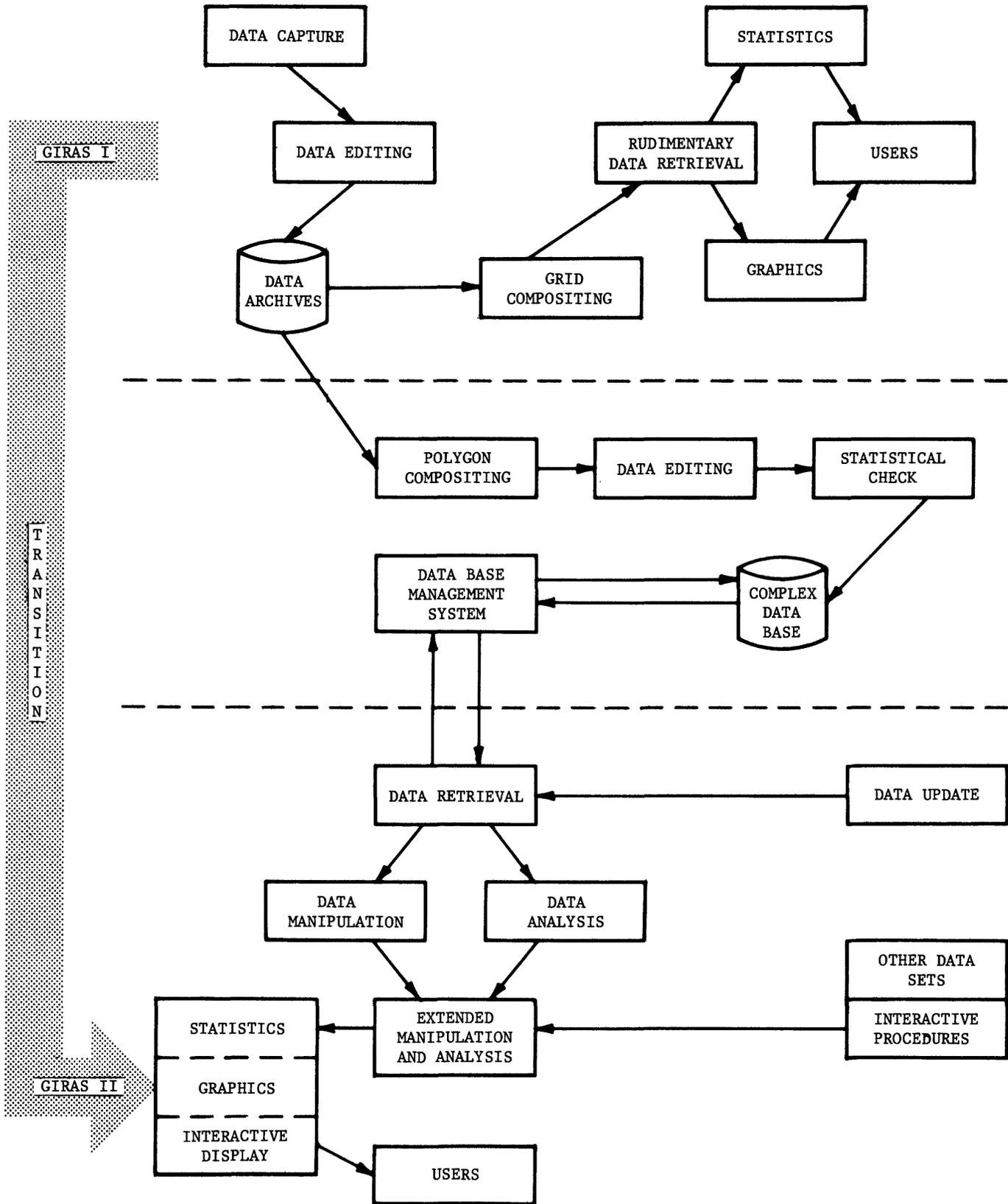


FIGURE 10. — System flow diagram of GIRAS I and GIRAS II.

tions in land use planning and decisionmaking. Probably the most critical change has been from a batch-oriented, sequential access computer system to an interactive, time-sharing, random-access computer system. A second design change is the shift from grid compositing to polygon compositing. Tests of polygon compositing have demonstrated that the user can move from the manipulation of only entire map sheets or sections of sheets to output software which will provide facilities for the selection of subsets of the data for manipulation, the query of the data base, or searches for particular data subsets through the hierarchical data structure. A third aspect of the GIRAS II design is the interposition of a "data base management system" between the complex data base and data retrieval. The need for a data base management system has become critical as increasing volumes of spatial digital information are added to the U.S. Geological Survey geographic information system.

REFERENCES

- Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E., 1976, A land use and land cover classification system for use with remote sensor data: U.S. Geol. Survey Prof. Paper 964, 28 p.
- Cooke, D. F., and Maxfield, W. H., 1968, The development of a geographic base file and its uses for mapping, *in* Papers from the fifth annual conference of the Urban and Regional Information Systems Association [1967]: Akron, Ohio, Kent State Univ., p. 207-219.
- Guptill, S. C., 1975, Spatial filtering of nominal data — an exploration: Univ. Michigan Ph. D. thesis, 210 p.
- 1977, An optimal filter for maps showing nominal data: U.S. Geol. Survey Jour. Research, v. 6. (In press.)
- Puecker, T. K., and Chrisman, Nicholas, 1975, Cartographic data structures: *Am. Cartographer*, v. 2, no. 1, p. 55-69.
- Tomlinson, R. F., Calkins, H. W., and Marble, D. F., 1976, Computer handling of geographical data: Paris, Nat. Resources Research, ser. XIII, 214 p.
- U.S. Bureau of the Census, 1972a, Census tracts — 1970 census of population and housing: U.S. Bur. Census Final Repts. PHC(1)-1 through PHC(1)-241. [1 v. for each Standard Metropolitan Statistical Area.]
- 1972b, County and city data book, 1972; U.S. Bur. Census, 1020 p.
- 1972c, Geographic identification code scheme — 1970 census of population and housing: U.S. Bur. Census Repts. PHC(R)-1 through PHC(R)-4. [1 v. for each of the 4 U.S. regions, Northeast, North Central, South, and West.]