National Mapping Program

USGS Digital Cartographic Data Standards

Land Use and Land Cover Digital Data

Public Inquiries Office
U.S. Geological Survey
504 Custom House
555 Battery Street
San Francisco, California 94111

Geological Survey Circular 895-E

US GeoData

USGS DIGITAL CARTOGRAPHIC DATA STANDARDS

USGS Circular 895-A: Overview and USGS Activities
B: Digital Elevation Models
C: Digital Line Graphs from 1:24,000-Scale Maps
D: Digital Line Graphs from 1:2,000,000-Scale Maps
E: Land Use and Land Cover Digital Data
F: Geographic Names Information System
G: Digital Line Graph Attribute Coding Standards

Questions regarding availability and ordering of US GeoData (all types of digital cartographic and geographic data produced and distributed by the U.S. Geological Survey) should be addressed to:

User Services Section
National Cartographic Information Center
U.S. Geological Survey
507 National Center
Reston, Virginia 22092

Technical questions and comments should be addressed to:

Chief, National Mapping Division
U.S. Geological Survey
516 National Center
Reston, Virginia 22092

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

LAND USE AND LAND COVER DIGITAL DATA

By Robin G. Fegeas, Robert W. Claire,
Stephen C. Guptill, K. Eric Anderson, and Cheryl A. Hallam

USGS Digital Cartographic Data Standards

Edited by Robert B. McEwen, Richard E. Witmer, and Benjamin S. Ramey

---

Geological Survey Circular 895-E

1983

United States Department of the Interior
JAMES G. WATT, *Secretary*

Geological Survey
Dallas L. Peck, *Director*

Library of Congress catalog-card No. 83-600581

Free on application to Distribution Branch, Text Products Section,
U. S. Geological Survey, 604 South Pickett Street, Alexandria, VA 22304

FOREWORD

In recent years, the disciplines of cartography and geography have undergone a rapid and striking reorientation as the techniques for digital collection and manipulation of data have evolved from fledgling laboratory procedures into dominant and driving forces that now pervade the disciplines. Digital techniques have provided a variety of new and powerful capabilities to collect, manipulate, analyze, and display spatial data. However, this evolution also has introduced a number of new and complex problems. One of the most pressing problems, and one which is receiving particular attention at present, is the issue of digital cartographic data standards.

The U.S. Geological Survey (USGS) has been actively developing digital cartographic and geographic techniques for over a decade and has taken significant steps to develop and define in-house standards governing the various types of digital cartographic data that are being collected and archived in a national digital cartographic data base. The in-house standards are expressed in the form of specifications documents that were prepared to govern collection of the data and in the form of user guides that were prepared for distribution with the data.

In an effort to fulfill lead agency requirements for promulgation of Federal standards in the earth sciences, the documents have been assembled with explanatory text into this USGS Circular consisting of separately bound chapters. This Circular describes some of the pertinent issues relating to digital cartographic data standards, documents the digital cartographic data standards currently in use within the USGS National Mapping Division, and details USGS efforts to define national digital cartographic data standards.

Chapter A is an overview in which the major issues involved in developing digital cartographic data standards are discussed and the activities of the USGS related to digital cartographic data production and standards development are described in detail. Succeeding chapters comprise the pertinent documents that establish USGS in-house standards for the various types of digital cartographic data currently produced by the National Mapping Division--that is, digital elevation data, digital planimetric data, digital land use and land cover data, and digital geographic names data.

This compendium of relevant material is prepared to serve as a benchmark and to assist ongoing efforts to establish acceptable standards and conventions for both Federal agencies and the public.

Dallas L. Peck
Director,
U.S. Geological Survey
Department of the Interior

R. B. Southard
Chief,
National Mapping Division
U.S. Geological Survey

PREFACE

This Circular is the result of the efforts of numerous individuals who have contributed to the research, development, and preparation of various digital cartographic and geographic standards for the National Mapping Division of the U.S. Geological Survey. The individuals named as chapter authors represent both the originators of the various concepts as well as the writers who expanded and clarified these ideas. Their contributions, either to the concepts or the writing, are of such magnitude as to warrant crediting as authors.

Atef A. Elassal was largely responsible for the original data structures and computer file formats that are used for the Digital Line Graphs and Digital Elevation Models. The attribute coding scheme was first developed by members of the Digital Applications Team under the direction of Robert B. McEwen. The Geographic Names Information System was conceived and developed by Sam Stulberg and Roger L. Payne. The Geographic Information Retrieval and Analysis System was developed by Robin G. Fegeas, K. Eric Anderson, Stephen C. Guptill, Cheryl A. Hallam, and William B. Mitchell. The small-scale Digital Line Graph data structure and attribute coding scheme was developed by Warren E. Schmidt and Michael A. Domaratz.

The Circular was compiled in part from various user guides and technical instructions of the National Mapping Division. These documents were originally prepared by several individuals; credit is acknowledged to G. Michael Callahan, A. Joan Szeide, William R. Allder, Vincent M. Caruso, Hugh W. Calkins, Donna Cedar-Southworth, and Cheryl A. Hallam. The compilation of the various guides, instructions, and other material into the Circular format was performed with major assistance by Clark H. Cramer, Eloise R. Byrd, and Cynthia L. Cunningham.

We acknowledge these substantial contributions that have led to this publication.

CONTENTS

	Page
Foreword	iii
Preface	iv
Abstract	1
Introduction	1
Source data characteristics	2
Land use and land cover map	2
Political unit map	3
Census county subdivision map	3
Hydrologic unit map	4
Federal land ownership	4
State land ownership map	4
GIRAS data structure elements	4
Creation of the GIRAS files	5
Applications	6
Manipulation and analysis	6
Display	7
Data volumes	7
GIRAS file format	8
The standard local UTM coordinate frame of reference	9
Map header	9
Section header	10
Arc records subfile	11
Coordinate subfile	11
Polygon records subfile	12
FAP subfile	13
Text subfile	14
Associated data subfile	16
Composite theme grid (CTG) data file format	16
Binary CTG data file format	16
Binary CTG map header	17
Character composite theme grid (CTG) file format	18
Standard character-formatted CTG data file	19
Character CTG map header	19
References	20
Appendix. Listing of CTG map header data	21

ILLUSTRATIONS

		Page
Figure 1.	Topological elements of a polygon map	5
2.	GIRAS file structure	8
3.	GIRAS map header structure	10
4.	GIRAS section header structure	10
5.	GIRAS arc records subfile	11
6.	GIRAS coordinate subfile	11

Page
7. GIRAS polygon records subfile 12
8. GIRAS FAP subfile 13
9. GIRAS FAP subfile creation 14
10. GIRAS text record subfile 15

TABLES

Page
Table 1. U.S. Geological Survey land use and land cover classification system for use with remote sensor data 3
2. Federal land ownership 4
3. Sample GIRAS data volumes from 1:250,000-scale map overlays 7
4. Map type codes for GIRAS data base 9
5. Sample text subfile for a Census County Subdivision GIRAS file 16

USGS Digital Cartographic Data Standards

LAND USE AND LAND COVER DIGITAL DATA

By Robin G. Fegeas, Robert W. Claire,
Stephen C. Guptill, K. Eric Anderson, and Cheryl A. Hallam

Abstract

The discipline of cartography is undergoing a number of profound changes that center on the emerging influence of digital manipulation and analysis of data for the preparation of cartographic materials and for use in geographic information systems. Operational requirements have led to the development by the USGS National Mapping Division of several documents that establish in-house digital cartographic standards.

In an effort to fulfill lead agency requirements for promulgation of Federal standards in the earth sciences, the documents have been edited and assembled with explanatory text into a USGS Circular. This Circular describes some of the pertinent issues relative to digital cartographic data standards, documents the digital cartographic data standards currently in use within the USGS, and details the efforts of the USGS related to the definition of national digital cartographic data standards. It consists of several chapters; the first is a general overview, and each succeeding chapter is made up from documents that establish in-house standards for one of the various types of digital cartographic data currently produced. This chapter 895-E, describes the Geographic Information Retrieval and Analysis System that is used in conjunction with the USGS land use and land cover classification system to encode, edit, manipuate, and analyze land use and land cover digital data.

INTRODUCTION

The U.S. Geological Survey (USGS) is currently producing land use and land cover maps and associated overlays for the entire 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.

These maps will help satisfy a long-standing need for a consistent level of detail, standardization of categories, and consistent use of scales of compilation for a type of data used by government land use planners, land managers, and resource-management planners. Once this benchmark series of maps is completed, updating of the maps will provide a much-needed tool for analyzing trends, problems in local and regional areas throughout the Nation, and changes in land use patterns.

The set of land use and land cover and associated maps consists of land use and land cover, political units, hydrologic units, census county subdivisions, Federal land ownership, and State land ownership (optional).

The land use and land cover map is compiled to portray the Level II categories of the land use and land cover classification system documented by Anderson and others (1976). The Level II categories of this land use and land cover classification system provide the user with a basic framework to which third- and fourth-level categories may be added.

The associated maps portray either natural or administrative information and provide the user with the opportunity to utilize the land use and land cover maps and data either individually or collectively to produce graphic or statistical data for the areas portrayed on the associated maps. This mapping system is constructed in such a way that the graphical and statistical land use and land cover data can be related to other resource fields such as soils, geology, hydrology, and demography.

To provide the data in digital form, the Geographic Information Retrieval and Analysis System (GIRAS) has been developed (Mitchell and others, 1977). The data structure used in GIRAS to store the information is the end result of a series of evolving structures and, as such, reflects the judgment by the USGS concerning the presentation and format of polygonal data. For those users better able to handle data in a grid cell form, data are also provided in a Composite Theme Grid (CTG) format.

SOURCE DATA CHARACTERISTICS

The characteristics of the digital cartographic data base for land use and land cover and associated maps reflect the parameters used in compiling the maps. The land use and land cover mapping program is designed so that standard topographic maps at a scale of 1:250,000 can be used as a base for compilation and reproduction. After the start of the mapping program, the USGS began compiling intermediate-scale topographic maps at a scale of 1:100,000. The USGS anticipates preparing land use and land cover and associated maps at a scale of 1:100,000 and will release them to the open-file system when 1:100,000-scale topographic map bases become available.

The 1:250,000-scale mapping format is generally a quadrangle unit of 1° of latitude x 2° of longitude. The 1:100,000-scale mapping format has been established as a 30' x 60' quadrangle, normally a quarter of a 1:250,000-scale quadrangle. Both series use the Universal Transverse Mercator projection.

LAND USE AND LAND COVER MAP

The basic purpose of this map is to provide land use and land cover data to be used either as data sources by themselves or in combination with the other data sets produced in the program. One of the basic sources of land use compilation data is the NASA high-altitude U-2/RB-57 aerial photo coverage, 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 maps 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 system by the USGS. Although compilation of land use and land cover data is performed on a film-positive base enlarged to a scale of approximately 1:125,000, the associated overlays are both compiled and digitized at a scale of 1:250,000.

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

All features are delineated by curved or straight lines that depict the actual boundaries of the areas (polygons) being described. The minimum size of polygons depicting all Urban or Built-up Land (categories 11-17), Water (51-54), Confined Feeding Operations (23), Other Agricultural Land (24), and Strip Mines, Quarries, and Gravel Pits (75) is 4 hectare (ha). All other categories of land use and land cover have a minimum polygon size of 16 ha. Those sizes also are considered the minimum sizes to which polygons are digitized. In the Urban or Built-up Land and Water categories, the minimum width of a feature to be shown is 200 m; that is, if a square with sides 200 m in length is delineated, the area will be 4 ha. Although the minimum-width consideration precludes the delineation of very narrow and very long 4-ha polygons, triangles or other polygons are acceptable if the base of the triangle or minimum width of the polygon is 200 m in length and if the area of the polygon is 4 ha. For categories other than Urban or Built-up Land and Water, the 16-ha

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-Brushland 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 Rock
		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
		84	Wet Tundra
		85	Mixed Tundra
9	Perennial Snow or Ice	91	Perennial Snowfields
		92	Glaciers

minimum size for delineation requires a minimum width polygon of 400 m. 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 scales of either 1:250,000 or 1:100,000. Source material for the political unit maps is from the Bureau of the Census unpublished maps entitled "County Subdivisions--Townships and Places" and from the Geographic Identification Code Scheme (U.S. Bureau of the Census, 1972a) and the County and City Data Book (U.S. Bureau of the Census, 1972b). The "County Subdivision--Townships and Places" maps are used also 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 nontracted "independent cities" in Maryland, Missouri, Nevada, and Virginia are given an eight-digit code reflecting the State and city codes.

CENSUS COUNTY SUBDIVISION MAP

The census county subdivision map depicts census tracts in Standard Metropolitan Statistical Area (SMSA) counties and the minor civil division or equivalent division in non-SMSA counties. This overlay map is digitized in the same manner as the other overlays. The map is based on data from a Bureau of the Census publication entitled Census Tracts that describes the census tracts of the SMSA-mapped counties (U.S. Bureau of the Census, 1972a) and on a Bureau of the Census publication entitled Geographic Identification Code Scheme (U.S. Bureau of the Census, 1972c). Each census tract is encoded with a four-digit number that 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 is based on the Hydrologic Unit Maps published by the U.S. Geological Survey Office of Water Data Coordination, 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 that 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 USGS has the responsibility for researching, obtaining, and formatting maps, plots, and other descriptive data related to Federal land ownership. Minimum size for the delineation is 16 ha. Ownership is encoded according to the agencies listed in table 2.

STATE LAND OWNERSHIP MAP

In instances in which the USGS 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 the State. Although this overlay is compiled to the same map base used for the other overlays, the polygons are encoded according to the referencing system used by the State.

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 Recreation Area)
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

GIRAS DATA STRUCTURE ELEMENTS

The GIRAS digital data structure was designed to handle large quantities of map data of the polygon type. The topological elements associated with polygon maps are shown in figure 1. A polygon is an area that is homogeneous in the characteristic being mapped. An arc describes a boundary either between two polygons or between a polygon and the outside of the map. Further defined, an arc begins at one node, or point common to three or more arcs (that is, an intersection), and ends at another node but does not pass through any node. A special case is a simple island polygon totally surrounded by a larger polygon. For purposes of digitizing, an arbitrary point on the boundary of the island is chosen as the beginning and ending node of the arc. Each polygon label on the map is a code, not necessarily unique, that identifies or describes the polygon in which it is placed or to which it points.

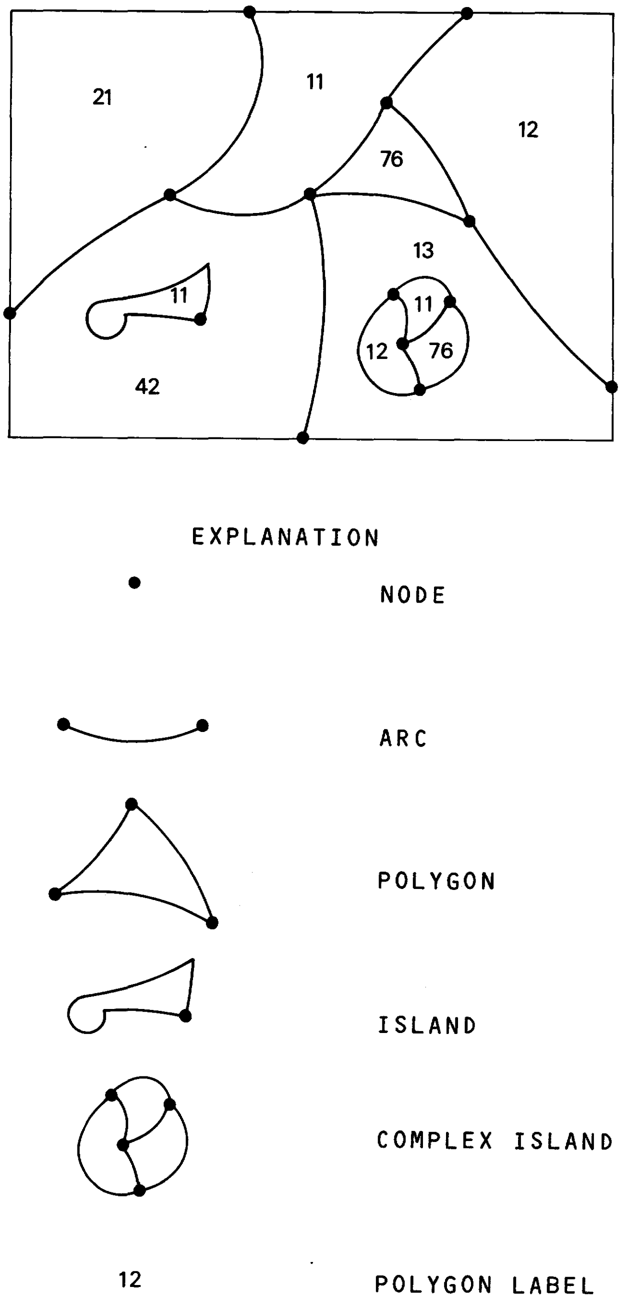


Figure 1.--Topological elements of a polygon map.

In the GIRAS structure, the common boundaries, or arcs, are digitized only once. The arcs are then linked together by editing software to form polygons.

Within the GIRAS data structure, the basic topological elements of a polygon map (arcs, nodes, and polygons) are all uniquely identified and cross-referenced to one another. The spatial location of an arc is given by a string of x,y points; the first point is its beginning (from) node, and the last point is its ending (to) node. The sequence of points of an arc defines a direction that, since the arc separates two polygons, determines a polygon to the right and a polygon to the left. A single node is the endpoint (first and (or) last) of three or more arcs (or the first and last endpoint of a one-arc island). A given polygon is spatially defined as the sequence of arcs that constitute its boundary, both external perimeter and any internal islands.

CREATION OF THE GIRAS FILES

The data capture procedure involves the conversion of the source material into a digital format. As defined by GIRAS, 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.

In digitizing, lines are not tagged in any way, and all that is required is the capability to recognize line intersections (nodes) during computer processing. Since land use and land cover maps consist entirely of polygons, the map is completely defined when each arc, including the arcs that are the boundaries of the map, has been digitized. Along with the line data, a file that contains at least one attribute and a point inside the polygon (represented by an x,y coordinate pair) for each polygon of the map must be entered into the system. These data are not in the GIRAS format at the end of the digitizing. Data are converted to the GIRAS format only when editing is finished.

After the necessary data have been captured, the data go through the following steps to produce a GIRAS file:

1. Conversion of the data file to the standard editing format;

2. Data reduction, by elimination of points not needed to define lines within a specified tolerance;
3. Splitting of data sets into sections, if necessary;
4. Limited automatic editing and error detection for the arc data;
5. Manual batch editing 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;
7. Manual batch editing of polygon labels data with returns to steps 5, 6, or 7, if necessary;
8. Edge matching of each section with neighboring sections and of each map with every available adjoining map;
9. Conversion of data files to standard GIRAS format; and
10. Transformation of data from table coordinates to the UTM coordinate system; where one coordinate unit equals 10 m.

APPLICATIONS

MANIPULATION AND ANALYSIS

Spatial data in the GIRAS format can be applied to individual problems through manipulation and analysis procedures such as:

1. Rotation, translation, and scaling of the coordinates;
2. Conversion to geographic coordinates and from geographic coordinates to specified map projections;
3. Conversion from arc-segment polygon structure to grid cells of a specified size;
4. Conversion to standard polygon format;
5. Production of area summary statistics from polygon or grid cell data;
6. Production of border (perimeter) and adjacency lengths of particular polygon types from the polygon data; and
7. Selection of a portion of a map for closer consideration by using procedures 1 through 6.

The first two procedures in the list deal with coordinate conversion. To apply the data to needs of various users, it is often necessary to be able to rotate and translate the coordinate system and scale it to the desired size. Similarly, a facility to transform the data to another map projection is desirable, particularly when supplemental data exist on a different projection.

A number of existing data systems utilize data stored in grid cells. Thus, the ability to convert the polygon structure to a grid-cell format can be very useful.

The land use and land cover and associated map data files were converted to grid cells of a specific size and orientation to permit their addition to a existing data base. This permitted GIRAS data to be used where the GIRAS data structure was inappropriate. GIRAS's capability to produce data in both polygon and grid-cell formats provides a flexibility whereby the needs of more users can be met.

Area summaries for a GIRAS data set may be obtained directly from the file. However, if a further breakdown of the information (for example, land use within each political unit for a data set) is wanted, it may be derived more easily from grid-cell formatted information than from an arc-segment file. Once all associated maps as well as the land use and land cover map for a 1:250,000-scale quadrangle sheet are in GIRAS format, they may be converted to one file of grid-cell information (an explanation of the CTG data file format begins on page 16) from which the more detailed summary described above may be obtained. For example, with the grid-cell (CTG) file, summaries of land use and land cover by county within a given drainage basin can be produced.

As with the grid-cell format data, the GIRAS format data facilitate production of certain types of information. An environmental study might find it useful to know the length of the border between two noncompatible types of land use or land cover (procedure 6). For example, if an area of industrial land borders a lake under study, the amount of lakeshore occupied by the industrial site might be of interest. This type of information can be derived easily from the arc-segment GIRAS files.

While GIRAS stores polygons as the arcs of which they are composed, many information systems that deal with polygon data store those polygons as complete entities. To bridge the gap between the two formats, conversion to a standard polygon format is necessary. Although this requires more storage than GIRAS's format, it allows the use of a simpler set of software for plotting, perimeter calculation, and area calculation.

Another technique for extracting information from the GIRAS files is the selection of an area smaller than the standard 1° x 2° map data file (procedure 7). It is often helpful to select data from the files for closer consideration. This windowing process reduces the amount of information handled by eliminating the portions of the map that are of no interest to the study. Once this smaller portion has been selected, any of procedures 1 through 6 may be applied.

DISPLAY

Computer-generated graphics may be used to augment the manipulations and analyses described previously. A computer-generated shaded color plot of the Level II land use and land cover for a data set along with a summary of the land use gives the investigator a spatial perspective of the distribution of land use and land cover over the area mapped. If it would be more helpful to see only one or more of the land uses displayed, they can be specially selected. If the lengths of borders between two specific land uses or land covers are under study, a plot showing only those two land uses or land covers would be helpful.

The detailed summary of land use and land cover by county might be illustrated by placing the outline of the county over the land use and land cover map. The pattern of land use and land cover for that county can be seen with the areas of the different land uses computed and displayed.

DATA VOLUMES

Large sets of complex spatial data, such as those handled by GIRAS, necessitate an efficient data structure. Table 3 shows some measure of GIRAS data volumes derived from records of data editing procedures. The numbers of coordinates (two coordinates, x and y, per point) are those defining the arcs and reflect the

Table 3.--Sample GIRAS data volumes from 1:250,000-scale map overlays

	Sample size (maps evaluated)	Average	High
Land Use and Land Cover			
Coordinates (no.)	54	165,384	435,540
Arcs (no.)	54	7,340	27,339
Polygons (no.)	51	2,986	9,865
Arc length (m)	50	60	139
Political Units			
Coordinates (no.)	47	3,526	7,834
Arcs (no.)	47	42	101
Polygons (no.)	54	16	37
Arc length (m)	34	9	13
Census County Subdivisions			
Coordinates (no.)	45	10,283	29,088
Arcs (no.)	43	419	2,259
Polygons (no.)	49	156	642
Arc length (m)	33	12	21
Hydrologic Units			
Coordinates (no.)	45	4,850	10,966
Arcs (no.)	43	36	72
Polygons (no.)	54	12	23
Arc length (m)	34	5	8
Federal Land Ownership			
Coordinates (no.)	33	5,492	13,896
Arcs (no.)	32	59	604
Polygons (no.)	37	31	222
Arc length (m)	24	4	16

minimum number of points needed; straight two-point line segments have replaced multipoint curves within the spatial resolution tolerance of the original graphic map. The actual physical size of a GIRAS file is basically a function of the numbers of arcs (NA), coordinates (NC), and polygons (NP): approximately (36NA + 2NC + 32NP) bytes, plus identifying and descriptive data as specified in the map header and text subfile sections. Arc lengths are given in terms of meters at the scale of the base map (1:250,000). The greater complexity of land use and land cover maps results in a large number of arcs and polygons recorded as compared to the number recorded for the associated maps.

GIRAS FILE FORMAT

The GIRAS file format is a specific physical implementation of the logical GIRAS data structure. The format was designed to optimize storage requirements, transfer operations, and sequential processing. As implemented in a computing environment based on 32-bit words, a GIRAS file is organized sequentially as 32-byte (8 bits per byte) logical records, blocked by whatever factor is appropriate for the particular storage device (disk or tape) and computing environment use. Each 32-byte record may consist of anywhere from 1 to 16 data-element fields, depending upon where, within the file, it is located. Each data-element field may contain only one of three types of data elements: (1) a 2-byte (16-bit) binary integer, (2) a 4-byte (32-bit) binary integer, or (3) a string of EBCDIC coded characters, 1-byte per character.

For ease of transfer to other computing environments, a GIRAS 32-byte binary file format may be translated into a card image character (coded in either ASCII or EBCDIC) file format. Each 32-byte binary record is expanded to an 80-character card image; a 2-byte binary integer field is converted to a 5-decimal-digit character field, and a 4-byte binary integer field is translated into a 10-decimal-digit character field. The EBCDIC coded characters within a binary-formatted file are translated without expansion, allowing two sequentially adjacent 32-byte records containing character data to be converted into one 80-character card image record.

A GIRAS file, either in binary or character format, logically consists of six or more subfiles. The general structure of a GIRAS file is shown in figure 2; details are shown in figures 3-8. The first six subfiles always exist (in the order shown in figure 2); the seventh, the text subfile, may or may not exist for data released by the National Cartographic Information Center (NCIC); the eighth, the associated data subfile, is meant for user-attached data and is never present in NCIC-released data. A map

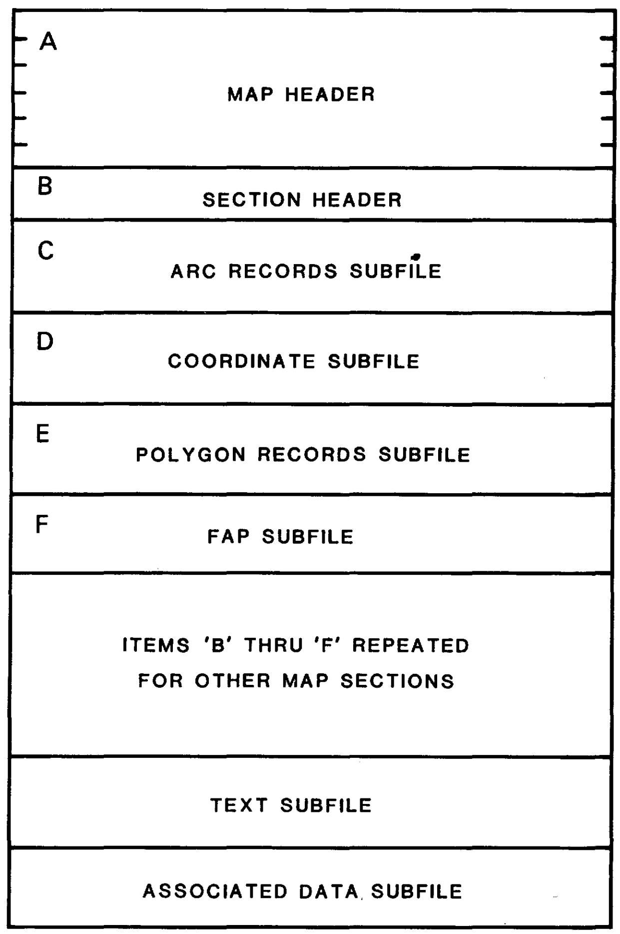


Figure 2.--GIRAS file structure.

data file may need to be divided spatially into several parts (sections) for processing purposes. For this reason, the second through sixth subfiles (fig. 2, B through F) may be repeated, one set of five subfiles per section. A text subfile (only one per GIRAS file) may follow the final section subfile.

THE STANDARD LOCAL UTM COORDINATE FRAME OF REFERENCE

A GIRAS file contains a number of different types of data elements--text, codes, identifiers, counters, pointers, and some derived measurement data. Most of these data elements will be explained under the descriptions of the various subfiles below. However, one type of data, the coordinate data, needs an overall discussion here. All coordinates stored within a GIRAS file are coded as 2-byte (16-bit) integers. The coordinate frame of reference is defined in the map header of the file by a projection code (MPJ, see fig. 3) and six control points. For each of the control points, the internal x,y coordinates are equated with the geographic (latitude and longitude) coordinates of the point. The resolution of an internal coordinate unit is indicated by the map scale (MSC) value in the map header. This value is the scale denominator of a graphic plot of the GIRAS file, if the data were plotted at one internal coordinate unit per mil (0.001 in.) on the plot.

A GIRAS file containing data based on a standard USGS 1:250,000- or 1:100,000-scale quadrangle (such as the land use and land cover and associated map series) is routinely stored in a local (to the map) UTM coordinate system (MPJ = "1" for UTM). Since 16 bits are not enough to store full UTM coordinates (which may exceed 4,000,000 m), the nearest 100,000-m UTM grid intersection, which is both west and south of all map control points, is used as a local origin (x,y = 0,0). Further, the resolution of an internal coordinate unit is set to 10 m (the MSC value = "393,701"). Using this coordinate referencing system, a GIRAS file may store data covering a 327,680-m square, more than enough for 1:250,000- and 1:100,000-scale quadrangle-based data.

MAP HEADER

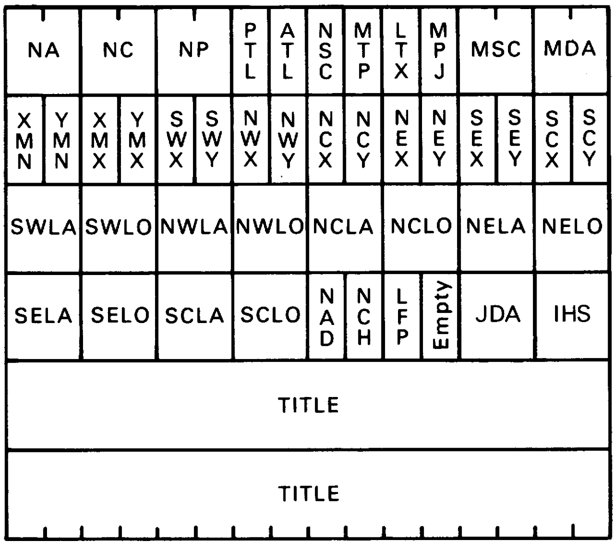
The map header (fig. 3) contains a substantial amount of information, including the amount of data in the file, the date of the source material, title information, and ground control information. The values in variables NA, NC, NP, and LFP represent totals of the corresponding values for each section of the map data file. For example, NA is the sum of the NAS (number of arcs in a section) values for all sections of the map data file. These values can be used to estimate the length of the file before analysis. The point tolerance (PTL) and arc length tolerance (ATL) are values used during processing and editing of the data to eliminate spurious or unneeded data. PTL is the width, in internal coordinate units, of a corridor that was used to delete unnecessary points from each arc of the map and reflects the relative accuracy of the original graphic lines. ATL is the minimum allowable length (in internal coordinate units) of any arc on the map. The possible map type codes (MTP) are listed in table 4.

Table 4.--Map type codes for GIRAS data base

Code	Map type
1	Land Use and Land Cover
2	Political Units
4	Census County Subdivisions
10	Hydrologic Units
20	Federal Land Ownership (optional)
40	State Land Ownership (optional)

The map date (MDA) is the year of the source material used to make the map, which is usually not the same as the year

MAP HEADER



Name	Length (bytes)	Description
NA	4	Number of arc records in map.
NC	4	Number of arc coordinates in map.
NP	4	Number of polygon records in map.
PTL	2	Point tolerance (in coordinate units).
ATL	2	Arc length tolerance (in coordinate units).
NSC	2	Number of map sections.
MTP	2	Map type code.
LTX	2	Length of text subfile.
MPJ	2	Map projection code.
MSC	4	Map scale denominator (1 coordinate unit per 1 mil).
MDA	4	Source map date.
XMN to YMX	2 ea.	Minimum and maximum x,y coordinates for map.
SWX to SCY	2 ea.	x,y coordinates of control points.
SWLA to SCLO	4 ea.	Latitude and longitude of control points.
NAD	2	Number of records of associated data.
NCH	2	Number of characters in TITLE.
LFP	2	Length of FAP subfile(s).
JDA	4	Julian Date of file creation.
IHS	4	Time of file creation, in hundredths of a second.
TITLE	64	Title of up to 64 characters.

Figure 3.--GIRAS map header structure.

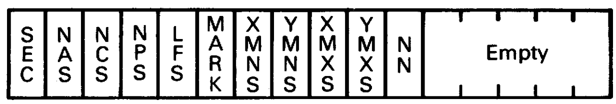
the map or data set is published. The number of characters (NCH) in the title gives information to ease the reading of the map title (TITLE) that occupies the first NCH bytes of the final two records of the map header.

SECTION HEADER

Because of computer constraints, the number of x,y coordinates (twice the number of arc points) of a section of a map data file is held to 32,000; the number of arcs, to 2,500; the number of polygons, to 1,500; and the total length of a FAP (file of arcs by polygon) subfile, to 6,000. Where these limits are exceeded within one map data file, the map area is broken into more than one section.

Four elements near the end of the section header (fig. 4; XMNS, YMNS, XMXS, and YMXS) are the coordinate limits of the section, and indicate the minimum and maximum x and y coordinate values within the section.

SECTION HEADER



Name	Length (bytes)	Description
SEC	2	Section number.
NAS	2	Number of arcs in section.
NCS	2	Number of arc coordinates in section.
NPS	2	Number of polygons in section.
LFS	2	Length of subfile assigning arcs to polygons.
MARK	2	Indicates processing history.
XMNS, YMNS	2 ea.	Minimum x,y coordinates in section.
XMXS, YMXS	2 ea.	Maximum x,y coordinates in section.
NN	2	Number of nodes in section.

Figure 4.--GIRAS section header structure.

ARC RECORDS SUBFILE

Each record of the arc records subfile (fig. 5) contains a pointer (PLC) to the x,y coordinates that represent the arc in the coordinate subfile. There is one PLC value for each arc, and it represents the position within the coordinate subfile of the last coordinate of that arc. For example, if the first arc contained 6 points (12 coordinates), its PLC value would be 12, and if the second arc contained 8 points (16 coordinates), the PLC value would be 28 (12+16). Following this pattern, the last arc would have a PLC value equal to the total number (NCS) of coordinates in the coordinate subfile. Along with the PLC value, each record in the arc records subfile contains the unique numbers (PL and PR) of the polygons that this arc separates and the types (PAL and PAR) of polygons between which this arc forms the border.

ARC RECORD

AID	PLC	PL	PR	PAL	PAR	XMNA	YMNA	XMXA	YMXA	ALEN	SN	FN

Name	Length (bytes)	Description
AID	2	Arc number.
PLC	2	Position of last arc coordinate in COORDINATE subfile.
PL	2	Polygon number of polygon to left of arc.
PR	2	Polygon number of polygon to right of arc.
PAL	4	Attribute of polygon to left of arc.
PAR	4	Attribute of polygon to right of arc.
XMNA, YMNA	2 ea.	Minimum x,y coordinates in arc.
XMXA, YMXA	2 ea.	Maximum x,y coordinates in arc.
ALEN	4	Arc length in coordinate units.
SN	2	Node number at beginning of arc.
FN	2	Node number at end of arc.

Figure 5.--GIRAS arc records subfile.

COORDINATE SUBFILE

The coordinate subfile (fig. 6) is simply a sequential listing of every x,y coordinate needed to represent the arcs of the map section. When a map is digitized, a series of x,y coordinates is recorded for each arc of the map. Although the direction taken by the digitizer during recording is not significant, that direction--once determined by digitizing--becomes important in the data structure. The direction of recording is referred to as the positive direction

COORDINATE DATA SUBFILE

X_1	Y_1	X_2	Y_2	X_3	Y_3									→
→													$X_{NCS/2}$	$Y_{NCS/2}$

Variable	Length (bytes)	Description
x_n	2	x integer coordinate value of point n
y_n	2	y integer coordinate value of point n
COORDINATE (PLC(I-1)+1)		is the x coordinate value of the first point of arc I. [PLC(0)=0]
COORDINATE (PLC(I))		is the y coordinate value of the last point of arc I.

The order within the COORDINATE subfile of the (x,y) points of a given arc determines the direction of the arc, and therefore the right and left of the arc. The first point in an arc (x,y) string is its "FROM" or "START" node; the last point in an arc (x,y) string is its "TO" or "END" node.

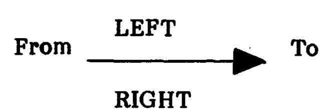


Figure 6.--GIRAS coordinate subfile.

for the arc, and each arc can be referred to by a positive or negative representation of its arc identification number (AID). For example, if arc 1 is read from its starting node (SN) to the node (FN) at the end of the arc, it is represented as "+1" or 1. If the arc is to be read from FN to SN (in reverse order from that stored in the Coordinate subfile), it may be represented by "-1." The need for this will become apparent when the FAP subfile is described.

To repeat, the order in which the arcs are recorded is not significant, but their order, once established, is important to the GIRAS structure. The order allows each arc to be accessed by using only the unique arc number (AID) and its PLC value. No pointers to the beginning coordinate of an arc exist in the coordinate subfile. For example, if the PLC for arc 1 is 16, then the 17th element of the coordinate subfile will be the first coordinate of arc number 2. This permits storage of the x,y coordinates in the smallest possible space.

POLYGON RECORD

PID	PLA	CX	CY	ATT	AREA	XMNP	YMNP	XMXP	YMXP	PERL	NIW	NIP

Name	Length (bytes)	Description
PID	2	Polygon number.
PLA	2	Position of last arc number of polygon in FAP subfile.
CX, CY	2 ea.	x,y coordinates of an interior point (arbitrarily positioned).
ATT	4	Polygon attribute.
AREA	4	Area of polygon in coordinate units squared.
XMNP, YMNP	2 ea.	Minimum x,y coordinates of polygon.
XMXP, YMXP	2 ea.	Maximum x,y coordinates of polygon.
PERL	4	Perimeter length of polygon in coordinate units.
NIW	2	Number of islands contained within polygon.
NIP	2	Number of the polygon containing this polygon, if it is an island.

Figure 7.--GIRAS polygon records subfile.

POLYGON RECORDS SUBFILE

The polygon records subfile (fig. 7) describes each polygon of the map section. The polygon identification number (PID) is the unique number by which each polygon may be referenced. The PLA value (equivalent in purpose to PLC in an arc record) points to a list, in the FAP subfile, of the arcs that comprise that polygon's borders. As with the coordinate subfile (referenced by PLC), the FAP subfile is a sequential list of information. For each polygon it contains the list of the numbers (AID's) of the arcs needed to create the boundary of the polygon. These values are listed first for polygon 1, second for polygon 2, and so on for each polygon. If polygon 1 of a section were the three-sided polygon with attribute 76 in figure 1, it would be composed of three arcs, and the PLA value for it would be "3." If polygon 2 were the polygon in figure 1 labeled 42, the situation would be slightly more complex because the FAP subfile entry for this polygon must also include the arcs that make up the outside boundary of the island within polygon 2. In this case, the value of PLA would be the number of arcs it takes to enclose the polygon (4) plus one place for a zero (indicating that what follows is a list of arcs making up an island) and one place for an arc number to represent the island (2) plus the PLA of the previous polygon (3). The PLA value for polygon 2 would then be 9. It also follows that the PLA value for the last polygon will equal the total length of the FAP subfile (LFS in the section header). A detailed explanation of the FAP subfile begins on page 13.

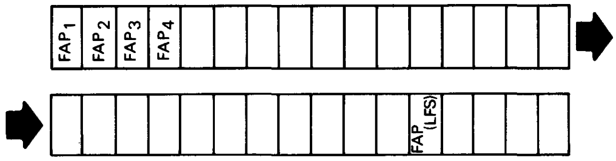
The next entries in the polygon record are the x,y coordinates (CX and CY) of a point inside the polygon. These coordinates do not represent a centroid or center of mass. They define a point inside the boundary of the polygon that was recorded during the digitizing process and used during the editing process to tie the arcs together to form the polygon. The polygon attribute (ATT) and polygon area (AREA) also are included, as well as the minimum and maximum x and y coordinates (XMNP, YMNP, XMXP, YMXP) of the polygon. The area of a polygon is useful information and is stored to avoid recalculation. The limits of the polygon allow the user to--in a simple way--isolate a polygon from much of the rest of the map section. The perimeter length (PERL) of the polygon can be used along with the total area occupied by the polygon to compute measures of the compactness of the polygon.

The final two entries in the polygon record are the number of islands (NIW) contained within the polygon and the number (NIP) of the polygon that contains another polygon as an island. The number of islands is helpful when used along with PLA to read the island entries in the FAP subfile. The NIP entry identifies a polygon totally surrounded by another polygon whose attribute can easily be obtained by looking at the value of ATT in the record for polygon number NIP. If the data were to be compressed and small polygons to be eliminated or combined with larger surrounding polygons, NIP would show quickly which polygon number and type that small island would become.

FAP SUBFILE

The FAP (file of arcs by polygon) subfile (fig. 8) is the last subfile that exists separately for each map section. It is accessed by way of an entry (PLA) in the polygon records subfile. The total length (LFS) of the FAP subfile is stored in the section header. To understand the contents of the FAP subfile, it is helpful to know how its contents are used to construct a polygon. The FAP subfile consists of lists of arcs, one arc-list per polygon. The arcs for a polygon with no islands are listed in clockwise order around the polygon from an arbitrary starting point. This starting point can be either the starting node (SN) or final node (FN) of the first arc

FAP SUBFILE



Each record of the FAP subfile consists of 16 elements (2 bytes each) for the FAP array.

Each element contains an arc identification number pointing to the arcs that make up a specific polygon.

FAP (PLA(I-1)+1) is the first arc bordering polygon I.
[PLA(0)=0]

FAP (PLA(I)) is the last arc bordering polygon I.

Within the FAP subfile, the identification numbers of the arcs constituting a given polygon are ordered clockwise around the perimeter of the polygon and counterclockwise around interior islands of the polygon.

A positive arc identification number indicates the polygon is to the right of the arc.

A negative arc identification number indicates the polygon is to the left of the arc.

A zero FAP element indicates that the arc identification numbers following it, and before the next zero in the subfile or the end of the FAP list, point to the arcs that constitute an interior island of the polygon.

Figure 8.--GIRAS FAP subfile.

in the list. If the starting point is SN for the arc (the polygon to the right of the arc), then the number of that arc is recorded in FAP as "+AID" or "AID." If the starting point is FN (the polygon is to the left of the arc), then that arc is recorded as "-AID." In figure 9, the direction in which the arcs of a polygon were digitized (positive direction) is indicated by the filled arrows. The list of arcs composing the polygon begins at the node closest to the internal point (CX, CY), and the FAP entry for this polygon is "-1 4 3 -2."

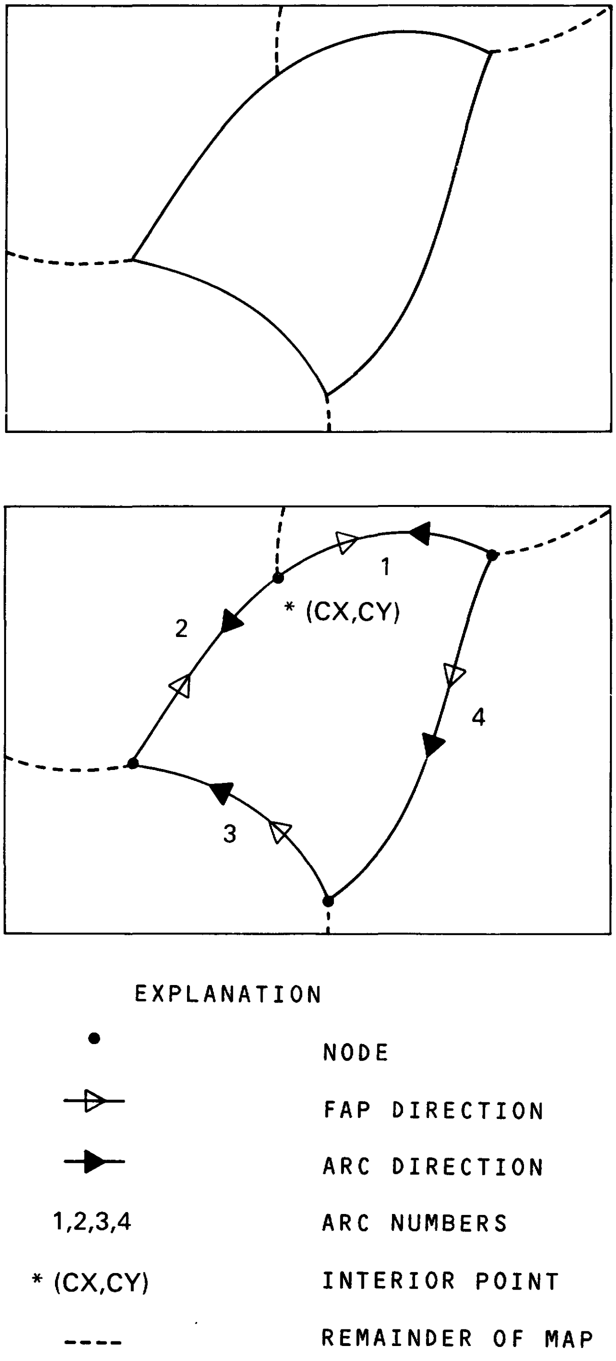


Figure 9.--GIRAS FAP subfile creation.

The FAP subfile contains a series of elements for each polygon whether or not that polygon is an island. If the polygon is an island, a list of the arcs that make up its outside boundary will also be listed in counterclockwise order in the FAP entry for the surrounding polygon in which it is an island.

This points out one factor not yet mentioned in describing the FAP subfile. In many maps there may be two types of islands--simple and complex. Examples of both island types can be seen in figure 1. A simple island is a polygon that stands alone, totally surrounded by one other polygon and directly bordered by only that polygon. A complex island is a cluster of adjacent polygons that, as a group, are totally surrounded by one other polygon. When the arcs that make up a complex island are listed--in counterclockwise order--in the FAP subfile, only the arcs that compose the outside boundary of the entire cluster are recorded. The individual identities of the polygons that make up the cluster are not maintained because the cluster is considered to be an area to exclude from the polygon being described.

The FAP subfile is crucial to the GIRAS data structure because it is the way individual arcs can be described as part of the polygons they compose. From this file the entire outline of a polygon can be obtained, including island areas within its boundary that are to be excluded. This information is used, for example, when computer-generated shaded color plots are made.

TEXT SUBFILE

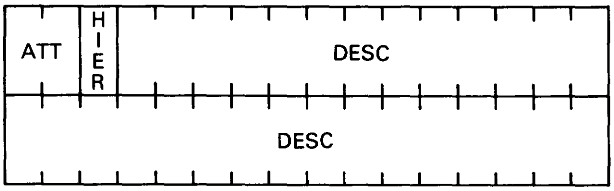
The text subfile in a GIRAS file is reserved as a place where numerical polygon codes are assigned textual labels. An individual GIRAS file includes only a single text subfile that physically follows the data for the various map sections.

Every record in the text subfile consists of three elements--an attribute code, a hierarchy code, and a descriptor (fig. 10).

A text subfile record format of (A4, A2, 13A2,/16A2) was designed to be compatible with the standard binary representation of GIRAS files. Each logical text subfile entry comprises two GIRAS 32-byte records.

Included in the text subfile are records for general codes that provide

TEXT RECORD



Name	Length (bytes)	Description
ATT	4	Polygon code, right justified in a 10-digit field.
HIER	2	A 5-digit parameter to indicate polygon code type (that is, specific, general, or special).
DESC	58	A label or descriptor defining the polygon code, left justified, in a 58-character field.

Figure 10.--GIRAS text record subfile.

labels or descriptors for the more general categories under which specific polygon codes come. Consider references made to Level I and Level II land use and land cover categories. A polygon coded as "13" is known to be Industrial. The digit '1' indicates that it falls under a broader category of Urban and Built-up Land. General code records also appear in the text subfile to provide definitions for these more general categories. For the above example, a text subfile record with "10" in the polygon code (ATT) field would have the descriptor Urban and Built-up Land. (The "0" is for positional purposes only and can be considered insignificant.)

The hierarchy code (HIER) indicates the degree of generality of a text subfile polygon code (classification level). Specifically, the hierarchy code is defined as the number of insignificant digits on the right-hand side of the polygon code. Our sample land use and land cover text subfile records would thus appear:

________10____1URBAN AND BUILT-UP LAND

________13____0INDUSTRIAL

The role of the hierarchy code is more appreciated when dealing with other classification schemes utilized by GIRAS that involve multiple levels.

Text subfile records are arranged in ascending order by polygon code. Consequently, more general codes precede general codes of greater detail, which in turn precede codes explicitly referenced in the file. This ordering can be considered an indentation-type structure, as illustrated in table 5 for a sample census unit subfile.

Special polygon codes are required where available codes are not suitable. The set of special codes used in GIRAS is defined at the end of the subfile, as also shown in table 5. Special polygon codes were intentionally assigned extreme values (greater than 2,000,000,000) to force them to the bottom of the text subfile.

The type of map stored in a GIRAS file has an impact on the nature of the text subfile. Land use and land cover and Federal ownership maps reference classification schemes that are attribute in nature. That is, there are a relatively limited number of possible codes (37 for land use and land cover and 26 for Federal ownership), and many polygons may have the same codes. The text subfiles for these types of GIRAS files list text records for all codes in the classification scheme. Consequently, all land use and land cover text subfiles are identical, as are Federal ownership text subfiles.

Census unit files, political unit files, and hydrologic unit files, alternatively, utilize codes that serve as unique identifiers. Two polygons rarely have the same code. Furthermore, it would not be feasible to list all possible codes. Consequently, only codes explicitly referenced in the file (and respective general and special codes) are defined in these types of GIRAS text subfiles.

The text subfiles for State ownership maps are not processed because coding schemes vary from situation to situation.

The length of the text subfile is given by the LTX parameter in the GIRAS map header. LTX refers to the number of logical text subfile records (that is, polygon code definitions) and includes general and special polygon codes.

Table 5.--Sample text subfile for a Census County Subdivision GIRAS file

42000000	6PENNSYLVANIA
42051000	3FAYETTE COUNTY
42051095	0LURERENE
42059000	3GREEN COUNTY
42059005	0ALEPPO
.	
.	
.	
.	
42059127	0WAYNE
42059130	0WHITELEY
54000000	6WEST VIRGINIA
54001000	3BARBOUR COUNTY
54001005	0BAKER
54001015	0ELK
54001025	0PHILIPPI
.	
.	
.	
1628000000	6PITTSBURGH
1628732000	2WASHINGTON
1628732000	0WASHINGTON
1628762000	2WASHINGTON
1628762000	0WASHINGTON
.	
.	
.	
1900000000	6WHEELING
1900020100	2MARSHALL
1900020100	0MARSHALL
1900020200	2MARSHALL
1900020200	0MARSHALL
.	
.	
.	
1900020900	0..........
2000000000	9SPECIAL CODES
2000000100	2AREA UNDEFINED BY CODING SCHEME
2000000101	0LAND AREA UNDEFINED BY CODING SCHEME
2000000102	0WATER AREA UNDEFINED BY CODING SCHEME
2000000200	2UNMAPPED AREA
2000000201	0UNMAPPED U.S. AREA
2000000202	0UNMAPPED NON-U.S. AREA

ASSOCIATED DATA SUBFILE

The final GIRAS subfile is the associated data subfile that allows the user to store information of his own that pertains to the file (for example, population information that pertains to a particular census map). A pointer (NAD) to this subfile provides a place in the map header subfile to store the number of records in the associated data file once it is created.

COMPOSITE THEME GRID (CTG) DATA FILE FORMAT

Digital data from all the overlays of a given quadrangle also are combined in a raster or grid-cell format as a Composite Theme Grid (CTG) file.

A CTG file is available in either binary or character format. In either format, the files are sequential and consist of fixed-length logical records, all of identical internal format, one grid cell per logical record. The grid cells are actually a regular point sample. The attribute codes at the center point of each cell are recorded from each overlay. The points are oriented to the UTM projection and are usually spaced 200 m apart in both east-west and north-south directions. The cell records are first ordered in the file by row from north to south, then within each row, by column west to east.

Header data, usually physically separate from the CTG data file, contain the following data necessary to complete the description of an individual CTG data file:

1. Title;
2. Number of overlays and which overlays are represented;
3. Numbers of rows and columns;
4. UTM zone number;
5. UTM Easting of the west edge of cells;
6. UTM Northing of the north edge of cells; and
7. Cell width in meters.

BINARY CTG DATA FILE FORMAT

Each logical record of a binary CTG file is 52 bytes in length. A record consists of 13, 32-bit (4-byte) binary integers in the following order:

Bytes 1- 4 = Row index, where 1 is the index of the northern-most row and index numbers increase by one for each row moving south (NOTE, due to a processing error, CTG files in which the State ownership is not coded will have all zero row index numbers; the row index is then a function of the sequential position of the record within the file);

5- 8 = Column index, where 1 is the index of the western-most column and index numbers increase by one for each column moving east;

9-12 = Land use and land cover code;

13-16 = Political unit code;

17-20 = Hydrologic unit code;

21-24 = Census county subdivision or SMSA tract code;

25-28 = Federal land ownership code;

29-32 = State land ownership code; and

33-52 = Null (binary zeros) fields.

If a given overlay has not been digitized, the codes for that overlay will all be zero. To be sure that a regular grid of cells (forming a UTM rectangle) covers the entire base map quadrangle, a "buffer zone" of cells with all zero attributes has been included in the binary CTG data file.

Binary CTG Map Header

The header data associated with a binary CTG data file are stored in a physically separate sequential file, modeled after the GIRAS formatted map header. The header consists of six, 32-byte logical records. All elements in the first four records are coded as either 16-bit (2-byte) or 32-bit (4-byte) binary integers. The title in the fifth and sixth records is coded as a string of EBCDIC characters (one character per byte).

Record 1:

Bytes 1- 4 = Number of rows;

5- 8 = Total number of cells x 2;

9-12 = Number of columns;

13-14 = Meaningless field (duplicate point tolerance);

15-16 = Cell size (width and length) in meters;

17-18 = Number of overlays merged;

19-20 = Map type code (see below);

21-22 = Projection zone number;

23-24 = Map projection code (should be "1" for UTM);

25-28 = Scale of a plot at one mil per cell width; and

29-32 = Source date of the land use overlay.

Record 2:

Bytes 1- 2 = Minimum column index;

3- 4 = Minimum row index;

5- 6 = Maximum column index;

7- 8 = Maximum row index;

9-10 = Column index for SW control point;

11-12 = Row index for SW control point;

13-14 = Column index for NW control point;

15-16 = Row index for NW control point;

17-18 = Column index for NC control point;

19-20 = Row index for NC control point;

21-22 = Column index for NE control point;

23-24 = Row index for NE control point;

25-26 = Column index for SE control point;

27-28 = Row index for SE control point;

29-30 = Column index for SC control point; and

31-32 = Row index for SC control point.

Record 3:

Bytes 1- 4 = Latitude of SW control point;

5- 8 = Longitude of SW control point;
9-12 = Latitude of NW control point;
13-16 = Longitude of NW control point;
17-20 = Latitude of NC control point;
21-24 = Longitude of NC control point;
25-28 = Latitude of NE control point; and
29-32 = Longitude of NE control point.

Record 4:
Bytes 1- 4 = Latitude of SE control point;
5- 8 = Longitude of SE control point;
9-12 = Latitude of SC control point;
13-16 = Longitude of SC control point;
17-20 = UTM Easting value of west edge of cells;
21-24 = UTM Northing value of north edge of cells;
25-28 = File creation date (a Julian date); and
29-32 = Meaningless field (time of file creation).

Records 5 and 6:
Bytes 1-32 = Title (text characters in EBCDIC coding).

Some further explanation is needed for some of the elements in the CTG map header:

1. The map type code (in bytes 19-20 of the first record) indicates which overlays have been included in the CTG data file. The code is formed by the addition (in base 10) of the separate GIRAS map type codes for each of the overlays:

   01 = Land use and land cover;
   02 = Political units;
   04 = Census county subdivisions and SMSA tracts;
   10 = Hydrologic units;
   20 = Federal land ownership; and
   40 = State land ownership.

   For example, the map type code for a combination of the first four overlays above would be 17; all six overlays combined have a map type code of 77.

2. The UTM Easting and Northing values given in the fourth record (bytes 17-24) are in whole meters and are values for the west and north edges of the cells, rather than the center point of the first (northwest corner) cell. The Easting and Northing values for a given cell may be calculated thus:

   Easting = (XORG-CW/2) + (column index)*CW

   Northing = (YORG+CW/2) - (row index)*CW

   where XORG and YORG are the Easting and Northing values in bytes 17-24 of the fourth header record, and CW is the cell width in bytes 15-16 of the first header record.

3. The control points usually define the 1° x 2° (for 1:250,000-scale base maps) or 30' x 1° (for 1:100,000-scale base maps) quadrangle on which the overlay data are based. The latitude and longitude values are given as positive integers of the form DDDMMSS, where DDD is degrees, MM is minutes, and SS is seconds. Western longitude values are given as positive numbers, increasing in value from east to west. The row and column values given for the control points are the indices for the cell whose center point is closest to the true position of the control point.

CHARACTER COMPOSITE THEME GRID (CTG) FILE FORMAT

For ease of transfer and readabiltiy, CTG files also are available in character-coded (ASCII or EBCDIC) format. Each binary integer value or code has been converted to a number, right justified (with leading blanks if needed) within a fixed-length field of an 80-character record. All character CTG data files consist of fixed length 80-character logical records, one grid cell per logical record. Two internal record formats are available--one with row and column index numbers such as the binary CTG format and the second (and standard) format with UTM zone, Easting, and Northing values in place of row and column indices.

The row and column number format is a direct copy of a binary CTG file with each 4-byte binary integer represented as a 10-digit character-formatted integer. The null fields (bytes 33-52) of the binary CTG data records are not copied. Each character CTG data record in this format has eight, 10-digit integer fields--row and column indices followed by the six attribute codes in the same order as in the binary CTG data file.

Standard Character-Formatted CTG Data File

Each logical record of a standard character-formatted CTG data file is 80 characters in length and consists of nine decimal integers, right justified (with leading blanks) within fixed-length fields:

Bytes 1- 3 = UTM zone number (this value should be the same in every record of a given CTG file); the first byte will always be a blank for zones in the northern hemisphere;

4-11 = UTM Easting value, in whole meters, for the sample point of the cell;

12-19 = UTM Northing value, in whole meters, for the sample point of the cell;

20 = Blank;

21-30 = Land use and land cover attribute code;

31-40 = Political unit (FIPS State/county) code;

41-50 = USGS hydrologic unit code;

51-60 = Census county subdivision or SMSA tract code;

61-70 = Federal land ownership agency code; and

71-80 = State land ownership code.

If a given overlay category has not been included within the file, the codes for that category will be zero (0). Since some misregistration of map overlays occurs, some of the cells along the edges of the 1:250,000- or 1:100,000-scale quadrangle may have codes for some overlays, but not others (the "other" code(s) will be zero). The standard character CTG data file will have only those cell records for which at least one of the categories is coded. This means that, since the 1:250,000- and 1:100,000-scale quadrangles do not form perfect rectangles in the UTM projection (lines of latitude curve and lines of longitude converge), a variable number of cell records will exist for any given row or column.

Character CTG Map Header

CTG map header data associated with a character CTG data file may exist in one of three states: (1) as a separate file, (2) as the first five records of the character CTG data file; and (3) as a printed listing. In either of the first two forms, the data exist in five 80-character records. The first four records are exact copies of their binary CTG header counterparts, but with the 16-bit (2-byte) integers of the first two records converted to five-digit integer fields and the 32-bit (4-byte) integers of records 1, 3, and 4 converted to 10 digit integer fields (right justified with leading blanks). The title is carried as the first 64 characters of the fifth record (with 16 blanks to fill the end of the record).

A listing of the CTG map header data, which is meant to be delivered to a user with the CTG data file on tape, lists all the elements (see the list under the description of a binary CTG map header) of the header (but not in exactly the same order as in an actual CTG map header).

A sample listing of CTG map header data (as produced by the binary-to-character conversion program CTGBTA) is shown in the Appendix.

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. Geological Survey Professional Paper 964, 28 p.

Mitchell, W. B., Guptill, S. C., Anderson, K. E., Fegeas, R. G., and Hallam, C. A., 1977, GIRAS--A geographic information retrieval and analysis system for handling land use and land cover data: U.S. Geological Survey Professional Paper 1059, 16 p.

U.S. Bureau of the Census, 1972a, Census tracts--1970 census of population and housing: U.S. Bureau of the Census Final Reports PHC(1)-1 through PHC(1)-241. [One volume for each SMSA.]

__________1972b, County and city data book, 1972: U.S. Bureau of the Census, 1020 p.

__________1972c, Geographic identification code scheme--1970 census of population and housing: U.S. Bureau of the Census Reports PHC(R)-1 through PHC(R)-4. [One volume for each of the U.S. regions: NE, NC, S, and W.]

APPENDIX

Listing of CTG Map Header Data

COMPOSITE THEME GRID CHARACTER DATA OUTPUT:

C T G B T A   RUN: JUNE 3, 1982   TIME 19:23:06

GRID CELL MAP HEADER INFORMATION:

TITLE: LAWRENCE, MO KS 1:250,000 QUAD LU PB CN HU FO SO

FILE CREATION DATE: 81084   TIME 0: 0

MAP TYPE: 77   PROJECTION: 1   SCALE 1: 7874016   MAP DATE: 1973

NUMBERS OF FILE ELEMENTS:

CATEGORIES	CELLS	ROWS	COLUMNS	ZONE NUMBER	WEST & NORTH EDGES: EASTING	NORTHING
6	485368	575	884	15	236900	4321100

DUPLICATE POINT TOLERANCE = 0   CELL SIZE IN METERS = 200

MIN COL = 1   MIN ROW = 1   MAX COL = 884   MAX ROW = 575

CONTROL POINT INFORMATION:

	LONGITUDE	LATITUDE	COL	ROW
SOUTH WEST	960000	380000	-1	557
NORTH WEST	960000	390000	17	2
NORTH EAST	940000	390000	883	21
SOUTH EAST	940000	380000	877	576

CHARACTERISTICS OF THE CHARACTER CTG FILE:

THE FILE CONTAINS ONLY GRID CELL (AND NO HEADER) RECORDS.

THE FILE CONSISTS OF 80 CHARACTER RECORDS, ONE GRID CELL PER RECORD.

UTM ZONE, EASTING, AND NORTHING VALUES ARE PART OF EACH CTG DATA RECORD AS THE FIRST THREE INTEGERS, RIGHT JUSTIFIED IN BYTES 1-3, 4-11, AND 12-19.

BYTES 21-80 OF EACH RECORD CONTAIN THE USGS 10-DIGIT INTEGER CODES, RIGHT JUSTIFIED WITHIN 10-BYTE FIELDS, FROM THE FOLLOWING OVERLAYS, IN ORDER:

LAND USE/LAND COVER, POLITICAL UNIT, HYDROLOGIC UNIT, CENSUS SUBDIVISION/TRACT, FEDERAL LAND OWNERSHIP, AND STATE LAND OWNERSHIP.

ONLY RECORDS WITH AT LEAST ONE NON-ZERO ATTRIBUTE ARE PART OF THE FILE. (A VARIABLE NUMBER OF RECORDS EXIST FOR A GIVEN ROW OR COLUMN.)