

**KANSAS GEOLOGICAL SURVEY**  
**OPEN-FILE REPORT 88-45a**

Mapping by Computer  
Introduction and GIMMAP View

by

Charles G. Ross

*Disclaimer*

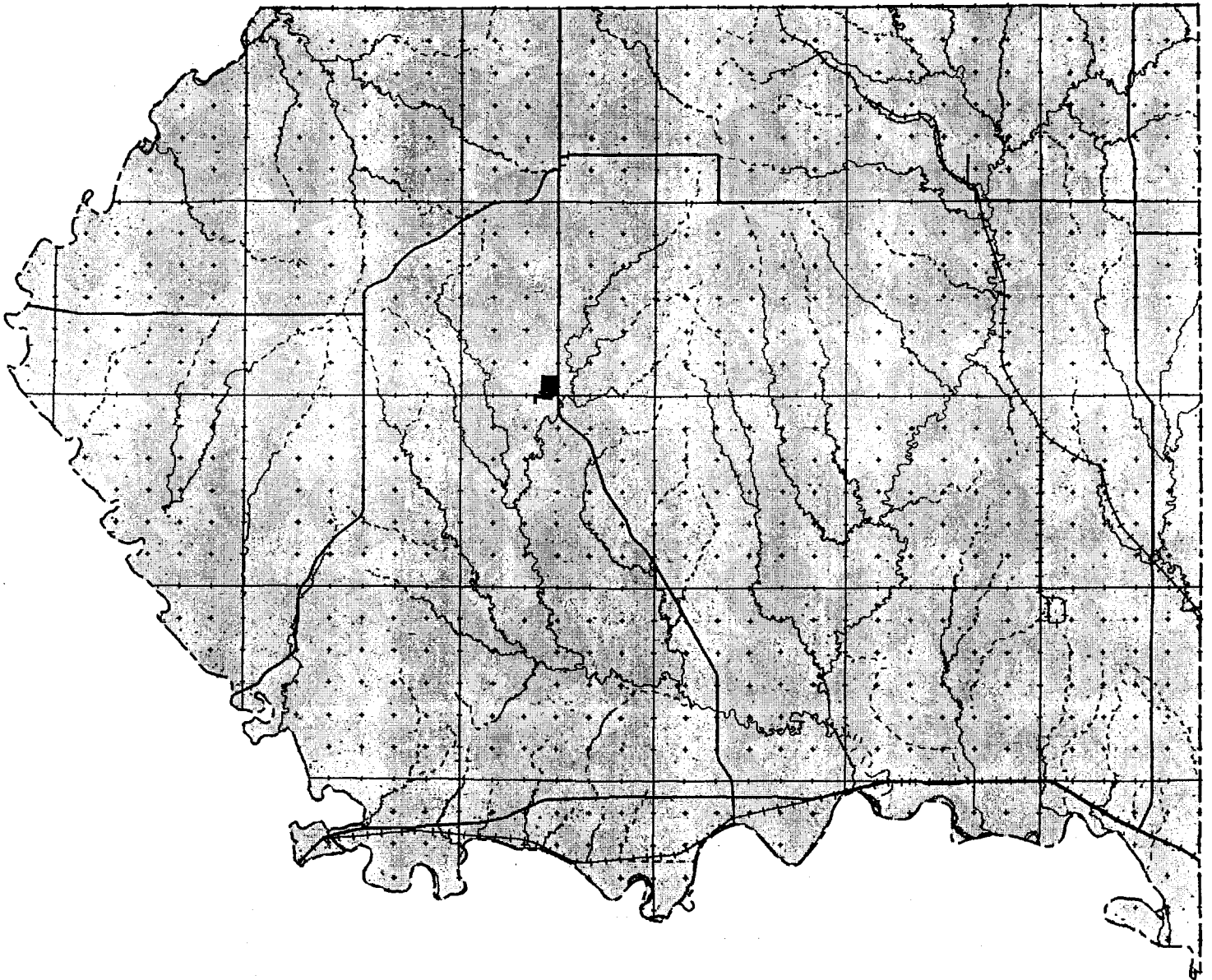
The Kansas Geological Survey does not guarantee this document to be free from errors or inaccuracies and disclaims any responsibility or liability for interpretations based on data used in the production of this document or decisions based thereon. This report is intended to make results of research available at the earliest possible data, but is not intended to constitute final or formal publications.

**KANSAS GEOLOGICAL SURVEY**  
1930 Constant Avenue  
University of Kansas  
Lawrence, KS 66047

# Mapping by Computer

## Introduction and GIMMAP view

by Charles G Ross



KANSAS GEOLOGICAL SURVEY

TO:

Joe's mom, Gene, Van, and Bear

"the best part about writing is the finishing"

(quoted by Steven Spielberg)

"All good writing is swimming under water  
and holding your breath."

F. Scott Fitzgerald

## Foreword

The compilation of this information about the Geodata Interactive Management Map Analysis and Production or GIMMAP ("jim-map") system has been a long process, beginning with the first programs written in 1977. As with the conquest of the great mountains, there have been many different attempts made over the years. This set of separate volumes has finally resulted from the union of the efforts of those people involved directly in the creation and use of GIMMAP with the support of the administration of the Kansas Geological Survey.

First, I must acknowledge the people who have contributed to the completion of this work.

I thank my wife, Gina, and my children, Catherine and Patrick, who are my reason for living. And I must credit Gina, whose love and encouragement helped in many ways to bring me to the light at the end of this tunnel.

Trang Cao initiated GIMMAP in Orleans, France while he was working for the BRGM there. He brought ideas and the foundation for the system when he was Visiting Research Scientist in the 1977-78 academic year. His ideas and good will in working with me made the potential for an automated cartography system a reality at the Survey. His work is a shining example of what can be.

David Collins helped to provide momentum, direction and many, small packets of inspiration, and helped greatly to organize and edit the final effort.

Finally, this work would not have been completed without the commitment to the definition and prioritization of the project by Dr. Lee Gerhard, Director of the Kansas Geological Survey.

Creation of these four volumes of GIMMAP documentation has been a long and arduous process, marred by difficulties which might have been avoided. Important lessons have been learned about the process of creating software documentation and I would feel remiss if some of these were not passed on to. The foreword may not be the best place to air these thoughts, but I feel they are of a fundamental importance.

Of the lessons learned, these might be the most important:

1. The first step is to clearly define in writing the task which is to be done. This includes identifying the target audience, the purpose and function, and the form and scope of the documentation.
2. For the job to be well-defined and well done, a group of people must participate in overseeing and performing the task at all stages. This group should consist of administrators, people who use or benefit from the software, the software author(s), and experts in software, education, technical writing and marketing. Strong and generous support for this group and the documentation authors by the administration in organization, project definition and the provision of resources is required for success.
3. The software author is probably not the best nor even a good choice to be the author of the documentation. The skills required to write software documentation are not the same as those required to write computer software. Generally and historically, an experienced writer with knowledge of the software is the best choice. The software author is the best source of technical information on the software, prepared as draft material for the software documentation during implementation of the software. Comprehensive documentation of a complex software system such as GIMMAP can not be properly prepared in scattered pieces of time by software authors who have too many responsibilities in creating, expanding and maintaining the software.
4. The group of experts (above) should estimate the cost of the project if it is to be done right. Money, time, personnel and other resources should be discussed and consensus reached. To do this, clear definition of the priorities of all affected personnel must be made. Overlapping expectations coupled with improperly low estimations of the size of the task have led to disappointments in the past.
5. The reason(s) for writing the software in the first place should be well understood and accepted by the group as the foundation of the discussions implied above.

- D R A F T -

P A R T I. I N T R O D U C T O R Y I N F O R M A T I O N

CHAPTER I.1 INTRODUCTION

Section I.1.1 Mapping By Computer

Topic I.1.1.1 Maps and Cartography

I.1.1.2 Computer Cartography

I.1.1.3 Types of Systems

I.1.1.4 Introduction to Hardware

I.1.1.5 Advantages of Cartography By Computer

I.1.1.6 Disadvantages of Computer Cartography

I.1.1.7 Geographic Information Systems

I.1.2 Computer Mapping With GIMMAP

I.1.2.1 "G I M M A P"

I.1.2.2 History of Development

I.1.2.3 A Computer-Assisted Cartography System

I.1.2.4 Philosophy of Development

I.1.2.5 Functions of the GIMMAP System

I.1.2.6 Operations and Database Administration

I.1.2.7 Introduction to Software

I.1.2.8 Expandability, Portability and Limits

I.1.2.9 Hardware Requirements With GIMMAP

I.1.3 About This Manual

I.2 THE GIMMAP VIEW OF MAP DATA

I.3 PROCESSES: THE BASIC SEQUENCE FOR GIMMAP

I.4 HARDWARE USED BY GIMMAP

I.5 DIGITAL REPRESENTATION OF GIMMAP DATA

I.6 SOFTWARE OVERVIEW (Diagrams and Paragraphs)

I.7 LOGICAL AND PHYSICAL DATA SEGMENTATION

I.8 ADDITIONAL PROCESSES AND SEQUENCES

II. BASIC SEQUENCE

III. EXTENDED FUNCTIONS

IV. PROGRAM LIBRARIES

V. ADDITIONAL TOPICS

VI. SYSTEM UTILITIES

VII. PROGRAMMING SECTION

VIII. APPENDICES

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.1 Maps and Cartography

According to Webster, **cartography** is the science or art of making maps. A **map** is defined as a representation (usually on a flat surface) of the whole or a part of an area. The area being represented is usually a part of a curved, 3-dimensional surface, which is most often the surface of the earth.

The art of making maps involves things like style and aesthetics. It is the selection of some options and the rejection of others. It is the collection of choices defining the content and form of the finished product. It involves the choice of line types (solid, dashed, hashed...), thicknesses and colors. It is the selection of media, map title and legend, ink, point symbols and area fill patterns or colors. It is the choice between emphasizing some features and omitting others.

In manual map-making, the art also includes the physical construction and implementation of these decisions, the transfer of mental images into the final or intermediate physical form of the map. This part of the art of making maps has been largely converted to science in **computer cartography**, the process of making maps by computer. However, the art common to both methods remains largely the same for both, though some differences do exist.

It is the science of map-making which is very much different between maps made manually and those made by computer. Most dramatic is the change in the tools used for production. Tools used in manual cartography include the hand-held pens and scribing tools, straight edges, compasses, french curves, erasers and opaquing brushes. In computer cartography, the tools include the computer, computer programs, monochrome (single color) and multi-color graphics terminals, manual and automatic digitizing tables, and line-drawing or color dot plotters.

Perhaps the only science which is common to both manual and computer cartography is the science used to gather or create the very information which is to be portrayed on the map. This process may include the use of computers as most scientific work does today, but at this stage, the work is analytical science and not yet in the realm of cartography. However, at the point that the gathered information passes to the cartographer, original data in digital form may be far more advantageous to the computer cartographer than to the traditional cartographer.

The need for human comprehension of the processes of map construction and production process remain, though it is now

posited in a "high-tech" cartographer or a computer expert with some working knowledge of maps. The knowledge required to operate the manual tools for map production is replaced by an even greater requirement for knowledge of maps and a computer environment.

-----

Arthur Robinson, in his Elements of Cartography (1984, 5th Ed., John Wiley & Sons, NY, p.4) refers to the map as an "instrument for recording, calculating, displaying, analyzing, and in general understanding the interrelation of things in their spatial relationship." In this it should be clear that the computer can perform or assist in performing the functions of recording, calculating, displaying and analyzing maps. The understanding function remains that of the people who are interested in the map.

It is also clear that the **digital map** (the computer representation of a map) enhances the above functions. While the digital map can always be used to (functionally) duplicate the original graphic form of the map, its digital form can also provide for endless extensions of calculation, forms of display and types of analysis. The digital map will also support functions never dreamed of with paper maps (see Advantages of Computer Cartography).

There is no restriction on the content of maps as defined above, but in practice there are three fundamental classes of maps (Robinson, p.7-10). These classes are general maps, thematic maps and charts. General maps are those which display the spatial relationships of a diverse set of geographical features such as roads, boundaries, hydrology and topographic (elevation) contours.

Thematic maps usually display varying magnitudes of a single attribute (population, rainfall, temperature, income) with each unit area displaying a single, discrete value or range of values. Maps emphasizing a single attribute such as geologic formation would be called general maps. And finally, charts are maps designed for navigation and (unlike other maps) are usually intended to be written upon in this function. They are often considered a special form of general maps.

All of the kinds of maps have been produced or approximated with computer cartography. Most have been done well with one computer cartography system or another, but unfortunately, no one system can make all kinds of maps well. The GIMMAP system described here is considered a general-purpose system in that it is designed to perform functions necessary for all kinds of mapping. It is designed to be flexible enough to expand to meet the needs for additional kinds of mapping through addition of new programs rather than through rewriting of the system.

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.2 Computer Cartography

The past three decades have witnessed the conversion of cartography from a strictly manual operation into one now wholly dependent on the computer, computer programs and associated high-technology machinery. The infusion of digital technology was predictable and inevitable in cartography as it was in virtually every other field. This was due to the promises of increased productivity, faster turnaround, greater flexibility and (perhaps) even lower cost in the long term.

From this fusion of sciences grew a field which joined specialists from cartography, other geosciences, computer science and enough engineers to make the machinery perform. Much effort was required to design, build and improve the hardware (machines) which was specialized to perform digitization (the conversion of map features into computer form) done by a digitizing table, graphic display of map data (maps drawn on a terminal screen), and plotting (the drawing of the map onto paper or other media under the control of the computer) performed via a digital plotter. These developments were necessary for cartography to be done by computer, and they were found useful in numerous other fields. But, by themselves, they were not sufficient to the development of computer cartography.

The specialized hardware could perform nothing without the development of computer software (programs and data) to instruct the machines. In fairness, the software could do nothing without the hardware on which to operate. Due to variances in hardware, application, resources, philosophy and perhaps other factors there emerged as many different systems (the sum total of the hardware, software, procedures and local data) for computer cartographic applications as there were people desiring to have those functions.

Because of this process, many basic functions overlapped among these many systems. However, it is the nature of computer systems that a significant price must be paid to transfer capabilities from one system to another. Thus, there continue to be significant duplications of effort among computer cartography systems.

The geoscientists and computer scientists responsible for developing this software and the procedures for operation referred to the new field by many different names. Included are automated cartography (popular and associated with the conference in the field), computer cartography (to the point), computer-assisted cartography (more precisely correct), map data

**processing** (perhaps from the image processing contributors), and **geographic data processing** (no doubt from the geographers). These terms are considered interchangeable here, but computer cartography and computer-assisted cartography will be used.

These names generally refer to the broad area in which digital technology (computers, digitizers, graphic display terminals and plotters) is applied to the making of maps and to the preparation and analysis of data related to the production of maps or other forms of display of spatially related data.

Some generalized systems (dealing with both map data and other forms of spatial data) have been referred to as spatial data management systems, and clearly provide functions beyond those of traditional cartography and computer cartography systems (though their cartographic functions may be sub-standard). Systems which provide full data management capabilities (add, delete, edit, display, analyze, query...) on attribute data (such as depths of wells, magnitudes of earthquakes, diameters of pipelines...) which is associated with spatial data, as well as provide standard map-making functions are referred to as **geographic information systems** (GIS). These will be discussed briefly, later in this section.

A computer cartography system must include a basic, minimal set of capabilities (specific functions will be given later in this section). Among these is the ability to digitize (convert to computer form) map features from source maps to create a coherent, digital representation of the input documents. On the other end, the system must be able to produce copies which are functionally, if not literally, identical to the original input document in one or more forms. And in the middle, the system must support addition, subtraction, display and modification of the digital map features.

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.3 Types of Systems

There exist many different kinds of systems to support computer-assisted cartography. There may be as many such systems as there are kinds of mapping. Within this variety of systems, there exists one basic factor which divides the systems into two separate groups. This factor separates the two groups in every aspect of the process from the initial philosophy of the map structure to the hardware, software and procedures to perform the basic computer cartography functions.

In the **vector** approach to mapping, all map features are considered to be points and lines, where lines are defined by the sequence of points which approximate the original map line with as little error as possible. Areas are defined by their (closed) linear boundaries.

In the **raster** view of mapping, the map is defined as an ordered set of rows (rasters) of evenly spaced dots, covering the area of the original map. The dots are coded to indicate the color of the map at the dot's location, or at least to show a black or white (feature or no feature) indication.

The vector approach, similar to the view of a draftsman or a manual cartographer, employs hardware which allows the digitizer to follow lines and to digitize points along those lines. The output device, called a **vector plotter**, usually positions paper on a drum or a flat bed and was developed originally to support automated drafting. Display devices were developed from the oscilloscope or cathode ray tube technology in which a beam of electrons is magnetically aimed to draw lines on a screen. The image is retained on the screen in this **storage** (tube) technology until the user requests a fresh view.

Hardware for raster processing was borrowed directly from that of the television to produce monochrome and full color displays, and required little time to perfect. Both the digitizing and plotting hardware for raster processing required considerably more development, but have become acceptable and even preferable for many applications. Digitization involves a laser scan of color-separated overlays (which may cause some problems) and requires CPU-intensive post-scan editing which is expensive in time and money. Raster, color plotting devices come in many forms and range widely in performance and cost. Improved quality has effectively made newer color plotters more affordable for many cartographic applications, especially where other kinds of color graphics applications may also be desired.

Considering the cost of hardware and software development, acquisition, and operation for raster-based systems, one may conclude that such technology might never have evolved had it not been for the ever-increasing use of satellite imagery for map-making. Satellite imagery has been used in many ways prior to the Landsat (and other) images which have become so well-known today. Furthermore, raster processing might still be in its infancy were it not for the development of the discipline of **image processing** (the science of extracting, editing, analyzing and otherwise processing digital pictorial images by computer), which began long before mapping applications from satellite imagery were being considered.

Both the raster and vector approaches have advantages and disadvantages. One historic problem with the raster approach is the amount of space required to represent a map. The raster view of the map is one of a set of rows of dots covering the area of the map. In the raster view, dots which represent the absence of features must be maintained along with dots which do represent real map features. Thus, maps with sparsely located features require the same (large) storage space as those with very dense features.

There are also many advantages to the raster approach. One is that raster processing has (nearly) always supported a color display of map data. Such a display not only looks more pleasing and portrays map features more effectively, it greatly enhances areal editing functions and provides an in-house **color-proofing** capability (a "dress rehearsal" to check color compatibility).

Vector maps have traditionally provided clear, sharp images with excellent **registration** (matching of the separate overlays produced for the four-color printing process). The vector approach produces publication-quality maps by creating high-quality color separations for the traditional map printing process. Furthermore, maps represented in vector form store a virtual minimum of information compared to their raster counterparts. It is this quality and efficiency which have kept the vector approach going, along with its lower cost and ease of availability. Finally, it is no small factor that the more "common sense" or "natural" vector approach to the structure of map data also lends to the continued preference of the vector method over the more complex and unnatural raster approach.

There are distinct and significant disadvantages to the vector system. Among them is the storage technology for display of the map data. In the storage technology, the map is displayed in whole or in part on the storage tube and does not change until requested by the user. At that time, the screen is erased completely and the entire map is redrawn. This is very time consuming.

In raster methods, a different display screen (based on the television) is used. On these raster or **refresh** (all dots are individually and continually updated with current information) screens, individual map features may be updated as they are changed by the operator, and without erasing or redrawing any other features. This update procedure occurs continuously and is performed by the hardware of the display terminal, with update information coming from the computer and the software.

Another distinct difference between the two kinds of screens is that the raster terminal can (usually) display more than one color, as selected by the program and user. This means that different feature types may be represented as different colors and are thus easily distinguished. On the other hand, the vector terminal generally provides a single color for display of all features. Different map features must be distinguished by (time-consuming) line types, by masking out all but desired features, or by some other technique in the software.

There are many such advantages and disadvantages between the vector and raster approaches. What can be done in one system, probably can be done in the other. It may not be quite as good or as fast or as inexpensive....at least not today.

With all the systems in use today and all the kinds of applications desired for computer cartography systems, it is clear that there exists no single system that can do all things. Cartography is art as well as science, and both art and science grow and change. Largely for this reason, any large computer cartography system must inevitably be a living system, requiring constant change, correction and improvement. As long as it is to be used in a robust environment by scientists and artists, the need for new and different mapping techniques will always be present.

As a natural result of the vector/raster dichotomy, there does exist a small but increasing group which adhere to a third approach to computer cartography. This approach is the pragmatic one: take from the vector and raster approaches that which you will and that which works best for you - build a hybrid system of components from both. Of course, for some the goal has become to make a system better than both. There is even a term for the data structure or map unit in one such system: the vaster, sort of a hybrid term for a vector-raster. At a minimum, many people and systems (GIMMAP included) have added color raster display terminals and color plotters to their otherwise vector systems.

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.4 Introduction to Hardware

The hardware requirements for a computer-assisted cartography system will vary greatly depending on the applications involved. For example, there are some systems based on the microcomputer which can perform many cartographic functions. Given a modest plotter, such systems may produce their own plotted maps.

In a general sense, the hardware required for computer cartography may be related to the basic functions which are inherent in such a system. Map data must be entered into the system; it must be organized, analyzed, accessed, modified and retrieved; it must be displayed and edited; and it must be prepared and output to make maps. To support these four functions, there are four basic areas of hardware.

(1). Map data entry is accomplished through the process of map digitization. This process is the conversion of graphical map data into computer bits (representing numbers which identify locations) which is accomplished through the digitizing system. In the vector system, this process occurs on a digitizing table, an electronic table (usually) containing a grid of closely-spaced wires. A current is activated in these wires indirectly by the creation of a magnetic field in the area, in turn generated by the user positioning the digitizing cursor above the point and pushing a button. The current in the table wires is sampled and a location in (x,y) table coordinates is sent to the computer.

Digitizing in raster mode requires a more complex and more automated system in which all of the dots of each raster (row) of the map are sampled to determine the presence or absence of map features. The most common device is a rotating drum on which the map overlay (only one feature type at a time since different colors confuse the system) is placed. A laser then scans each row, recording or transmitting the sequence of bits which represent the on/off value found for each spot on the map. The transmitted sequence of values then represents the digitized map for the given overlay.

In both vector and raster digitizing, there is a terminal connected to (or in parallel to) the digitizing device for communications and to initiate operations. In vector digitizing, there may also be a graphics display terminal to show the progress of the operation. In other settings, this terminal may be used to perform interactive digitizing and corrective editing at the point of the digitizing process. Such a terminal may operate either in vector or raster mode.

(2). The display of the digital map is performed on a graphics display terminal. The many functions achieved with this display include modification or addition of map features and attributes, verification of color compatibility prior to publication, viewing and examining the displayed "electronic" map for the same purposes as one displays or examines a paper map, and preparing and editing color separations for publication.

These functions may be obtained on either a vector or raster graphics display terminal. The vector terminal draws from point to point, displaying only the map features on the storage tube. The vector terminal has no memory of the features it has drawn and can not reproduce the map image without repeated instructions from the computer. The image (usually) exists in a single color and individual features may not be updated without the entire image being refreshed by the computer.

The raster terminal displays full-color images of maps, displaying **pixels** (from "picture elements", one pixel on the display screen corresponds to a single dot in the line or raster, and displays a single color). The technology for raster terminals comes directly from that of the television, coupled with a memory to allow the computer to update parts of the image at any time. Thus, the pixels on the screen are being updated continuously from this memory, and map features or individual pixels in the memory may be modified by the computer without requiring a complete image update.

The line work of the storage tube is precise and fine, but lines on the raster terminal may look jagged like a set of steps. Edges that serve as boundaries between different colored areas may be similarly affected. This unwanted effect, known as **aliasing**, is the result of the raster representation of straight lines, often with too few pixels or too large size. The number of pixels in a row and number of rows of pixels in the screen determines the **resolution**, which is a measure of the density of pixels in a unit area. The aliasing problem may be reduced through increased resolution and through special software techniques which attempt to reduce the distractive effect by altering the colors of the pixels involved.

Along with the graphics terminal, (assumed to include a keyboard), are two additional pieces of hardware which can greatly enhance the utility of the graphics terminal. As mentioned earlier, any kind of graphic input device may be connected to the graphics terminal to allow the user to point to features and locations on the screen by moving a physical object in space.

The physical device is usually a **thumbwheel cursor** (one vertical and one horizontal wheel) or **joystick cursor** (as in an airplane), a **digitizing tablet** (a small digitizing table), or a **mouse** (a roller ball device such as those used in video games). These devices allow the user to position a screen-displayed

**crosshair cursor** (a vertical and a horizontal line segment forming a "+" to identify a unique screen location). The point selected may be used by the software to identify individual features of the map.

The second device, a **hardcopy** unit, is used to make a direct, immediate copy of the graphics terminal screen. For the monochrome display of a vector terminal, a (usually) black and white image is produced on paper. The color image of a raster screen may be copied in monochrome or in color, though in some cases the color will not be consistent with the screen.

The hardcopy is made at the push of a button or under control of the software. The (approximately) page-size copy is not an accurate or properly scaled map in most instances, but may serve as an intermediate product. As such, it is useful for checking completion and color combinations, layout and content, error documentation, and in some cases may serve as a final product. Techniques used in the monochrome hardcopy units are fairly traditional. In the color hardcopy units, techniques are essentially those used in the color plotters described below.

(3). The output function is performed by the digital plotter. The plotter produces finished maps or intermediate separations for four-color publication. There are vector plotters of various kinds (flatbed, drum, etc.) which produce line drawings as instructed by the software. The line drawings may use multiple pens or overlay plots to produce different colors for different features (as created by the software), and may vary the size of the pens for different line weights.

In vector plotting, different media may be used at varying cost to produce different effects (a two layer opaque/clear mylar called **scribecoat** is cut with jeweled pens as the first step in creating printing materials). Area coloring may be approximated by using color shading lines, created by the software. Annotation, generated through the software, may be very high quality in vector plotting, as is the other linework.

Color, raster plotters have reached a level of quality that has brought them into a useful range in many applications. Though they remain quite expensive (as do high quality vector plotters), their ability to directly plot solid areas of (user-defined) colors is highly desirable, and their increased resolution (which reduces aliasing and improves colors) has improved the line quality to acceptable levels. Color plotters are now fully capable of producing acceptable final products for many applications.

There are many technologies for producing color plots. In general, the paper is moved past a coloring head which has the ability to transfer or cause the transfer of a dot of colored ink at all selected dot locations (pixels) for a single line (raster) of the map at a time. As each new line approaches the head, new

information describing the on/off pattern for the line is given to the head. The image is treated in this fashion for all the lines of the map image and for each of the primary colors, to produce the user-mixed colors selected for the whole map.

Perhaps the best aspect of the modern color plotters and color graphics terminals is that while they do operate in raster mode, it is only an internal restriction. Both the terminals and the plotters now contain their own sizeable memories and processing capabilities, and both can convert a vector image into a raster image (usually) in reasonable time. In both cases, there are some limits on the local memory and thus on the amount of map data which can be so handled. This restriction seems to be more severe in the graphics terminals than in the plotters, but where restrictions are not too severe, the gain in functional capabilities is significant and desirable.

The significant difference between vector and raster plotting hardware has been the trade-off between the direct color of the raster system for the publication-quality linework of the vector system. In vector plotting, any map that is to be published will be of very high quality, a level not obtainable in the raster world. But, as color plotter resolution improved, the higher quality brought more and more applications into the print-only-on-demand kind of operation. Now, the vector world shares in all the improvements of the raster color plotters, while it retains the advantages of publication-quality vector plotters as well.

The fourth function of the hardware is the most central and overlaps the other three functions listed above. All the other functions and hardware built to support the other three functions all rely heavily on the hardware for this central function. As listed above, it is the organization, analysis, access, and modification of map data. This function is performed by the computer and the brains of the operators, map designers and the designers and system programmers of the computer cartography system.

The functions of the computer all require the software and usually the intervention and decision-making ability of the operators. Included in the tasks of the computer/people/software "system" are the conversion, communication, construction, modification, addition, deletion, transformation, protection, correction, recovery, organization, selection, collation, reconstruction, annotation, analysis, enhancement and acquisition of map data.

Computing systems range widely from microcomputers to main frame computers. Microcomputers perform sophisticated operations and produce surprisingly acceptable cartographic products. Peripheral equipment like digitizers and plotters now exist in smaller sizes and at reasonably small cost to provide complete

functionality (as defined above for the computer cartography system) within a microcomputer environment.

The mini- and super mini- computer systems provide much greater capabilities at much higher speeds to produce much better and more useful products - at much higher costs. Mainframe systems, costing more still, often rely on the same peripherals for digitizing, displaying and plotting as used on the smaller minicomputers. Thus, they provide a significant increase in speed and resources, and may result in much higher productivity for installations with greater throughput requirements, but generate products of no higher quality than can be obtained in a minicomputer environment for comparable costs.

There are some desired aspects for the computing system in a computer cartography environment when significant, sizeable cartographic products are to be produced, including the construction and maintenance of a large cartographic **database** (direct-access files referencing each other through the direct-access mechanism - the record number).

In these circumstances, the computing system should have a minimum memory size which can allow the computer cartography system to operate well in all extremes of the mix of computing that exists at the site. The processing speeds should be as fast as possible, and adequate secondary storage should be available for the proliferation of mapping files and cartographic databases which is bound to occur. Finally, the basic system should incorporate magnetic tape drives, a high-speed line printer and adequate terminals for all operations.

The computing system need not be a stand-alone system for computer cartography operations only, but can be a general purpose system providing services to a multiple-user community, provided its basic resources are sufficient to maintain adequate speed for an **interactive** (computer operations requiring user input and decision-making) computing environment for its users.

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.5 Advantages of Computer Cartography

In comparison with the practices of manual cartography, cartography by computer has a number of distinct advantages. These advantages may be categorized by dividing computer cartography operations into "input" and "output" operations and an area in between for "internal" operations.

On the input side, the original art work of drawing or tracing or digitizing lines from a source map is similar in manual and computer cartography. However, the manual artist is generally working directly on the final output product at that point, while the digitizing operator is creating a temporary, intermediate product.

The manual artist may make repeated corrections to the map, but eventually may cause some deterioration in the quality of the map. In the manual operation, this editing ends when the map is turned over for final production, but computer cartography editing may be repeated until complete satisfaction is obtained. Some errors made during manual compilation may require restarting the map from the beginning, but computer editing errors usually result in no more loss than the current edit session. There is no limit to the addition, deletion or modification of map data in the cartographic database.

A second advantage of computer cartography is that of accuracy in location of features. There are several ways in which a computer-assisted cartography system may be superior to manual techniques if all operations are performed properly. On the input side, the vector (manual) digitizing systems are very accurate, but can produce results no better than the quality of the work done by the people who operate them. Raster-based automatic digitizers are clearly quite accurate without concern for the operator's abilities, but can not be compared directly with manual digitizers due to the considerable difference in the nature of their products.

Additional accuracy on the input side is provided to the vector digitizer by production of very accurate plots for comparison to the original documents, by further re-digitizing, and by interactive editing techniques. Importantly, once the map features are accurately represented in the digital map, they will remain that way indefinitely or until someone deliberately modifies them.

One other advantage on the input side is that there now exists a growing set of sources of digital cartographic data (the need for which is another advantage for those with computer

cartography systems). Although the usefulness and quality of such data may be in question, it is certain these sources will become increasingly more valuable, but only to those with computer cartography systems.

There are advantages of computer cartography which are internal, somewhere between the clearly input and clearly output functions. Among these are the potential for spatial analyses of the map data by computer program, an option not available in manual cartography. Using a graphics display terminal, the digital map may become an "electronic map", useful for rapid inspection, analysis and variable displays of maps derived from the original map data by a user at the graphics terminal. Residing on secondary storage (such as disk) on the computing system, the map data is potentially accessible to a variety of users as copies of databases or directly through the cartographic databases themselves. And finally, the database itself is a new kind of cartographic product, valuable for all current and future mapping in its area.

With excellent fidelity of reproduction among them, the greatest advantages of computer cartography lie on the output side of the operations. Map data in a cartographic database is reusable without loss of quality. Reproductions of maps of equal quality to the original may be produced (plotted) in large quantity very quickly. Custom-made maps and maps derived from one or more cartographic databases may also be produced quickly. In manual methods, production of such maps may require as much effort as the original source map(s) did.

One of the best effects of computer cartography is that the map databases provide a source for an endless series of different maps which may be derived from the original map data by means of the software and with comparatively little effort. Among the variable options include coverage boundaries, sets of feature types, map projection and scale, annotation and legend content and form, line and area color mixtures, and point or line symbolization types. All these may be altered at will, and new maps may be plotted within minutes or hours.

Computer cartography systems offer the possibility of map production virtually on-demand. In advanced systems, end-users may become directly involved in the production or (at least) the design and retrieval stages of production. Production of digital files from which the database may be easily reconstructed offer the potential for transfer of digital map data between sites and for digital archival of map data as well. Finally, maps produced from digital databases may serve as base maps of high accuracy and locational control on which new map features may be drawn for digitization and eventual addition to the database.

The final item under advantages of computer cartography is the cost, not clearly an advantage per se. The cost of a computer cartography system including the special hardware for

digitizing, displaying and plotting map data, and the cost of acquiring or developing software is very high in comparison to that of manual cartography. In addition, there is the consideration of the cost of maintaining a staff to operate the system, manage the database resources, design maps, organize the operations, and to extend or correct the software. While much of the data input and editing operation may be performed by relatively unskilled and low-paid staff (e.g. students), this may be offset by the concurrent need for higher-paid, computer programmers and analysts needed to extend, document, correct and recover from errors in the system among other things.

On the face of it, cost can not be considered an advantage for computer cartography over manual cartography. In fact, the cost is clearly higher in computer cartography. On the other hand, it must be noted that the two forms of cartography are not really comparable. Computer cartography offers many valuable capabilities which are not possible in manual cartography and it produces far more valuable assets, most importantly the digital cartographic databases.

~~Take for example the fact that~~ a sophisticated computer cartography system may produce a very large set of digital databases covering a large area of interest for its organization. From then on, the system may be used to produce a map of any selected area within its bounds containing any desired set of map features in the area which are annotated and symbolized as desired and plotted in a selected projection system at nearly any scale in a matter of hours or days at most.

Providing such capability with manual methods would probably require an extremely large staff of manual cartographers, each responsible for some small area, if it could be done at all. Even if it were possible, the final map would likely be less accurate and would be more likely to contain errors than the computer cartography map.

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.6 Disadvantages of Computer Cartography

While there are many obvious (and otherwise) advantages to computer cartography over manual cartography, there are also a few disadvantages. The most prominent of these, as discussed above, is the expense of acquiring and operating the hardware and software which comprise the system. While there is clearly a cost in establishing and operating a manual cartography system, these costs are minuscule in comparison to that of just purchasing a digital plotter.

It can be argued that digital equipment is so well built today that its life expectancy probably approaches that of much of the manual cartography equipment. However, in truth digital equipment tends to become outmoded and replaced long before it ceases to function, thus widening the gap in expenses between manual and computer cartography.

As versatile and effective as computer cartography systems are, there are limits to the types and sizes of maps that they can produce. Often, the software or the hardware possess various limits in sizes which restrict the number of map elements or the physical size of the map. Digital plotters are limited in one or both dimensions, thus limiting the physical dimensions of the largest plot. Color display terminals may limit the number of features which can be displayed or the number of edges defining an area. These limits may be restrictive, but manual cartography methods have similar restrictions.

Smaller computing systems may have limitations on memory which prohibit the processing of large maps (disregarding certain techniques to be discussed later), and the low speed of operations in computers and graphics terminals may discourage or prevent computer mapping in some cases.

There are many such limitations in the hardware and software of computer cartography systems which may cause limitations in the types and sizes of the maps which can be produced. However, in many cases, methods may often be found to achieve the desired results without concern for the limits. One example is the map which contains too many features for the system. Such a map may often be produced in spite of the limitations through **physical and logical data segmentation** (subdivision by area or class), a technique which is discussed later in this chapter.

In addition, there are a number of (perhaps) smaller disadvantages which accompany the production of maps via the computer cartography system. One, another financial concern, is the overhead of operating the computer cartography shop. In the

computer cartography shop, a much greater use of paper and other materials is made to produce "intermediate" maps. These products are used for edit-completion testing, color and layout proofing, verification and interpretation of computer-generated data, and many other applications. Such operations do cost, but are not feasible in manual cartography.

Finally, there is a set of issues, almost social in nature, which arise from the use of computers and associated digital hardware for making maps. These issues are common in the application of computers to many other fields as well, where they have rarely been addressed fully.

The use of sophisticated hardware as the basis for a cartography system creates a dependency on that hardware to achieve the map-making goal. When the hardware does not work, maps can not be made. If the hardware performs in error, the operator must detect and correct the error, a process which may be more difficult than in manual methods. In practice, hardware failures occur infrequently, and hardware errors which are not grossly obvious occur even less often.

Correction of (usually operator-created) errors in computer cartography is required at different stages, and always depends on timely detection of the error. Often, there are different operators involved in different stages of map production. The operator who made the error may not be the one to correct it.

Operators do not usually follow a computer-produced map from its beginning to its end as they might in manual cartography. Without this follow-through from beginning to end, operators may lose some of the sense of accomplishment and pride which might otherwise improve error detection and other quality factors in the production of the map. Ironically, when errors remain in the final plotted map, manual cartographers may be called upon to correct them.

## I.1 INTRODUCTION

### I.1.1 Mapping By Computer

#### I.1.1.7 Geographic Information Systems

The term "geographic information system" is one which has been applied loosely and with little consistency over the past decade. This may be attributed in part to the rapidly developing field in which it applies, where truly remarkable progress has been made in a very short period of time. It may also result in part to the widely varying backgrounds of the researchers and the practitioners brought together for its development. Only recently, with the commercialization of some leading-edge systems does a more clear definition emerge.

As recently as 1980, nearly any computer-based system designed for mapping applications or analysis of geographically based information might be termed a geographical information system or GIS. The collection of some eighty systems presented as "full geographic information systems" in Calkins, H. and Marble, D. (Computer Software For Spatial Data Handling, v. 1, 1980, Int'l. Geographic Union, Commission on Geographic Data Sensing and Processing, for U.S. Department of the Interior, Geological Survey) includes many systems which could not qualify as geographic information systems by current standards. At that time, a GIS seemed to be any system in which spatial data was manipulated, and a computer cartography system such as GIMMAP might well have qualified at that time.

The thrust of the modern GIS is to handle the complete geographic "entity" in a computer-based system. The geographic entity is multi-dimensional, with a location specification in two or three dimensions (on the map or the earth's surface). Other dimensions describe properties of "attributes" consisting of numerical or descriptive values associated with the entity. The locational dimensions may also be considered as attributes of the entity and treated similarly. If desired, further complexity may be added by considering temporal changes in spatial dimensions or attributes.

Thus, both spatial and non-spatial information about geographic entities are to be built, edited, maintained, analyzed, retrieved and used together to produce maps and other significant forms of output to enhance the geographical (or geological) knowledge base.

Spatial information is simply the location of the entity. It may be in the form of latitude and longitude describing a unique position on the surface of the earth. It might also be in Cartesian coordinates (x,y) resulting from a map projection (such as UTM or Polyconic) of the earth's surface to the plane of a

map. It may exist as state plane coordinates (a disguised and fixed form of map projection), or even as a cell or raster pixel or grid node row and column number perhaps from a satellite image.

In any case, spatial information is that which is necessary to locate objects for the production of a map or other graphic image (such as on a graphics terminal) in at least its most basic form. Beyond that, spatial information is necessary for determining spatial relationships and properties (distance from, area of...) which are in turn required for analytical operations on spatial properties and even more valuable analyses involving non-spatial attributes as well. Questions such as "What percentage of the surface area of Douglas county is covered by the Lawrence formation?" may be answered through such analyses.

Non-spatial attributes greatly outnumber spatial attributes, which are restricted to location. Rivers and lakes have names, highways have designations (US, Kansas...) and numbers, railroad lines are labelled by the name of the company (Santa Fe...), counties and states all have names. Both natural and artificial geographic features have many physical properties of interest, each of which may be an attribute in a GIS. Hydrologic features have depths, widths, surface area, elevation, measurements of flow and water quality, mineral content, temperatures and so on. And all at many different locations associated with a single geographical entity. All these properties may be used in conjunction with location to perform sophisticated analysis and retrieval for map production.

The GIS incorporates both spatial and non-spatial attributes to provide analysis and retrieval functions employing a combination of both types of attributes. This basic idea is very powerful and offers a high-level interface between the user and a vast store of geographic information through the tools of a computer-cartography system joined with a database management system for the (non-spatial) attribute information. An overriding system of software blends the two parts into an integrated whole called the geographic information system. This GIS accomplishes the desired goal of treating the geographic entity as a whole, with both the spatial and non-spatial aspects brought together for complete analysis and map production.

The GIMMAP system described in this document is not (yet) a GIS, but rather is a computer cartography system in which only a minimal set of attributes is maintained. Its emphasis is on the spatial characteristics and on production of maps of publication-quality inwith great variety of form and content. This emphasis is not always found in the commercial GIS, whose primary focus is often the combination of multiple data sets for an area and manipulation of attributes as described above. In this arena, preservation of high-quality and adherence to proven cartographic principles is often secondary to the flash of new, unprecedented forms of graphics for geographic application.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.1 "G I M M A P"

The GIMMAP (pronounced jim-map) system was originally titled GIMMS for Geodata Interactive Management and Mapping System by the founder of the system Dr. Tho Trang Cao of the Bureau de Recherches Geologiques et Minieres (BRGM) of France in his stay as Visiting Research Scientist at the Kansas Geological Survey for the academic year 1977-78. During Dr. Cao's stay, it became known that another computer mapping system, predating GIMMAP also had the same acronym GIMMS. The acronym was then changed to GIMMAP, but with the same interpretation, namely Geodata Interactive Management and Mapping System. Months later, GIMMAP came to mean Geodata Interactive Management Map Analysis and Production system.

#### Geodata

The GIMMAP system was developed at the Kansas Geological Survey for the purpose of providing computer cartography to meet the growing demand for in-house research mapping and publication-quality mapping of research results at the Survey. Thus, the emphasis for GIMMAP is on geologic data, though the general approach of the system is applicable to geographic data of many kinds.

Maps have been produced with the GIMMAP system for surface geology; subsurface geology; microearthquakes and fault data; oil and gas fields, wells, pipelines and even field names; magnetic field strengths; and numerous hydrologic features both above and below the surface of the earth. Thus, GIMMAP supports geologic mapping, but it also supports most kinds of geographic mapping to provide additional background or "base" data (state and county lines, highways, hydrology, legal boundaries...) ~~to provide~~ <sup>to provide</sup> locational reference ~~for all~~ maps.

#### Interactive

The interactive nature of the GIMMAP system is emphasized by the "I" to suggest that the user of the system will sit at a terminal and react to and guide the system with on-the-spot decisions about the course to take. Many operations in GIMMAP function this way, but some require only a few initializing choices from the user and then proceed without further input. Such operations are called "batch" operations, as they operate in the absence of the user.

Interactive operations in GIMMAP include digitizing of original map data; editing of points and lines; selecting colors for areas (polygons); and creating and editing map symbology such as titles, scales and names. Still other operations (such as retrieval of map data for plotting) require large amounts of user selection in an initial phase of operation, but then complete their tasks in a batch mode.

### **Management**

One important function of the GIMMAP system is nearly invisible to the user. This is the creation and management of the four to nine files comprising the digital cartographic database for a map area. In GIMMAP, these functions are performed largely by the software (programs) in ways not always known to the user. Additional functions may be incorporated in the computing system command or "macro" language which perform many basic steps which are also hidden from the user.

Database creation occurs primarily under the control of batch programs, with initial user selections obtained via the system macro. Management of these databases, such as the nearly 2,000 comprising the Kansas Cartographic Database (KCD) at the Kansas Geological Survey, requires not only program maintenance of the internal bookkeeping records and used data records, but also requires the execution of numerous system commands under control of the macros to insure proper operations on the desired files and nothing else.

### **Map**

The map is the central focus of the GIMMAP system. From the input of raw map data to the output of plot coordinates of map symbology, special point symbols, annotated lines and areas to be filled with color, all aspects of map content and form are taken into account. The quality and content of the original data must be preserved or improved in the process so that a duplicate or better may be produced. The digital map outlives the original, and serves cartographic functions never possible with paper or mylar maps. If properly created and maintained, the "electronic" map will continue to be more useful in the future.

### **Analysis**

Map analysis within the GIMMAP system is limited to the functions required by its applications to date. There are two fundamental analyses required in building the database. The first is the analysis of the topology of the arcs to build the collection of unique node locations along with the node/arc network for construction of the Node file. Second is an analysis of paths through selected subsets of this network to generate the boundaries of the zones or polygons of the map. In addition, a number of metrics are generated through spatial analysis to provide characteristics like length, perimeter and area.

## Production

For most users, the primary function of the GIMMAP system is to produce maps, and this is the main goal of the system. As stated above, GIMMAP has been used to produce a wide variety of maps, all with publication quality whether that was required or not. The most recent application of the system was to produce a suite of maps for the state of Kansas site proposal for the new Superconducting Super Collider project.

In addition to map production, other important cartographic products must be noted. First, an increasing demand for digital data has caused GIMMAP to produce data files to be transferred directly to other users with computers and mapping needs. Still other forms of digital products include a database to support conversions between the Legal (township-range-section) system of reference and the more universal geographic (longitude-latitude) system.

The software (Legal-Geographic = LEO) used to perform these conversions is also a cartographic product, as is PROJCT, the map projection software providing conversion between geographic and cartesian (x,y) coordinates. ~~Software has become a commodity with a potential income.~~ Finally, the digital cartographic database itself, the central element to all the other products, is of great value due to the complete range of applications for which it may be used. The inevitable sophistication of users in their application techniques will easily rise to that required to use the database directly. ←

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.2 History of Development

The seeds of GIMMAP were planted in the few years <sup>immediately</sup> preceding the summer of 1977. In those few years, a graduate student working under the direction of a staff member wrote a thesis on the design of a relational database management system which was to be applied to map-making and geographic data processing (as it was then called) at the Kansas Geological Survey. Along with the thesis, the student (with the help of a staff programmer) wrote a collection of programs to support digitizing map data from USGS quadrangles.

Both the theoretical approach and the implementation of the program were wholly inadequate and unworkable for the task at hand, as was discovered by this author when replacing the graduate student in September of 1977. At nearly the same time, Dr. Tho Trang Cao had begun a one year appointment as Visiting Research Scientist, to pursue research into automated cartography which he had begun at the BRGM in Orleans, France.

By November of 1977, Dr. Cao, a mathematician, and the author, a computer scientist and mathematician, had joined forces to begin building a computer cartography system for the Survey. Dr. Cao had already produced two geologic maps in France with a set of programs designed to build a topological structure to model map data and to use that structure to create plots which could be used to make printing plates for publication.

This system worked for small data sets only, without consideration for the large, digital databases which would eventually be created. It clearly demonstrated that the structures and the ordered processes which Dr. Cao had devised were sufficient to make maps by computer. That system made GIMMAP a real possibility, and provided the basic structures and processes which are still a part of GIMMAP today.

New programs were written to build the basic procedures with the structures (modified only a little at that time) into the Survey computer, with the goal of providing map-making abilities to fit the needs of the Survey. A set of programs provided database construction, interactive graphical editing and selective retrieval for plotting. With these, maps could be entered, edited and plotted to fit the original documents. Zone (polygon) building and editing software followed, providing the capability to produce boundaries of zones of like color in separate plot files, a step necessary for publication of color maps such as the geologic maps produced by the Survey.

The first product of GIMMAP was the simplified geology of the Lawrence East (USGS) quadrangle of Kansas. The map was produced only as a single color proof to demonstrate the state of the GIMMAP system and was completed in 1978. The following year, a second project of greater complexity was published in quantity for sale at the Survey. This map, the geology of the Lawrence West quadrangle, required considerable modification to the GIMMAP system.

In 1978, Dr. Cao returned to France where efforts were made to continue the program on limited resources. Apparently, the large contingent of manual cartographers on the staff were able to restrict support for the program. Eventually, a shift was made to a large program in remote sensing, and automated cartography was abandoned entirely. During this period, the Survey (through this author) kept the French updated with new versions of GIMMAP on a regular basis. In fact, a contract for the marketing of GIMMAP was signed between the two agencies.

Since the departure of Dr. Cao, the GIMMAP system has undergone continuous growth and change. The entire system has been re-written on two separate occasions, the most recent being in 1984-5 to accomplish three tasks. These tasks were (1) to change the language from FORTRAN5 to FORTRAN77, (2) to make extensive modifications to the data structures, and (3) to incorporate (long desired) changes and additions in algorithms and options, many of which required the data structures changes to be possible.

Over this same period, the research staff (one staff and one student) also maintained all of the GIMMAP software (cumulatively re-writing the system piece by piece in that time); attempted to create and maintain enough documentation in and out of programs to be able to operate and repair, and created and trained a small staff as a production unit, which in turn applied GIMMAP to the various, increasing requirements for production mapping.

In the previous 5 years, the research staff had been forced to double as the production staff as well. In those days, production was generally restricted to a single large project at a time. Each map required development of new software, and the involvement of research staff in production was beneficial to a more complete system. Furthermore, this involvement insured proper attention was paid to development problems in production as well as to the user-interface, error-correction, modification for form as well as function and extension of the software. On the other hand, all cartographic databases require management and maintenance, responsibilities which fell on the research staff.

The result of having a production staff has been the creation of numerous valuable products. Beyond a series of published maps and in-house research maps, a large number of cartographic databases have been established. Most of these databases belong collectively to the **Kansas Cartographic Database**

(KCD), providing statewide cartographic coverage for Kansas. The KCD contains much of the information contained in the series of topographic (7.5' quadrangle) maps published at 1:24,000 by the U.S. Geological Survey. Included in the KCD are:

1. State and County lines
2. Township and Range lines
3. Section corners
4. U.S. and Kansas highways
5. Named Streams
6. Lakes of 1/4 Section or larger
7. Railroads
8. Oil and Gas fields

These features are complete insofar as they appear on the source maps. In addition, partial coverage exists for other features:

1. Elevation contours
2. Surface geology
3. Sub-surface geology
4. Unnamed Streams
5. Small lakes

In the early years, progress of the GIMMAP system was steady and easy to measure. The atmosphere was one of isolation and (as always) self-direction, with the occasional map publication or in-house completion enough indication of progress to those above. The charge - to build a computer-based system to perform Survey mapping - was general to say the least. Efforts to elicit more specific description of operation or definition of functions for the system beyond the generic "it must make maps" seemed to fall on deaf ears. This left the research staff with the freedom and the responsibility to design the entire system. Thus, it was made to be as general and unlimited as it could be.

GIMMAP began on a (Data General) NOVA 1220 uniprogramming (one user at a time) computer, whose operating system (at least) was the product of a subsidiary of the Ball jar company. This system had only 32K (about 32,000) words of 2 bytes or 16 bits each. Most microcomputers of today have 10 or more times that much memory, with processors ??????????????nearly as fast????????? In spite of this, the first GIMMAP system performed many of the same functions as it does today, including the primary functions of database construction, graphical editing, zone processing and retrieval for plotting.

More than one of the programs (GIMMAP now consists of more than forty separate programs) required nearly the limit of the memory, enforcing an efficiency consciousness and adoption of logical and physical data segmentation principles. The effects of both of these philosophies remain with the system today. The slow processor demanded almost endless attempts to improve algorithms to make programs run fast enough for true interaction.

For example, the graphics management (GRFMAN) library has passed through at least five major reconstructions or alterations to make it perform at today's acceptable speeds.

The NOVA computer was followed by a succession of larger (Data General) computers which brought multiprocessing (and the ability and problems associated with multiple simultaneous operators), timing changes in communications (and in graphics, thus requiring adjustments in interactive graphics programs), new versions of the programming language FORTRAN (which required modification of all programs, and which sometimes had its own errors to get around), virtually unlimited memory (a blessing or a curse? requiring modification to extend data limitations and considerable thought on what limits to use), a system macro language to simplify operations in the environment (but a whole new kind of software to maintain), and an ever-increasing competition for system resource (which in turn increase only long after they become scarce).

Processing on the NOVA was limited to about 500 arcs and 400 nodes at best, while current limits are arbitrarily set to allow 10,000 of each and could be higher. Algorithms in those early days sometimes relied on performing a sort on the disk. No such limitations apply today. Additional space and CPU capability often provide the opportunity to do additional functions not possible before. However, current limits have never been tested. It seems there is always the natural limit of map complexity or at least of sub-area (of the map) complexity which continues to encourage the application of data segmentation well below the current limits.

Today, the GIMMAP system runs on a Data General MV 20000 "super" minicomputer at the Kansas Geological Survey with adequate disk storage, many graphics terminals and two plotters. Though the system was originally envisioned as one which was to be operated directly by any potential user, this turned out to be not the case. As the system grew in function, it grew in complexity. Each new development required additional knowledge to properly operate the system, and frequently increased the complexity of affected components already in the system. Even before this became clear, a separation of production operations by the research staff and eventual establishment of a production arm had already occurred. As of about the fourth year, GIMMAP had become a system to be operated only by a staff of trained operators.

Many new programs have been added to the GIMMAP system as it and its computing facility grew. Among these are programs to provide map projection and deprojection (essential for correction of source map distortion and joining multiple maps together to form a larger coverage); to display areas of maps in multiple colors for color checking prior to publication, for study as an 'electronic map', or to plot the colored areas directly on an in-house color plotter; for interactive placement and editing of map

symbology (titles, names, numbers, scale bars, north arrow and legend information); and automatic annotation of features by types to produce variably-sized point symbols (+, \*, square, circle, triangle, well symbols...), line annotations (solid, dashed, hashed, double,...); and area shading with solid or dashed lines at a selected angle and spacing.

In recent studies, following up on studies made in the first three years of the GIMMAP development, consideration has been made to build onto the GIMMAP system the software required to produce a geographic information system (GIS). The basics of this idea originated with the founder of GIMMAP, Dr. Tho Trang Cao in his stay at the KGS in 1977-8. Unfortunately, the resources at the time (and in the foreseeable future) negated any possibility of such a development.

Today, the resources are basically in place for such a development. The computer cartographic functions exist in the GIMMAP system to a great extent. Acquisition of a relational database management system, now only a matter of time, will provide a set of powerful tools for manipulating attribute information, just as it was planned by Dr. Cao. The future history of the GIMMAP system then lies in this path towards the development of a GIS.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.3 A Computer-Assisted Cartography System

The GIMMAP system was developed in a time when many groups were in the process of creating computer programs and systems of programs for the general purpose of applying information science or computer science techniques to the design and production of maps. There arose a number of terms to describe the process which these people were attempting to do.

Among these terms were **computer cartography**, defined in Teicholz and Dorfman (1976 see THESIS) as "...the application of computers towards the design and production of maps"; and the term **geographic information system**, defined in Brooks and Pease (1978 see THESIS) as a "computerized system designed to store, process, and analyze spatial data"; or **geographical information system**, defined by Tomlinson (1972 see THESIS) by combining a previous definition for an information system (Dueker and Horton, 1971 see THESIS) with the provision that "the data be referenced in a manner which will allow retrieval, analysis and display [based] on spatial criteria". Tomlinson adds "...it is generally agreed that a geographical information system contains data with location identifiers, that these data are manipulated and retrieved on geographical criteria and that the output generally takes the form of graphical presentation".

It should be noted that the current concept of a geographic information system or GIS goes well beyond the two definitions of the terms most closely similar ten years ago. At that time, the treatment of the spatial aspects of geographic data was the focus of development. Now, both the spatial and non-spatial attribute information must be combined.

Other terms for this field of study and for the developing systems included , "automated cartography", "geographic data processing", "map data processing", "spatial data management systems", and "computer-assisted cartography". These terms were not strictly interchangeable, but we shall use computer-assisted cartography system to further define the functions of GIMMAP. The term "computer-assisted cartography system" most accurately reflects the true nature of the GIMMAP system. It is a computer-based system designed for cartographic application, where the computer is employed as a powerful tool to assist in the process. Most of the difficult and time-consuming work of cartography is performed by the computer (with the possible exception of the digitizing phase), with the operator making choices to guide the work toward the desired results.

Some basic functions of a computer-assisted cartography system are listed below. The GIMMAP programs which are involved in the specified activities are given below in bold for each of the listed activities.

1. Accurate data capture (digitizing) and correction to **syntax** (rules defining form and structure)

**MAPDIG** - map digitizing  
**SYNEDT** - syntax editing

2. Immediate plotting of clean (by the syntax) digitized data for gross check of digitizing errors

**DIGPLT** - digitized data plot

3. Generation of a cartographic database, the digital model of the original map

**PARGEN** - point and arc generation  
**NODGEN** - node (intersection, endpoint) generation  
**GRFCHK** - graphical (error) checking

4. Data entry via non-digitization with data in geographic (longitude-latitude) coordinates

**PROJCT** - projection (conversion to x,y)  
**SYNEDT, PARGEN, NODGEN, GRFCHK**

5. Interactive viewing and editing of cartographic data for the purpose of correcting errors in lines and points

**GRFEDT** - graphical editing

6. Graphical display of cartographic data directly selected directly from the database

**GRFEDT**  
**PLTGEN** - plot generation

7. Cartographic data update (graphic and major attribute) for the addition of new map data to existing data

re-entrant **PARGEN, NODGEN, GRFCHK**

8. Secondary database construction (**zones** - polygons) for coloring areas and preparing for color separation or color display or plotting

**ZONGEN** - zone generation

9. Secondary database construction (symbols, calligraphy) for creation, placement and editing of symbology (scale bar, title, legend...)

**CYMBAL** - symbology editing and plotting

10. Production of intermediate plots for edit-completion testing

**PLTGEN** - plot generation

11. Selective retrieval of line and point features based on location and major attribute values

**PLTGEN**

**MAPGEN** - map (plot, transfer) generation  
others

12. Enhancement of point and line features with automatic (selected) symbols and line types (double line, circle, square, triangle, hashed or dashed line...)

**MAPGEN**

13. Joining together ("quilting") of adjoining areas to form a new database or map with correct positioning and scaling of all features

**MAPGEN**

**PARGEN, NODGEN, ...**

14. Creation of symbology overlays separated by associated feature types for map production

**CYMBAL**

15. Creation of map files in original (but clean) digitized data form for transfer or archival

**DIGGEN** - digitized data generation

**ARCHIVE** - archival of map data

16. Coloring, marking and linking of areas for proper boundary selection, and selective plot creation

**ZONEDT** - zone (polygon or area) editing

17. Creation of high-quality, color-separated scribe-coats (materials used to eventually produce printing plates) for map production

**MAPGEN**  
**PLTGEN**  
**ZONEDT**  
**CYMBAL**

Many additional functions are required for a functional computer-assisted cartography system. There are many other functions of varying importance which could be listed and for which the GIMMAP program names could be given. These additional programs and functions will be covered later in this section.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.4 Philosophy of Development

*keyword* { The development of GIMMAP has gone through two distinct phases. These phases are distinguished primarily by the considerable differences ~~between the~~ computing systems of the early and late phases. The capabilities, speed and capacities changed greatly between these systems through a series of four (or more) changes in hardware. These changes all had a direct and significant effect on the execution of GIMMAP programs, and a considerable impact on the philosophy which guided the development of the system.

In the early stages, when the limits and range of functions of the mapping system were unknown, the very small computing system was adequate for the prototype development which was occurring. The system was adequate (given the high quality Xynetics flatbed plotter) to prepare color separation materials and a color proof of the Lawrence East geology. As the prototype was refined and the more realistic Geologic Map of the Lawrence West Quadrangle was prepared, the limitations of the computing system became very apparent.

These limitations were in most aspects of the computer system, but most important in the CPU-based restriction on the size of internal memory. The memory was only 32K words or 64K bytes of memory, a limit equivalent only to that of the original dinosaurs of the microcomputers which were just beginning to be marketed. But, unlike most of the subsequent microcomputers, the memory of our computing system could not be expanded.

9 ( From a programming viewpoint, the limitation in memory translated into a limitation on the number of map features which could be handled in a single map unit. This created a conflict with our view that any limitation on map features should be great enough to accommodate all mapping in the state of Kansas. The existing limits were in direct conflict with this view.

Attempts to avoid this conflict took many forms, but usually sought noble approaches which reduced memory requirements and lowered memory overhead at all costs. This valiant effort even attempted to convert an internal (in memory) sort into an external disk sort used in the NODGEN program to identify and catalogue unique node locations. This gross misuse of the computing system was rejected, though it could accomplish its goal at the expense of voluminous disk accesses, requiring very large execution time.

More successful approaches involved finding ways to get from the starting point to the same ending point without a large need for internal memory. These methods usually involved several sequential operations which required extra CPU time, but did not exceed memory limits or reasonable disk access. Unfortunately, there were times when such innovation was not found. In such places, limits had to exist.

Another significant limitation occurred as the result of the GIMMAP system operating in interactive mode for many significant parts of the map-making process. Most of these interactive operations involve the display of map data for inspection and correction by an operator. As maps grew more realistic (compared to the original simplicity of the prototypes), their display became more time consuming.

It should be noted that the display of a map produces what is perhaps the most comprehensive demand on the computing system and the software (GIMMAP). Map data is retrieved from the database via disk read operations. It is brought to internal memory where as much as possible may be stored. Significant arithmetic operations are performed by the CPU to prepare it for display. And then it is transferred to the display terminal via communication lines and is accepted, interpreted and displayed by the terminal.

Concern for the amount of time required to display maps has resulted in important changes in the GIMMAP system throughout its history. This concern arises from the fact that an operator who must wait an inordinate amount of time between editing operations will become disinterested and bored and will not perform good work. In this particular case, the unacceptable limitation on speed has been significantly altered to improve the display speed of maps. Furthermore, increases in display speed also provide an increase in the size or number of elements to be displayed, provided no other limitation further restricts that possibility.

It was clear from the start that some limitations would always exist on the amount of map data to be treated as a single unit. It would not be possible on any existing (then and now) computer system to treat all map data (even in Kansas) as a single unit. Even if memory, speed and display restrictions were erased, there are still limits on the feasibility of properly treating such a database by the operations staff. There are theoretical solutions to such problems, but they amount to nothing more than the approach we have taken. Our approach is to apply the principles of logical and physical segmentation of map data in the state of Kansas.

Another consideration in our philosophy of development was the differentiation between interactive and batch modes of operation. It is our preference to fully automate activities which could be done without operator intervention. However, when decision or judgement is required, the operator must perform the

task. Whenever artistic or other choices are preferable to some artificial default, then an operator is involved. Often times the line between these cases is not clear and has changed as technology has changed.

Often, choices in such matters were not available when the situations were first encountered, due to a lack of manpower for software development. Eventually, software would convert an unpleasant or difficult interactive task into a batch operation with very little operator involvement. One example is the syntax editing (error detection and correction) operation which originally required a programmer with tables of states, tokens and potential actions in hand to react to and correct all syntax errors at the terminal as they were detected by the software. Today, these errors are detected and corrected automatically by the SYNEDT (syntax editing) program.

Though it was not our philosophy in the beginning, we came to learn a very important truth about software development. This fact is that programs which are still being used are, in general, never finished. That is, the work which programmers do is never complete for a single program, much less for a system of programs involving a database. As long as demand for products of the single program or system exist, and as long as the programs are used, there will be work left undone. This work will be to correct errors in the software, to modify the functions or the methods of the existing software, and to extend the function of the software through creation of new programs or routines.

This work will be ever-changing even as the needs and desires of users of the software change. Thus, software is dynamic. Better yet, **software is a living entity** just as is the body of scientific knowledge itself. As the ideas of the users change, so changes the software. And as in any system, changes in one part may require or support changes elsewhere. Once enough changes (critical mass) are made, large, perhaps complete program or system rewriting must follow. The only mechanism to stop this process is complete abandonment of the software or some artificial prohibition against change.

The effect of this reality is quite important. While observers may think that the programmer is turning out one program after another like some series of children's novels, he is in fact building an inverted pyramid of related, inconstant blocks. As each block gets "finished," he can not empty his thoughts and calendar and then proceed to the next block. But rather, he has added more weight to his responsibilities since he is supporting all these blocks which came before, and each of those blocks has also increased in weight.

Each "completed" program requires less attention for a while. It's weight in the pyramid is lessened. Later, it may increase as changes to that program rise in importance. Once the changes are finished, the responsibility drops greatly. But for

most or all programs in the system, there will come a time when work is required or highly desired. And work on a single program in a system often requires work on one or more related programs. Other consequences of the living nature of software may be drawn, but are not relevant here.

Another aspect which guided the development of GIMMAP was the image of the user of the system and uses to which he would put the system. In fairness, we were as unbiased in this concern as we could be. This was not our fault. We tried to gather the opinions of the research staff of the Survey, for whom this system was intended. Unfortunately, unlike today, no one knew what they wanted of a mapping system. No help was to be found anywhere beyond the general objective: to make geologic maps for the Survey.

This vacuum allowed us to design the system with as much generality as we could. From the data structures through the completion of the basic sequence, the system was designed to meet every need and every application we were aware of. There were, of course, some possibilities which were overlooked. One idea for a mapping application was suggested by Visiting Research Scientist, Michael J. McCullagh of Nottingham University, whose contouring package for small computers provided some pleasant diversion (e.g. the development of a triangulation algorithm for contouring from a non-rectangular grid - which won the author and Dr. McCullagh an award).

Dr. McCullagh suggested the inclusion of **Z-values** (a third-dimensional or attribute value for points) in the records of isolated point arcs. His idea was that such points may be digitized and edited in GIMMAP and then presented for contouring. The idea was added to the Tomorrow file (a file of potential software correction, extension, modification and creation) and eventually incorporated in the redesign of the data structures when the change was made to FORTRAN 77.

This episode is one which suggests that the process of our initial development, effectively in isolation, worked better than it might have with user input. Today, we have incorporated the Z-values into the system in the sense that they may be entered, pass through and may be used on the map production end. Many potential editing and retrieved functions have never been developed. In fact, there has never been a legitimate use of the functions which exist and no request for those anticipated. But, the potential exists.

On the other hand, there have been many occasions where users have suggested real, current needs requiring software developments which became solid and useful parts of the GIMMAP system. These requests were fulfilled because the capabilities had (in almost every case) been anticipated in a theoretical sense and had been included on the list of desired or needed software prior to the request. The actual request was the

necessary prioritizing mechanism to trigger construction of the desired software. Most (GIMMAP) software development in recent times has been of this type due to the scarcity of programming resource.

The original work done on GIMMAP (at the Survey) was performed in an era when the watchword of the day was "user friendly." This term referred rather vaguely to the notion that users of computer software were people, not machines. In this vein, programming society adopted this phrase to set some very loose guidelines for what today may be referred to as the **user interface**. The goal was to improve programs so that they would communicate with users at appropriate times and in appropriate, or even friendly, ways.

Queries for information should be clear. Answers by users must be easy to give, and the program must be tolerant of typing and other mistakes. Explanations for options should be available. Help and Menu systems we are so familiar with today were just being created. The need for operators to memorize a list of commands, or to carry a large manual about at all times was replaced by the menu of commands which could be obtained easily from the program.

Of course abuse may arise from either extreme. A program which provides too much help or too many explanations is not very friendly. Programs must provide a minimum of help in general, but be able to produce all that is required when it is needed. The GIMMAP system today provides a Help and non-Help mode for editing operations. In the Help mode, requests and options are spelled out and selections are verified before modification actually occurs. In the non-Help mode, the operator is presumed to be an expert user who always knows what to do. Of course, the mode may be changed at any time.

One unfortunate result of this striving for the ideal user interface is the rising complexity of the software designed to simplify things for the operator. For example, in one program, routines were written to perform input, interpretation and error-checking of information from the operator terminal. Many kinds of information were to be entered through this software, and it becomes rather involved. Later, addition of a small option in part of this program turned out to be extraordinarily difficult and time-consuming due to the early "simplifications."

The desire for user-friendly software arose in the early days of GIMMAP when the final or end user of the system was expected to be a researcher desiring a map. Within a few years when the basic sequence had been worked out in its initial form, and by the time of the publication of the Oil and Gas Field Map of Kansas (Paul and others, 1982), a separate production unit had been set up.

Prior to this, the research staff had "contracted" to do only one map at a time. Each map was useful to development, because each map required development of software not previously available. These maps were produced by the research staff in conjunction with developing the GIMMAP system.

By the time the production staff was created, it was abundantly clear that the demand for computer mapping greatly exceeded the staff's ability to produce maps. This remains true today. What was then a forgotten but important element was the fact that the system had grown beyond the reach of the (what was then called) "casual user.". This important fact was true because of the complexity of the GIMMAP system which in turn was due to the unexpected complexity of the mapping process itself.

This fact does not set well for some. While the trained operator is certainly capable and hard working, why can not the researcher who is arguably more capable and equally hard working also operate the system? The answer is that, with training, the research may also learn to operate the system. But, unlike the operator, the researcher will not use the system between needs. Then, are researchers to be retrained (at no small expense) every time they need a map? With or without documentation, occasional users of the system will always require retraining unless they use the system on a very regular basis.

All the developments and philosophical viewpoints expressed above have been tied closely with the current state of the hardware - the computing system available at the time. Most of the concerns raised above arose when the computer was quite restricted. Once the system grew to virtually unlimited size, new decisions had to be made.

The new system and the acquisition of the new programming language (FORTRAN 77) provided the opportunity to increase the limits placed on the amount of map data in a single map unit. At this point, the only real objection to such a change came in consideration of the possibility of marketing or otherwise distributing the system where memory limitations may restrict or prohibit its usage.

Their concern, and the decision itself were circumvented for a while but the creation of a new GIMMAP system based on a new set of data structures, with the higher limits made possible by the new system. Then, both systems were maintained until the portability concern diminished. At some point it was felt that considering the expensive and specialized peripherals required for GIMMAP operation, anyone wishing to acquire it would probably have an adequate computing facility. The "super" minicomputer became required hardware.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.5 Functions of the GIMMAP System

The basic functions of a computer-assisted cartography system are outlined above (in "A Computer-Assisted Cartography System", Topic I.1.2.3) and most of these functions are embodied in the **basic sequence** of programs in GIMMAP. The basic sequence is a partially-ordered set of programs designed to take map data from one or more source maps to create and maintain a digital cartographic database and to generate from that database one or more high-quality map products which parallel or duplicate the original documents.

The basic sequence includes operations which usually follow a natural order defined by the state of the map in physical or digital form. These are: (1) digitizing with the **MAPDIG** program; (2) syntax correction by the **SYNEDT** program; (3) generation of the map database with the **PARGEN**, **NODGEN** and **GRFCHK** programs; (4) interactive editing with **GRFEDT**; (6) plotting to check for edit completion with **PLTGEN**; (7) database extension with **ZONGEN** and **CYMBAL** and area editing with **ZONEDT**; and (8) retrieval by feature types for final plots created by the **MAPGEN** program.

This basic sequence is performed for all maps in the GIMMAP system. The natural order of operations listed here may be altered by multiple digitizing sessions, plot checking of clean digitized data, and numerous cycles in the editing and completion plots cycles or the production of final form plots.

The first form of **extended** (database) **generation** is the automatic generation of **zones**, the areas of the map, from the information in the basic database about arcs and nodes. The second form of extended generation is the creation of **symbolology** and its associated **calligraphy** (pen strokes), the collection of names, titles, scale bars, map frames and legend information which may be added to the graphic data to enhance identification and comprehension. Extended generation involves the programs listed above, and creation of additional files (Zone, Border, Label, Symbol, and Calligraphy) in the database.

Zone generation (**ZONGEN**) is automatic (if the Node and Arc files are correct), but requires an interactive process (**ZONEDT**) to mark and color the zones and to relate **islands** (zones which are surrounded by another zone) to surrounding zones. These actions allow for proper separations of zones by colors. On the other hand, symbolology generation is performed in an interactive setting (**CYMBAL**) with the operator placing symbolology on the image of the map which is generated from the cartographic database.

Beyond extended generation is another commonly used function in GIMMAP operations. **Quilting** or concatenation of multiple, adjoining maps to form a single map product at a common scale and projection is accomplished by combining various capabilities of GIMMAP software and the uniform structure of all cartographic databases in the GIMMAP system. Given a list of names of map databases to be so combined and a set of specifications for the form of the product, a single program (**MAPGEN**) combines the features of all the named maps through the (**reprojection**) process to a common reference system, clipping off unwanted features outside a rectangular window.

The quilted product may be a set of plots for publication or other use, or may be a single file of information in the same (**digitized data file** or **DDF**) form as data from the digitizing table. In the latter form, the file may enter the basic sequence of programs and be used to create a new cartographic database covering the area of the quilted product as defined in original specifications. This secondary or **intermediate database** may have value of its own (e.g. a county map made from quadrangle maps) or may be useful for creation of symbology for the quilted area map at the specified scale.

In addition to the basic sequence, extended generation, and the quilting function there are numerous **auxiliary functions** which are performed by the GIMMAP software. All of the auxiliary functions serve an important role in creating various forms of map products, but are certainly optional and are not required for all map products. Included in these functions (with associated programs in parentheses) are the following:

- Display of colored area maps with assigned colors (**COLORZ**)
- Creation of plots with shaded area fill of zones (**SHADEZ**)
- Symbology creation directly on existing plot file (**MAPTTL**)
- Creation of selected data in DDF file from database (**DIGGEN**)
- Creation of plot/DDF file with Lat/Long lines/ticks (**TICTOC**)
- Archival of complete database in DDF file (**ARCHIVE**)
- Direct map database examination and modification (**RAFPRN**)
- Reference system conversions:
  - Geographic (Lat/Long) to Cartesian (x,y) (**PROJCT**)
  - Cartesian to Geographic conversion (**DEPROJCT**)
  - Legal (twshp./range/section) <--> Geographic (**LEOCVT**)
- Map project request and accounting system (**MAPREQUEST**)

Direct plot creation of map symbology (**SYMPLT**)  
Creation of plot file of specified rectangles (**BOXER**)  
Creation of plot of **fiducials** (registration marks) (**FIDGEN**)  
Direct examination of Hershey symbols database (**HERSHEXAM**)  
Creation of plot of complete Hershey alphabets (**HERSHPLOT**)  
Direct examination/modification of LEO database (**LEOQUERY**)  
Report feature counts for multiple map databases (**SUMMIT**)  
Creation of plot file for quilted zone marks (**ZMARKS**)

## **I.1 INTRODUCTION**

### **I.1.2 Computer Mapping With GIMMAP**

#### **I.1.2.6 Operations and Database Administration**

There are many aspects to operation within the environment of the GIMMAP system. Many things must be learned, many varied skills must be brought together. Each piece of hardware must be operated properly and every program must be understood well enough to produce the correct results for all map projects. Content and design of products must be carefully considered and the proper sequences of operations must be followed to get from source documents to the desired products. Only a subset of the GIMMAP operations will be discussed here.

Operation of GIMMAP (described here in terms of operations at the Kansas Geological Survey) is oriented toward two basic goals. The first goal is to create and maintain a comprehensive set of accurate and complete databases containing cartographic data covering the area of interest. The second goal is to use the digital cartographic databases to produce high-quality maps and other cartographic products to meet the mapping needs of the organization.

The first part of operations, the creation and maintenance of a set of cartographic databases may be viewed as a complex exercise in **database administration**, which will be developed further below. In this part, a set of programs called the **basic sequence** and other operations are used to accomplish the task. The collection of operations in the basic sequence and for other operations is:

1. **data input - or digitizing**, which is the conversion of maps in graphic form into their digital or computer form
2. **database construction** - building the complex, structured files containing the digital model from the raw digital data
3. **graphical editing** - correcting graphical errors in the database using a display of the map on a graphics terminal
4. **zone creation and zone editing** - any areas to be produced in color or with shade lines are constructed and edited
5. **quilting** - merging two or more adjacent maps into a single map through reprojection, creating maps or a database
6. **symbolology placement** - creation/editing of names, labels, scales, titles or other symbols directly on the map image

7. map design - all aspects affecting production including content, scale, colors, line and point types must be set
8. retrieval of data - map features are selected by type and separated into files by production criteria (color, width..)
9. plotting - plot files are drawn as overlays with changes in pen type/color/width on the selected media (paper, mylar)

This sequence of operations involves numerous pieces of specialized hardware including the digitizing table, computer terminal, graphics terminal, and plotter. These devices and their basic use are described in the chapter on hardware.

In the application of GIMMAP at the Kansas Geological Survey, these operations are governed by a set of special "programs" called **macros**. These are programs containing operating system commands to the computer as one might normally enter at the terminal to effect various functions. Since the same sets of the commands must be repeated frequently for the basic sequence, macros have been written to save time and to simplify (on the surface) operations.

In fact, all GIMMAP programs are instigated through use of these macros (at the Kansas Geological Survey) since they would all require repetitive and unnecessary typing otherwise. In the database construction process alone, up to five GIMMAP programs are executed through a macro called GRAND. This macro not only sets up appropriate files and other basic necessities, but also gathers the necessary answers to questions asked by the programs and passes this information to the programs as they run. Thus these five programs are executed in "batch" mode after being initiated, in the absence of the operator. Macros do require some programming effort. The GRAND macro consists of more than six hundred lines of commands and data.

At their most basic, macros require the selection of the database and selection of the current or "working" copy versus the back-up or "save" version (if one exists). This option provides a safely maintained version in a known state with the option of extending work on the working version and is highly recommended for large data sets or computing systems which suffer various unstable conditions in which work may be lost.

The name and indication of working/save (by "W" or "S") are used to locate the database, determine if it is available, and temporarily rename its files to protect against concurrent use by another operator. This and other information is appended to the "history" file of the operator along with other information pertinent to the execution of the selected GIMMAP program(s).

In the second part of the operations goals, the creation of maps and other products, the collection of basic or primary cartographic databases (e.g. county patch databases at the Kansas

Geological Survey) are quilted together to form maps or databases of adjoining areas defined by a mapping project. If databases are created, it is to perform further operations prior to map production at the new scale. Symbology placement is one such operation commonly done this way since automatic reprojection and re-scaling of symbols is not completely supported in GIMMAP.

Information is kept for project databases and on all aspects of the project both in operators written files and in a computer-based system called MAPREQUEST. This system is intended to record the many selections made for the form and content of the final map project(s) and other critical information associated with the project, its production and its requestor. A database (MPRDB - map project request database) stores the information as obtained through the MAPREQUEST program. The selections may be "frozen" or fixed in final form when all decisions are complete. Reports of the contents of each project request or a survey of general information in the database is produced on request.

Looking at the primary cartographic databases (of which there are 2,379 **county patch** databases covering Kansas for the Kansas Geological Survey) and the project or secondary databases as individual units or collectively as a single large database of cartographic information, the task of database administration is considerable. The database administrator, being either a single individual or a group, is responsible for a number of oversight functions in the cartographic database environment.

Each new database requires the usual paper accounting and recording system entries to keep track of its characteristics, uses, access limitations, content and history of operations. It also requires the creation of a set of files of a specific type. These files must be named according to a scheme which promotes easy access and provides easily recognized, unique names for all files in the system. For example, all files associated with a county patch at the Kansas Geological Survey include in their names a unique **base name** which consists of the 2-letter county abbreviation and the row and column numbers (both 2 digits) of the USGS quadrangle map in which the county patch lies. The row and column numbers are uniquely set by the imposition of an arbitrary, fixed grid of rows and columns for the state.

The files created for the database must be created with an appropriate size which, at the least, does not waste unused space. Furthermore, consideration must be given to the eventual content and, to some extent, the form of the information to be placed in the database. The source (maps, databases or other) and process of obtaining the content must also be decided and practical matters of accuracy and validity must be guarded. A set of back-up files may be created and maintained with a simple naming scheme to differentiate the two. At the Kansas Geological Survey, the original work files include the letter "W" at the end of the name for "working" files. Back-up files which may be used to restore a previous state end in "S" for "save".

The database administrator must understand the database structure, the complete set of GIMMAP files (at least those which are to be run on a given database), operational "macros" to execute various programs, and operations (including production) well enough to recover from any errors which occur. At the simplest, this requires knowing how to rename files and when such renaming is all that is required for recovery. Such errors are often system or operator generated. At other times, lengthy research and analysis is required to determine what values must be changed directly in the database. These modifications require a working knowledge of the data structures in detail as well as knowledge of the RAFPRN (Random Access File PRInt and modify) program. Errors of this kind are usually program generated, but may be complicated by subsequent operator actions.

A third kind of correction requires modification to the entire set of cartographic databases at all levels. This kind of change happens infrequently, but requires careful consideration and planning and sometimes results in very time-consuming actions. There are at least two categories of these unusual corrections, both of which are best explained by examples from the Kansas Geological Survey.

One example of unusual, system-wide database modifications occurs when some aspect of the original source or form of input data for the databases is found to be in error. One such problem occurred at the Kansas Geological Survey after most of the source (USGS) maps had been digitized for the initial phase of database construction. The problem arose about how the Lambert Conformal Conic projection was used by the USGS in their preparation of the 7.5' quadrangle maps, the primary data source.

For nearly two years, GIMMAP had software to support the Lambert Conformal Conic projection, but no information had been obtained from the USGS to indicate where the two standard parallels of the projection had been located in the construction of the 7.5' maps. In lieu of this information, the assumption was made to assume the standard parallels at the positions which minimize distortion across the map. These locations are one-sixth the distance in from the north and south edges of the map. Such an assumption was reasonable and workable until the correct positions were learned. Eventually, the actual locations were discovered and it became necessary to reproject all the data in quadrangles projected by the Lambert projection.

Another example of unusual, system-wide correction due to some error in the source map data will arise in the future. The new standard for the model of the earth (NAD83 - North American Datum 1983) will shift all data, requiring correction in all USGS quadrangle maps. The USGS is showing the new, shifted corners in relation to the old corners on new printings of the (old) maps. Fortunately, unlike the USGS which does not have all quadrangles in digital form, transformation to NAD83 should be fast and easy once the Kansas Geological Survey obtains the proper information.

The second type of universal database correction at the Kansas Geological Survey was required not by data error, but rather by modification to the **data structures** of the GIMMAP system. The data structures parallel the record content of the database files and provide the digital representation of map features in all pertinent facets. The data structures determine how programs perform their operations on map features, providing efficiency in some cases and unacceptable restrictions in others. The data structures combine the information necessary to do all possible functions in a form designed to achieve high efficiency in all effected operations.

The data structures are changed to remove restrictions on functionality, to increase speed through improved efficiency (or other technique), or to provide for the addition of needed or desired information in the database (thus increasing function). These changes allow for the addition of whole new functions and increases in speed of operation which may allow painfully slow operations to become useful and efficient. Sometimes unneeded information is removed from the data structure because operations have changed, making parts of the data structure redundant or simply unnecessary.

Changes to the data structures are made infrequently, if at all. The price that is paid for changes is probably modification of all programs which access the affected files and modification of all affected files in all databases from the old structure to the new. Such a change may or may not be easy to generate. A radical change was made in GIMMAP data structures at the Kansas Geological Survey on one occasion. A year was spent in changing all the software (many additional functions were added in that process) and all databases were converted to the new structure (with plenty of tape back-ups of everything).

Other functions of database administration generally concern the computing facility environment. The cartographic databases are stored on magnetic disk drives to provide the direct and rapid access required for GIMMAP program execution in a real, interactive environment. Space requirements must be carefully watched to assure new databases necessary space for creation and existing ones space to grow through operations. Long-term use may eventually result in fragmentation of the disk which will slow operations for all operators. The database administrator must be trained on many aspects and many levels of the computing environment and must continually inspect and protect the many resources needed for automated cartography.

Finally, one important function of database administration and operations is to **archive** cartographic databases to protect the new resource which is created by automated cartography. Archival means that a permanent back-up is made for each cartographic database and that the archived database is as current as the original was when the archival occurred. Whenever changes are made in the original (or at least at frequent,

regular intervals), the archival process is repeated. The archived database may be used to replace the original in the event it is accidentally deleted or damaged. The archived data may also be copied for transfer of the database to another site. Archival may be supported automatically through the operating system of the computer, with or without macros, or may require some operator involvement.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.7 Introduction to GIMMAP Software

The GIMMAP software described herein is a snapshot of the GIMMAP system used at the Kansas Geological Survey at the time of this writing. The software frozen for this document is also frozen on a magnetic tape. In truth, this system is not complete and undergoes a nearly continuous development. This single aspect of software, that it is **dynamic** and never really finished, is a most important and most overlooked fact.

It should be clear from the beginning that application and operation of a complex mapping system can not exist in an environment which excludes continuous software development. No program or system of programs can remain unchanged in a dynamic environment such as a computer cartography shop. Furthermore, ask an experienced programmer of complex systems if the programs are ever finished. If the answer is "YES" then the programmer probably dislikes the subject of the programs immensely, or wants to start work on something else for another reason.

Solutions to newly discovered problems and the creation of capabilities to support new applications will always require new software development. Many database administration functions, including occasional radical, system-wide modifications (such as data structures modifications, earth model changes, source map structure changes) all require software development as well as sophisticated knowledge of computer science issues. Acquisition or upgrading of any specialized hardware (digitizer, graphics terminals, plotters) will also require software development.

So the software outlined here is presented as a part of a dynamic package which is continually growing and is continually modified. Of course, for the purposes of this documentation, it is frozen as a package which does no more nor less than described (except for the expected but unintended omission). This is intended as an introduction to the software. As such, only a listing of programs (etc.) in certain groupings with very brief descriptions of function is provided. In terms of software, the GIMMAP system consists of at least five kinds of objects. These five kinds are:

1. Approximately forty **programs** written in the FORTRAN77 programming language. Programs instruct the computer to perform various tasks in the complex operations of the map-making process. Programs are separate, independent and unique entities which perform their own set of specified tasks when proper instructions are given by the operator.

2. Several **libraries** containing **routines** (partial programs) written in FORTRAN77 to provide specific functions repeated for two or more programs, saving unnecessary duplication. These libraries contain routines which contain instructions for the computer, but are not independent. The routines are grouped into libraries by functions such as map projection, database maintenance, color terminal/plotter graphics, and vector terminal graphics.

3. Three **source databases** provide information repeatedly to various programs. Each of these databases is a set of files containing important information in a structured form with interconnected pointers linking the files. The information in the files is directly accessible by programs as it is needed. The databases contain pen strokes for the Hershey alphabets, section and township (legal system) locations in geographic coordinates, and map project request information.

4. Numerous system **macros** written in the command language of the computer's operating system provide control over file, programs, and the entire operating environment. Nearly all program executions will be under the control of macros where they are available. They are used to coordinate and control cartographic databases, to sequence programs, to provide operator selections to programs automatically, to convert programs from interactive to batch mode, and to provide the expected environment for program execution while protecting against unwanted duplicate access.

5. The fifth form of object is more dynamic than the other four, and is created and often released dynamically according to user needs and project completion. This is the set of special **source files** which provide information to programs which is too lengthy to type for each execution, but which changes by project or through time. The single largest example of source files are **dictionary files** or just **dictionaries** which contain definitions of objects or attributes of objects. The most common example is the dictionary files containing color mixtures for display of colored map areas (zones) on terminals or plotters.

The programs are categorized below by general function. The **basic sequence** referred to here and elsewhere is the set of programs usually executed in a certain basic sequence to proceed from an input document map to production of essentially a copy or reproduction of the same map containing a desired subset of map features in an acceptable form. Other categories of programs refer to operations which are more or less supplemental to or supportive of the basic mapping goals.

## Basic Sequence Programs

- MAPDIG - Digitization of map data, creating the **Digitized Data Format (DDF)** file from original graphic form map
- SYNEDT - Automatic correction of **syntax** (digitizing rules and expected form of the DDF file) errors
- DIGPLT - Direct plot of clean (syntax-edited) map data from a DDF file
- PARGEN - Conversion of DDF data into the initial database form, generating the Arc, Point, and Qplot files
- NODGEN - Comparison of arc endpoints to recognize unique nodes (intersections and termination points) and generation of the Node file from the Arc file
- GRFCHK - Analysis of basic database (Arc, Node, Point, and Qplot) to locate graphical and logical errors reporting all and automatically correcting most
- GRFEDT - **Interactive graphical editing:** a process in which errors introduced into the map database in source documents or in the digitizing process may be found, examined and corrected while viewing the digital representation of the map on a graphics screen, with windowing by area and masking of features by codes, and simple pointing to identify or modify features
- PLTGEN - Generation of edit-completion plots of selected features separated for multi-color plotting to aid in checking the database for errors or completion
- MAPGEN - Creation of color-separated plots or DDF files for creation of intermediate databases, for map production with windowing, reprojection, quilting (joining of maps of adjacent areas, and automatic annotation of line and point symbol data
- ZONGEN - A "face-finding" process in which paths through the map network (embodied in the Arc and Node files) are traced to recognize the boundaries (or sequences of **border arcs**) around the **zones** (areas or polygons) and generation of the Zone and Border files
- ZONEDT - Interactive editing of zones to properly locate identifying marks (most created automatically), to assign codes signifying colors for production and to link **islands** (zones surrounded by other zones) to make topological relationships between zones exist in the database, and to create and edit **labels** (symbology linked to zones), creating the Label file

CYMBAL - The process of interactive placement and editing of map symbology directly on the digital image of the map and creation of the Symbol and Calligraphy (pen stroke) files, using display techniques similar to those for interactive editing above with windowing and masking (separate masks for features and symbols) and functions for setting and changing the font, content, angle of rotation, height, reference point and justification (left, right, center) about the reference point

### Auxiliary Display/ Plotting Programs

- BOXER - Create a plot of one or more rectangles (boxes) for map frame, legend boxes etc.
- CALPLOT(T) - Create a color plot for the Calcomp plotter of zones colored according to zone colors in the database
- CLASSY - Perform editing operations on classes of symbology by feature codes to change fonts or character heights
- COLORZ - Produce an interactive color display of zones on a color terminal based on color dictionary and windows
- FIDGEN - Generate a plot of fiducial (registration) marks at locations defined in (projected) cartesian coordinates
- FIDGENLL - Generate fiducial marks at locations specified in geographic (latitude-longitude) coordinates
- KONTIKI - Create a color display of oil and gas fields in Kansas with windowing (KONTIKIBOX) with color copies
- MAPTTL - Interactive creation and edit of map symbology for an existing plot file (not a database)
- OGNAMES - Produce a plot containing the field names from the oil and gas fields from the state of Kansas map, as reprojected to fit another map at another scale
- PALETTE - Generate a complete palette of colors described in the Cyan-Magenta-Yellow system in two parts
- POSTER - Generate a plot of point symbols from locations given with symbol type and size
- SHADEZ - Generate plots of shaded zones (areas or polygons) of a map with selectable angle and dashing pattern
- SYMGEN - Select symbology from one database, reproject it and add it to another database with height selection

- SYMPLT - Generate plots of selected symbology directly from the database, using a non-graphics terminal
- TICTOC - Create a plot file of tick marks of specified length or full lines of a latitude and longitude grid clipped to fit a neat line with optional DDF output
- X2C - Convert a Xynetics vector plot file to a raster plot file for the Calcomp with optional weeding and optional tape output for the Calcomp
- ZMARKS - Produce a plot of plus marks representing the zone marks from a set of selected databases which have been reprojected to fit a quilted map of the area

#### Generating DDF Output From the Database

- ARCHIVE - Produce an archival of all data in a cartographic database in DDF form with minimal overhead
- DIGGEN - Generate a **digitized data format (DDF)** file from an existing database with feature selection by codes
- DIGGENZONE - Generate one DDF file from a single database ordering the arcs to be consecutive for each zone in the database selected by color range or zone number

#### Reference System Conversion Programs

- LEOASCII - Produce a file on tape containing the complete set of township and section corners for Kansas from LEOBASE, the Legal system database
- DEPROJCT - Interactive or batch conversion from **projected (cartesian) coordinates (x,y)** to **geographic coordinates (latitude-longitude)** for Lambert Conformal Conic and Modified Polyconic projections
- GIMMAP.LEOCVT - General purpose (public) conversion software for reference system conversion between the Legal system in Kansas and the geographic reference system
- GIMMAP.PROJCT - General purpose (public) conversion software for conversion from geographic system locations to the cartesian coordinate system (x,y)
- LEOCVT - A single routine which converts between the **Legal reference system (township-range-section)** in Kansas and the **geographic reference system (latitude-longitude)**

LEOQUERY - Interactive interrogation of LEOBASE, a database of geographic coordinates for all Legal system corners (township and section) in Kansas

PROJCT - Interactive or batch conversion from the geographic reference system into the Legal reference system for the state of Kansas

### Utility Programs

ASORT - Rapid sort into ascending order of alpha-numeric lists stored on disk or tape files

A2FELIX - Conversion of the original form Hershey alphabets to an intermediate (Felix) form

COUNTY - Duplication of DDF output for county patches on both sides of a county line including the county line, and possibly a township or range line

HERSH - Conversion of Hershey alphabets to the final GIMMAP form from the intermediate (Felix) form

HERSHEXAM - Interactive examination of the Hershey alphabets database (two files) with coordinate locations

HERSHPLOT - Create a compact plot containing all the 23 Hershey alphabet sets numbered and ordered by keyboard keys for comparison and selection of graphic symbols

MAPREQUEST - Create, edit, maintain and report information on all map project requests and production including features, area and projection, color and other details

QSORT - Use the Quickersort technique to quickly sort a list of integers stored in a disk or tape file

RAFPRN - Direct, interactive examination and modification of nearly all values in all database files

RENEWQ - Regenerate all or a part of the Qplot (display) file from values in the Point file

SIMPLE - Simplify or generalize the interior points of a selected arc according to the Douglas or other method

SUMMIT - Prepare a summary report of the contents of a set of map databases with counts by feature codes

## GIMMAP Libraries

The GIMMAP libraries are listed below. Five libraries are listed here, but others might well be added. The LEOCVT and the GIMMAP.LEOCVT packages listed above are actually single routines which are really libraries. The GIMMAP.PROJCT library, also listed above, performs basically the same functions as the PROJCT library listed below, but with some changes for "public" use. On the other side, the libraries for color graphics and plotter operations, CALCOMP, TEKLIB and RASTER are only applicable where those components exist.

CALCOMP - The Calcomp color plotter graphics library, provided by Calcomp and modified internally, supports all graphics operations on the color plotter

GRFMAN - The Graphics Management library, written internally to provide graphics functions on the vector graphics terminal (Tektronix) and vector plotter (Xynetics) including the symbology generation software

PROMAN - The projection management library supports both projection and deprojection, the conversion between geographic and cartesian coordinates (internal)

RAFMAN - The random (direct) access file management system written internally to provide the basic accounting and organization structures on database files

RASTER - The graphics management library provided for Raster Technologies terminals by Raster Technologies, provides support for all graphic operations

TEKLIB - A library of graphics routines for Tektronix color graphics terminals from Tektronix to support all graphics functions including COLORZ and KONTIKI

## Special GIMMAP Databases

The special GIMMAP databases are listed below. Some special source data files which might be considered as databases have not been included. Among these is the LEOASCII file of ordered section and township corner locations in geographic coordinates. The LEOASCII file is a single file usually put on tape for transfer of the Legal system corner locations. The databases listed below each consist of one or more direct-access files which contain information which may be examined interactively or used by some GIMMAP program. In some cases, the database may be modified through special or routine functions.

HERSHEY - The Hershey database consists of two direct access files containing the entry pointers for each of 23 alphabets and the pen strokes for all of the characters in each of those alphabets. The Hershey database is the product of the A2FELIX and HERSH programs acting on the original Hershey alphabets file

LEOBASE - The LEOBASE database is a single direct access file containing the locations of all the section and township corners in Kansas. It is organized with a complete township of corners in each record and ordered within a township from west to east with townships ordered from north to south, supporting conversions between the geographic and Legal reference systems in Kansas with a twin database LEOASCII described above

MPRDB.KGS - The Map project Request Database for the Kansas Geological Survey contains groups of direct access records, each containing information about production for a single map project

#### The Use of Macros in GIMMAP

For nearly every program, there is a macro to control the program execution, to rename files used by the program and set other factors in the environment, to recover properly from program or operator errors and to communicate needed information to the program from the operator or elsewhere. For example, the EDIT macro renames files, changes directories and executes the GRFEDT program using the specified database with a back-up copy created by the EDIT macro to provide exit options (keep the old copy or replace it with the new one).

In some cases, a single macro controls the execution of more than one program, passing temporary file names and other control information from one program to the next. The GRAND macro at the Kansas Geological Survey controls a basic sequence of programs (SYNETD, PARGEN, NODGEN, GRFCHK, and PLTGEN) involved in the construction of the cartographic database. This macro consists of some twenty-two parts, totalling more than five hundred lines of commands and data, all written in the command language of the computer, saving the operator a great deal of unnecessary typing.

The secret of the macro is that it allows for "arguments" or variables to be passed to or entered as input to the macro. Thus, a single set of command instructions may be applied to any selected database and answers to questions asked by the programs may be entered and varied for each individual operation using the same macro to perform the operations. The full set of GIMMAP macros as used at the Kansas Geological Survey will not be detailed here, but rather are demonstrated in separate materials on operation of the GIMMAP system.

## Dictionary Files

Some programs in GIMMAP provide an option for input directly from a special "dictionary" file to provide preset controlling information or definitions of objects or attributes of objects to the program. Such input is usually considerable in size and would be required entry through the keyboard every time the program is used without a dictionary. Such information may then be placed in a sequential access file whose name may then be specified to the program in place of entering the content of the file at the terminal. Generally, dictionary files may be created and edited (to change or add entries) with any editor on the system and may, in some cases, be edited within the program which accesses them. Further, it should be noted that some of these dictionary files may actually serve more than one program.

Among the programs accessing dictionary files are:

- CALPLOT - The dictionary file used by the CALPLOT program defines each zone color in the database as a mixture of percentages of **primary colors** (cyan, magenta, yellow and black) and includes a unit size to set the coloring patterns on the plotter
- COLORZ - The dictionary file provides definitions for each of the zone colors as percentages of three primary colors (cyan, magenta and yellow) for display on the color, raster graphics terminal
- SHADEZ - The dictionary file defines a shading pattern for each of the zone colors by associating a line type (solid or dashed) and specification (dash and gap length for dashed lines) with each zone color

There exist other files in the GIMMAP system which function in a similar fashion as dictionary files. One good example is the "Names" file used by the MAPGEN program. The Names file provides to MAPGEN the list of all databases (in the form of Arc file names) which are to be used as source for the production map area. Such a file may be used repeatedly over the course of production and may contain hundreds or even thousands of entries. Having the Names file (which may be created by the operating system on many computers) saves redundant typing and time. While the Names file serves a similar function as the dictionary file, it contains a simple list of single items rather than the multiply-associated items in the definitions of a dictionary.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.8 Expandability, Portability, and Limitations

The GIMMAP (software) system (an integrated collection of programs, macros, libraries and other forms of software providing functions, and dictionary files and databases supplying source information required by the software) is a moderately large and complex system. As it exists, GIMMAP provides many basic and auxiliary functions of computer cartography in a set of programs numbering approximately forty, and relying on a set of three or more source databases. As such, the GIMMAP system today does not provide all the desirable functions in a computer cartography system.

Many additional functions might be added to the GIMMAP system due to the nature of the data structures (which define the content of the database files) which model or represent all the important aspects of the map features in the digital database. The database provides efficient access to all the information required to perform many functions not currently available in the GIMMAP system. It is the rich and (topologically) structured cartographic database which promotes the efficiency and thus the speed form most functions.

It should be noted that one limitation of the system comes from a basic philosophical decision. The limitation is that some potential speed of functions is declined as the direct result of the philosophy to never assume that programs can always store all needed data in their internal memory. That is, operations on complete data sets (such as all arcs) for a database which are to be performed internally (all data in the computer memory at one time) require some upper limit on the amount of space required. For such operations, this limit effectively imposes a limit of its own on the amount of map features in the database. GIMMAP declines to place such limits, and avoids internal operations which require all elements in memory at once.

There are many capabilities and extensions which might be desirable for the GIMMAP system as it now exists at the Kansas Geological Survey. Among these are:

1. Expansion of available map projections beyond the Lambert Conformal Conic and the Modified Polyconic.
2. Capability to edit and retrieve isolated point features based on their z-values, and to generate point symbols for plots with both magnitude and direction components.

3. Selection of map data bounded by any arbitrary polygon defined in cartesian, legal or geographic coordinates and automatic generation of county or Legal system boundaries.
4. Automatic removal of unnecessary points and reduction of points required to accurately represent linear features across scale changes or at data entry through **generalization** by the Douglas or other method.
5. Integration of symbol generation and editing tools to add computer-assisted creation and editing of map legend items and abilities beyond current (CYMBAL) to include area color fill or shading, and point or line annotation (a la MAPGEN)
6. Rapid "on-the-fly" or prepared quilting and color display of data from multiple databases providing a VIEW function of real areas for simple inspection and hard-copy
7. Various kinds of editing operations using a color raster terminal to display the map data may include some or all of the graphical editing done in vector (GRFEDT) as well as the assignment of colors and island linkage for zones

These functions and extensions may be added to GIMMAP easily by the creation of new programs and libraries or by modification of the existing programs. GIMMAP may be expanded considerably because it has been designed and developed as a collection of units or **modules**, each of which performs a specific set of functions. The modular approach has many advantages (compact and easier to understand functional units, small units which are less likely to exceed memory or other limits, a division of labor translating into operations and thus allows for specialization of skills and for parallel operations in some cases,...).

Any number of new programs may be added to the system where functions are to be increased. These may well be a sequence involved in certain operations, but rarely is there a need for simultaneity of functions which must exist in a single program. Necessary sequences in program functions rely on the results of one function being available to the next function. The key to this is the existence of the cartographic database, a dynamic central repository where each program function reacts to one state of the database to produce a new state.

!!

The system is expandable from the size of data sets which must always have some arbitrary limitations placed upon them by the computing language and environment. Limits are set on the maximum number of arcs (10,000) for node generation and editing (10,000 nodes) and for the total number of zones (10,000) in the zone editing process. Other entities in GIMMAP (interior points, border arcs for zones, **node degree** - number of arcs incident at

the node, number of points associated with an arc, or numbers of symbols and associated calligraphy records) have no theoretical limits (they are "virtually" unlimited) due to the nature of their structures in the database.

The existing limits, listed above, are arbitrary and could be raised even in their current setting at the Kansas Geological Survey or for different systems, but they do represent practical limits for reasonable operation response times by the computer and for reasonable comprehension by the operator. Simply put, the limits probably exceed the complexity of map which is best done in a single piece. At this point, segmentation of the document into two or more pieces, each to be processed separately, becomes a profitable option.

The idea of logical and physical segmentation of map data derives from the need to process maps of large size, area, content or complexity within the natural limitations of man and machine (and software). Segmentation of data provides the ability to process nearly any map (that can be plotted) by the venerable "divide-and-conquer" technique, now a mainstay of the computer science field in the form of "binary search".

Transfer of the GIMMAP system to computing environments other than the current one at the Kansas Geological Survey would no doubt require modification and adaptation. Even changing only the peripherals for digitizing, graphic edit and display, or plotting would require many changes to the software or even a complete re-writing of the affected parts since there is no real standard in computer graphics. Although portability (the ability to transfer a program or system from one computer to another without extensive modification) was a design goal of GIMMAP and a great part of our development philosophy, it became clear that we could not meet that goal in this area.

The major obstacle to portability lies in the dependence on graphics hardware from the digitizing end to the plotting end. Fortunately, there is general equivalence of function (a de facto standard) within large groups of users of each kind of equipment (e.g. digitizers or plotters). For example, most digitizers transmit the cursor location in (x,y) with one or more control characters when the cursor button is pushed. Vector terminals generally require a sequence of control characters to locate positions on the screen or to generate control functions, with (x,y) locations easily mapped to the sequence of characters and vice-versa.

Most vector terminals support one interactive cursor and several kinds of line types including temporary. Raster terminals usually provide for vector input to describe polygons, several color selection schemes, and polygon fill functions which are easily initiated. More so than vector terminals, raster terminals require their own software packages which perform all these generic functions through host subroutine calls.

At a minimum, graphics terminals will draw and move to specified points. Vector plotters and most raster plotters will do the same. At a minimum, this allows for the simplest form of map production. Other functions on the color, raster plotters generally parallel the functions of color raster terminals and are available on most color plotters.

However, translating functions from one brand of hardware to another, while generally possible, often requires significant modification, if not regeneration, of the software library or interface if it is a part of GIMMAP. Resulting changes in the calling software in GIMMAP programs using the hardware through the library or interface are also significant. If the library or interface is provided by the manufacturers, then only the GIMMAP calling software need be changed.

The significance of changes in the calling software depends strictly on the degree to which the general functions are present and the extent of parallelism in the two interpretations and implementations of those functions. At best, only routine names will be changed in the calling software. At worst, major changes in operation will require large scale re-writing of the calling software. Re-design should not be necessary.

From one standpoint, transporting the GIMMAP system to another computing system would be much less difficult. The programming language, FORTRAN77, is likely to be present in the environments where computer mapping is likely to be performed. If not present, it could probably be obtained quite easily. The Data General version of FORTRAN77 (used at the Kansas Geological Survey) adheres well (though not completely) to the standard for which it is named. Variations from that standard are documented in Data General literature (FORTRAN77 Reference Manual, Data General Corporation). Conversion to a different implementation of the standard should not require great effort. Constructs which are irregular in the Data General compiler and constructs likely to be missing from other versions (e.g. recursion) are avoided in GIMMAP wherever possible.

One final consideration to be noted is that some adjustments in the GIMMAP software may have to be made to provide reasonable execution of timed events in the interactive graphics programs. (Of course, performance in other areas may require adjustments and changes in limitations on data set sizes as well.) In the interactive graphics programs, at least three different uses are made of temporary, flashing vectors ("write-thru" for Tektronix) to briefly identify features to operators. Such displays rely on a fixed number of cycles of re-drawing vectors to produce the desired effect. The number of cycles may well require alteration due to computer, terminal speed variations and changes in the system load.

## I.1 INTRODUCTION

### I.1.2 Computer Mapping With GIMMAP

#### I.1.2.9 Hardware and Environmental Requirements

The GIMMAP system may be used to provide the functions of computer cartography in many varied settings, but there are a number of basic (mostly) hardware-oriented requirements which must be met to operate the system. Because computer cartography and GIMMAP both deal with computer graphics, certain special devices are required. These devices comprise the primary hardware restrictions on the environment for GIMMAP. Other kinds of restrictions and requirements also apply.

The basic system for GIMMAP requires a computer of modest size. Considering the trade-offs between computer memory and limits on maps (discussed earlier), this computer might even be one of today's microcomputers with their appreciable memory capacities. Of course, GIMMAP is not designed for nor currently configured for micros whose lower processing speeds might become a problem with larger map data sets. At the Kansas Geological Survey, GIMMAP is operated on a Data General MV-20000 "super" minicomputer with virtual memory. The original GIMMAP software system at the Kansas Geological Survey was developed on a Data General NOVA computer with only 64K bytes of memory, considerably less than today's micros.

The computing system for GIMMAP must have an ample store of direct access (disk) memory for program and cartographic database storage and should have an archival (secondary) storage system such as magnetic tape (perhaps floppy disks could suffice for a micro system) for databases not in current use or for back-up of databases and other files on the disk. A tape drive may also provide the best means for transfer of data and receiving data from other sources.

The computing system must also include one or more (non-graphic) terminals and a line printer to produce hardcopy output for reports from GIMMAP programs as well as program listings for programmers or system/database managers. A high-speed printer is clearly preferable in an environment in which there will be multiple operators or other competitors for computing resources. At the Kansas Geological Survey, the GIMMAP system has always existed in an environment of multiple GIMMAP operators and strong competition for resources. The system currently averages around sixty processes during business hours.

On the software side, the computing system must support the FORTRAN77 programming language, because the entire GIMMAP system is written in that language. Other variations of FORTRAN may suffice as well, but may require some modification. An operating

system which supports a directory structure for files and an easy method for compiling lists of files by name template is quite desirable. A good text editor may be used in many ways to prepare and correct program source code and data in dictionary files. The option of system macros (executable files of system commands) also simplifies and shortens the task of executing programs repetitively in an operational environment.

Beyond the computing system and the programming language requirements, there are three basic hardware components which are considered necessary for GIMMAP operation (though it might be possible to contract for two of the three). In-house digitizing will require a digitizing table which generally does not require any special controller beyond that which is built into the table itself.

On-site plotting of map products will require one of a variety of plotting devices which also (usually) come with their own controlling device. The plotter may be one of several kinds, but must accept vector map data directly, a function performed even by most raster plotters today.

The third area of required hardware is the graphics terminal used to display map data directly from the digital database. There are many kinds of graphics terminals which might be used by the GIMMAP system (Tektronix is used at the Kansas Geological Survey). However, basic operations require a graphics terminal which accepts vector map data (though it may be displayed in a raster form), and any terminal must be capable of returning a location selected by the user in vector (or vector equivalent) coordinates.

Digitizing devices come in many forms and sizes. The proper digitizer for a GIMMAP environment is one which provides four or more function buttons (though earlier versions of GIMMAP have run on systems with a single function button) and with an active digitizing area which comfortably holds the largest map sheet to be digitized. This stipulation is not strictly a requirement for GIMMAP, because large maps may be digitized in (segmented) pieces that are joined together later by the software provided they have adequate control points to govern accurate registration in a common reference system.

Digitizing tables use many different techniques to locate points, but most have a table with wires imbedded in both the x and y directions and a moveable crosshair cursor used to select points. Most tables rely on generating an electric current in a circular coil of wire (in the cursor) to create a magnetic field which is strongest at the point to be digitized (electromagnetic induction, discovered by Faraday). The magnetic field created by the cursor induces a current in the wires of the digitizing table, strongest in the wires at the center of the cursor coil which is the point being digitized. Current in the table wires is sampled to determine the x and y location of the cursor.

Digitizing tables vary in size from page or "tablet" size up to six or more feet in length. In nice models, the table surface may be backlighted for better viewing of transparent media and the surface height may be changed by a motor. Table surfaces may also tilt to any angle to improve access to all parts of the surface when digitizing large documents. All these digitizing devices have one thing in common: they transmit (x,y) locations of the cursor center when the proper button is pushed (or held for continuous "stream" digitizing).

It should be noted here that some digitizing systems provide some editing capabilities as part of a smart, turnkey system for digitizing. These systems are generally well designed and well implemented to perform the tasks for which they were intended. They can perform some of the editing functions necessary to prepare DDF input for entry to GIMMAP. However, they are not designed with GIMMAP syntax and topological constraints in mind, and can not replace editing facilities within GIMMAP. Some of these systems might well be incorporated to assist operations in a GIMMAP setting and could prove quite helpful. Digitizing in the GIMMAP system was designed to simplify the task to maximize speed and effectiveness. For this reason, no editing functions were incorporated.

Digitizing systems in general may operate in "point" mode (exactly one point digitized with each push of the cursor button) or "stream" mode (continuous digitization of points as the cursor is moved along a document with the cursor button held on), and additional control characters (selected by cursor buttons) may also be sent with each location. GIMMAP digitizing uses only the point mode of digitizing to promote careful and digitizing, to facilitate node identification and to designate the operator as the source of the **generalization** (reduction of the information required to represent features, usually by selection of a set of necessary points and rejection of others) of the map. Stream mode digitizing usually requires considerable generalization of features usually in the form of gross **weeding** or reduction of points by applying mathematical measurements with thresholds for rejection of points.

Graphics terminals are used to display map or other graphic data for such purposes as interactive examination, editing for removal of graphic errors, assignment or modification of non-graphic attributes, designation of colors for production and selection of features for production. Graphics terminals may be categorized as either **vector** (displaying features as sequences of points, addressed by cartesian coordinates parallel to those of the digitizing table) or **raster** (features are displayed as groups of "pixels" for picture elements or the smallest unit of display on a TV-like screen, where the whole picture is a matrix of cells each assigned a color whether the cell represents a feature or the background).

Vector terminals usually display features in a single color (monochrome) against a dark background. They generally use the "storage" technology in which lines drawn on the screen remain there until the entire screen is erased. Selective erasure is not possible except with the temporary vectors mentioned above. Some vector terminals are actually raster terminals in which conversion from vector to raster and raster to vector is rapid enough to be acceptable. In such systems, individual features may be erased without erasing and redrawing the entire image.

The vector terminal image is parallel to the image which is created by the digitizing table and to the digital image stored in the cartographic database in the GIMMAP system. Thus changes in the database may be directly translated to changes in the displayed image. Similarly, changes indicated by the operator at the terminal may be directly translated into appropriate changes in the cartographic database. No conversion from raster coordinates is required by GIMMAP.

Perhaps the most significant feature of vector terminals has long been their advantage over raster terminals in the resolution (the size of the smallest addressable unit for display) of the display screen. High resolution provides a clearer, sharper image which displays separate points individually rather than as a single point. Low resolution, often found in color raster display terminals, tends to produce rougher "jagged" lines and features less distinguishable. Achieving higher resolutions with color, raster terminals is possible at considerably higher cost.

Vector terminals are used for all the basic functions of the GIMMAP system and thus are required for operation of the system. Specific manufacturers are supported by existing GIMMAP software, but devices by other manufacturers are functionally equivalent and software could be adapted to support them as well. Color, raster terminals generally will provide vector **emulation** (to the operator, the terminal appears as a color vector terminal), but at a much higher cost. Color, raster technology with conversion capability between vector and raster (in reasonable time) is expensive and should not be purchased solely to provide basic functionality with GIMMAP.

Raster terminals are used in extended functions in GIMMAP where displays of map areas called **zones** (polygons) is required. Display of colored areas and colored lines has the advantage (not available with vector terminals) of clearly distinguishing the areal (and other) features of a map, and can promote effective examination (for editing or selective retrieval for production) and comprehension of the map information. Electronic examination of such maps can provide advantages over traditional paper maps by allowing scan, pan and zoom functions (with masking to select subsets of features by feature codes) usually not available on paper maps. Color, raster terminals have many attributes which place them ahead of the vector graphics terminals for displaying characteristics for map display and examination.

Drawbacks to color, raster terminals are only their cost and perhaps the screen resolution. To offset these deficiencies, they offer not only color, but also the ability to **refresh** (to continually update the display from a store which is changeable by the user, thus updating single features by themselves) the image constantly with no great cost in time. This characteristic alone could save a great deal of time in editing and other display operations, if they are done in raster mode. However, GIMMAP currently performs all editing in vector mode.

Color, raster terminals could easily be used for assigning or editing colors for areas, testing color combinations using an electronic **color proof** (trial production of a map or part of a map to test colors, content and form) of a map prior to production, and examining a colored map for errors in color selections. Of these, GIMMAP at the Kansas Geological Survey has been used to perform the latter two functions. Color assignment in GIMMAP is currently performed on the less expensive vector terminals in a procedure which is adequate for success (colors are represented by color numbers posted in areas), but which might better be adapted to color, raster terminals.

Both vector and raster terminals require software to interface the terminals with the host computer and software (GIMMAP). Such software requires modifications or parallel versions of software within the GIMMAP system. Although there is general functional equivalence among terminals of the various manufacturers, addition of specialized hardware or changes of existing peripherals will require development of interface software or (at least) modification of GIMMAP software to fit the device.

A useful and perhaps necessary companion device for both vector and raster graphics terminals is the **hardcopy** device which produces a paper (or mylar) copy of the image displayed on the terminal screen. Hardcopy devices may be provided by the manufacturer of the terminal or by others, and usually produce black on white images for vector terminals and color images for color terminals (though often not the full range available on the terminal). The hardcopy device is a very useful documentation device for graphics images especially in the editing stage, and the hardcopies may even suffice as useful map products (see the discussion below on edit and illustrative maps).

Like digitizers and graphics terminals, digital plotters are available in many varied forms, each of which provides some different plotting capabilities. The plotter's basic function is essentially the inverse of the digitizing devices, namely to convert a digital, electronic (map) image into a physical, graphic image on a medium such as paper, mylar or film. As with digitizers and terminals, plotters come in different sizes and in the two basic modes of graphics: vector and raster. (Raster digitizers do exist but were not discussed above due to their expense and hence their rarity).

Vector plotters (in forms such as flatbed, drum, roller,...) use pens to draw vectors from point to point while raster plotters produce essentially continuous rows of colored dots across the width of the image. The vector plotter translates vector data directly to pen (or paper) movements generating the vectors and produces colors by selecting pens (often by operator) of the desired colors (for lines, not areas). Raster plotters which provide color (some raster plotters produce only a grey scale, monochrome image) can plot color, raster data directly and usually include the capability to convert vector data, appended with color information, directly into lines in the raster image.

Plotters come in many different sizes, shapes, prices, and levels of quality. Plotters must be selected to perform the complete range of functions which are needed (or may be needed in the near future) for users at their installation site. In a rich environment requiring automated cartography, there are a number of significantly different kinds of map products which must be generated on the plotter if they are not to be done by manual means or farmed out for expensive production at external shop.

*check*  
The first type of map product is the "edit map". Maps in this category arise from one of several sources; have no value as a finished product; are a necessary intermediate step to test completeness, form, correctness or design; and generally are created using inexpensive plot media (pens and paper) with low priority. Edit maps, created (possibly on the hardcopy device) to check completion and correctness of digitizing sessions at the database end of GIMMAP operations are a major source of these products. Intermediate plots for all map projects made to examine the map in content and form prior to final production also supply a large source of edit maps.

A second type of map product in an automated cartography shop is the "illustration map". Illustration maps range from page size, blackline maps directly useful for publication in journals to blackline or full color maps of page size or larger, created solely to display or demonstrate cartographic (and other) information locally or to transfer results of research or study in map form to an external agent. In both of these cases, the distinguishing characteristic which classifies these products as illustration maps is that they are produced essentially as a single map or in very limited quantity: basically one-of-a-kind products for a single, limited purpose. As such, the quality of their production, from content to plot media, may vary depending on use from poor to publication quality.

A third type of map product is the "sale map", which is a high-quality, finished map produced specifically for sale. The sale map is different from the publication map (see below) in several ways. First, it is produced in-house, directly on a (probably) color plotter. As such, its technical quality is a step below maps produced by the printing press in the traditional publication process. Sale maps are usually not produced in large

quantity, but rather are produced essentially on-demand or as needed to maintain a small supply.

The nature of current plotting hardware required to produce sale maps of this kind is that the media (both paper and inks) are substandard. Furthermore, there is no clear indication that maps produced directly on these plotters will be improving enough in the near future nor that plots appropriate for production of positives for publication (mylar) will be effective. Producing sale maps on-demand or in short quantity will save in the cost of unsold maps in the long run. However, the method used to produce such maps is very much more expensive than publication methods when appropriate quantities are produced. Hence, the buyer's cost of this inferior product actually increases.

*wait*  
Producing sale maps on-demand offers the possibility of selling continuously updated, "current state" maps at all times. Of course, providing such updates implies that (1) there is some new information with which to update the map and (2) someone is continually incorporating (digitizing, appending, editing...) the new data into all databases used to support current state maps. Perhaps the only real advantage to on-demand mapping is its clear advantage in customizing maps for different users, at a price.

"Publication maps" are those produced by the traditional four-color process using printing plates to effect separation of colors by primary colors (cyan, magenta, yellow, and black) and a printing press. The automated cartography shop must provide negatives (map features are white or clear while the background is black or opaque) or positives (features are black or opaque, the background is white or clear) of the color separations for the high-quality, full-color map. Negatives are often produced as stable-base, mylar "scribecoats" in which the plotter cuts away opaque material in one layer to form lines and areal boundaries in a negative form.

Only the best of flatbed plotters are able to generate the high-quality scribes required for publication. A less accurate technique is the production of positives as regular plots on stable-base material with very good line quality. Positives may be converted to negatives (required for the creation of printing plates) through a the "contact negative" process. Scribes and positives must be produced at very high quality on stable-base material to reach publication standards. Few plotters are really capable of these standards. (One other expensive kind of plotter reaches these standards by production of negatives by using a photohead to expose film negatives directly in a vector format).

Publication maps are produced in quantity and for the purpose of sale in a significant quantity. These maps require the best of quality in content and form and are expected to last for some time. These maps may change very little over time or they may change in content and form on an annual or more frequent basis. They are usually reproduced every few years as their

stocks dwindle and as significant updates are performed. In the automated cartography environment, publication maps which are to be regenerated may actually require only moderate updates to the existing database, thus lowering the cost of publication. In some cases, however, it is preferable to redo such a map from new source documents which in turn were partly generated from the old database.

Consideration of the mix and importance of these different types of map products will help to define the needs for one or more types of plotters in the automated cartography shop. Basic vector plotters can produce edit maps and most illustration maps. Sale maps require a color, raster plotter of fairly high quality and expense depending on the need for color and the level of quality required. Plotters which can produce high-quality color separations for traditional publication maps are the top quality, flatbed plotters; expensive, vector-based photohead plotters; or very expensive, laser plotters (not discussed here). In the hierarchy of map types, a plotter which produces the higher-level maps (e.g. publication) can also provide the lower levels.

-- D R A F T --

Proposal For Development of a  
Geographic Information System  
at the Kansas Geological Survey

C. Ross, Automated Cartography

September, 1987

September 1987

K G S ... G I S

### A Geographic Information System at the KGS

The proposal is to build on the GIMMAP computer cartography system which we now have to create a higher-level system which could be used directly by geologists as well as by a trained staff to better use the vast amounts of geologic and cartographic data available to the Survey to produce not only new and better maps, but different and more effective ways to view, analyze, display, model, and otherwise utilize geologic data in Kansas.

This idea was first introduced to the Survey by Dr. Trang Cao of the B.R.G.M. in 1978 as the Visiting Research Scientist. The idea was proposed to Charlie Ross as a potential topic for his Master's thesis. Following Dr. Cao's return to France, studies were conducted by C. Ross and Joe Brentano on the feasibility of creating a relational database management system as a part of the GIMMAP system to manage attribute data associated with spatial features.

The result of two separate studies was that such a system would be large and complex, perhaps exceeding the size of the GIMMAP system for computer-assisted cartography. This conclusion arose from the fact that the computing facility at the time was restricted to 32K (approx. 32,000) computer words (of sixteen bits or 2 bytes), a restriction that was already being severely strained by the existing GIMMAP software.

It was felt that when the computing power was sufficient, such a system could be built and, connected with GIMMAP, would have a great many powerful functions. However, at that time, we were constrained by the need to produce concrete results with the existing system. The GIMMAP system had to be able to operate as a real, applied system as well as it did on paper. Thus, work on the relational, attribute-handling system was set aside.

We now have the computing power to create such a system. We certainly have the desire on many parts to create such a system, especially on the part of the administration. These are crucial pieces which were also missing in the past. We have both the knowledge and the experience in sufficient numbers to be able to do the job. We can do the job with these ingredients, given the full commitment of the Survey.

COMPONENTS OF A KGS GEOGRAPHIC INFORMATION SYSTEM

**Computer Cartography System:**

*	GIMMAP
***	+ modifications
*****	
*	+ primary extensions
*	
*	+ secondary additions
*****	
***	
*	

**Link Mechanism(s):**

*	structure
***	
*****	software
*	
*	policy
*	
*****	
***	
*	

**Attribute Handler & Spatial-Attribute Interface:**

	relational database management system (RDBMS)
	+ primary software development
	+ desired software development

**Special Applications Systems:**

	large peripheral geologic applications packages
--	---

Primary modifications and extensions, required software and the RDBMS are to be developed initially. Secondary additions and desired software will be developed in the second stage. Special applications systems are to be considered throughout development, but implemented primarily in the third stage.

### **1. Representative List of Current Cartographic Capabilities**

Computer-assisted digitization of map data  
No left-right coding or node identification for digitization  
Automatic rectification of input data to mathematical model  
Automatic topology recognition and database construction  
Interactive graphical editing of digital maps  
Automatic generation of polygon (zone) boundaries with island recognition and calculation of metrics  
Interactive color assignment preparation for printing  
Interactive placement and edit of symbology (names, scale..)  
Selective retrieval by feature type and area  
Automatic point symbolization (+,\*,circle,..,geologic etc.)  
Automatic line annotation (hashed, dashed, double...)  
Map Concatenation, joining together adjacent maps  
Clipping to restrict maps within a specified rectangle  
Generation of Latitude and Longitude grids and tick marks  
Automatic topological and other consistency checking  
Deletion, addition or modification of points and lines  
Automatic shading of selected polygons in variable patterns  
Polygon reference point, statistics and neighbor linkage  
Changes in mathematical projection and scale

### **2. Necessary Extensions and Modifications to the GIMMAP System**

Clipping to arbitrary boundary (Entity-in-polygon testing)  
Polygon update with local regeneration only  
Intact reprojection of polygonal features  
Special polygon overlay operations (union, intersect...)  
Raster/Vector/Grid conversion software  
Grid and 3D surface displays  
Generation of buffer boundary from entity  
Rubber sheeting for entry of poorly controlled map data  
Accuracy, source, date and location control mechanism  
Raster multiple map overlay functions (union, intersect..)  
Store, maintain and display in raster or other form  
Automatic reclassification of polygons (due to union..)  
Spatial operators (adjacency, proximity, connectedness, superposition, containment, generalization...)

### **3. Secondary Additions and Modifications to GIMMAP**

Additional projection/deprojection software options  
Amateur Input Package: (Depending on Target User)  
"spaghetti digitizing", automatic clean-up  
automatic node generation, edge-matching  
automatic topology generation and check  
automatic overshoot clipping

### 3. Secondary Additions and Modifications to GIMMAP (cont'd.)

- Additional geologic point symbols and line annotations
- Color/Raster annotation and symbolization (title, names..)
- Additional area selection options (legal, arbitrary polygon)
- Graphical method for selection of areas
- Linear and possibly areal generalization
- Convert map output to other system standard formats
- Create and display special graphics (scatter diagram..)

### 4. Required Development for Attribute Handler

Basic Functions on Attribute Tables: (provided by RDBMS)

- Build new table
- Edit values in a table
- Insert new attribute columns
- Add/delete entries (rows)
- Union, projection, selection of tables
- Delete, erase, purge tables
- Mix or merge non-union compatible tables
- Display, analyze table
- Retrieve values
- Prepare reports
- Retrieval of spatial features based on attribute values for one or more attributes in a single table
- Retrieve spatial features based on relational operations on attribute values in a single table
- Retrieval of attribute values based on spatial entities
- Retrieval of attribute values based on spatial relationships
- Retrieved entities to be used for display, update and plot
- Display and edit attribute values associated with selected features
- Interface/Query language bridging the query language of the RDBMS and the graphic entity side

### 5. Desired Software Development for Attribute Handler

- Retrieval of spatial features based on values of one or more attributes in one or more tables.
- Spatial feature retrieval based on multiple attributes, multiple tables, relational operations and combined spatial and attribute values and relationships.
- Retrieval of attribute values based on spatial and attribute values and spatial relationships.
- Creation of charts, graphs, diagrams from tables
- Varied analysis capabilities for attribute data
- Combine attribute and graphic restrictions to make maps
- Derive new maps/tables from others (dissolve, composite..)

**5. Desired Software Development for Attribute Handler (cont'd.)**

Retrieve map/table data via boolean operations  
Interface/Query language, additional desired functions  
Derive new maps through attribute/spatial operations on  
multiple overlays for a given area  
Pull-down menus for interface language  
Macro language for frequently combined operations  
Keyboard Template for interface language

**6. The Spatial-Attribute Link Mechanism**

In order to provide the functions outlined above, the system must clearly provide the means for the software to perform at least two basic functions:

- a. given a spatial entity, as defined by its unique record number and/or other unique spatial identifiers, locate all relevant attribute data pertaining to the spatial entity. Or at least find the attribute data which has been specified.
- b. for any entry of interest in an attribute table, find the associated spatial entity via its unique record number and/or its complete spatial description.

Accomplishment of these tasks must be not only effective, but considering the volume of data involved in mapping activities and in attributes associated with mapping, it must be quite efficient. Furthermore, the link must not be overly wasteful of other resources such as would occur in duplicating spatial data for each associated attribute reference in each table.

The link mechanism must allow independent updates changing the structure and entity configuration in the spatial databases to be reflected in all attribute tables associated with the map.

The link mechanism is tentatively suggested to be a cross-index structure, whose details are to be worked out through a combination of acquiring the RDBMS and a more precise view of the functions of the system.

## 7. Special Applications Subsystems

These are major systems in their own right which are to be adapted for incorporation to the basic GIS system. Starred items are those which already exist in some form at the Survey. Some of the items listed have greater priorities, some may not be desirable in this environment. Still others are anticipated.

- Image processing
- \*Contouring - grid
- Modeling
- Decision-making
- Site selection
- Grid Cell Analysis
- \*Proximal mapping (Thiessen polygons)
- Network analysis
- 3-D Surface
- \*Transect diagrams
- Digital terrain modeling, analysis, mapping
- Interactive Contouring
- \*Contouring - triangulation
- Planning

## 8. Entities and Structures to be Supported

- Point (Isolated, Interior, Control, Tick, Label)
- Arc (Simple line, Polygon Border)
- Zone (Polygon - neighbored, island)
- Symbols (Names, Title, Scale, Labels)
- Grid (Rectangular, regular cells)
- Raster (Full image, rows of pixels)
- Surface (3-D image)
- Raster Areas (Contiguous areas of single-valued pixels)
- Terrain Models (?)
- Tables (Attribute data)

## 9. Issues and Concerns in Development

- Public vs. Private Tables
- Public vs. Private Copies of Map Databases
- Copy vs. Direct Access
- Accuracy, Security, Standards, Control
- Access levels/ Privacy
- Documentation procedures
- Concept of Target Operator/User
- Development Resources
- Comprehension of Desired Applications
- Design considerations for internal vs. external use
- Domains and overlap of raster and vector operations

September 1987

K G S ... G I S

**10. Projected Schedule**

Primary Stage (0-18 mo.)

- 12 mo. Infant system capable of many basic functions
- 18 mo. Prototype system complete with primary functions

Secondary Stage (18-30 mo.)

- 24 mo. System redesign completed and implementation begun

Tertiary Stage (30-42 mo.)

- 36 mo. Primary and secondary system function available with some applications incorporated
- 42 mo. Most desired applications incorporated

**Computer Mapping/GIS Development at KGS**  
(A Report for the Director)

by Charles G. Ross

Mapping by computer (beyond simple contour mapping) started at the KGS under the auspices of the ARES (Automated Resource Evaluation System) program in the late 1970's. At that time, a map digitizing program was developed and focused on the 7.5' USGS topographic map series for obtaining map data in Kansas. Plotting software was developed to convert digitized data into a form for graphic display on a mechanical, Gerber flatbed plotter.

Despite state and K.U. restrictions against agencies other than central computer sites buying computers, the Survey obtained its own minicomputer, a NOVA 1220 with 32 KiloBytes (KB) of memory for a single user. This was accomplished by referencing it as a controller for the digitizing equipment. Subsequent upgrades (which were no problem) brought a multi-user system (still only 32KB per user) which gave the Survey the capability to develop local software of all kinds.

Had the Survey not acquired a computer in this fashion, it might have become a part of the state computing network to have access to digital computing machines. But, in hindsight, it seems clear that much of the Survey's current software might never have been developed, including the mapping system. The computer network has advantages for some in Topeka, but could never have provided the Survey with the capabilities required to perform the analysis, communications, monitoring, interactive processing, graphics, and mapping functions required at the Survey. In fact, a centralized, state network for any (e.g. GIS) kind of computing functions must be considered with caution and skepticism, in direct proportion to the magnitude of the promises for success and usefulness to all.

The next step in the development of computer mapping at the Survey was the initiation (still under the ARES program) of a plan to build a relational database management system (RDBMS) and couple it with mapping software (the graphical side) to provide the power of relational processing to make maps (and to finish a thesis). This system would incorporate attribute data for map features along with the graphical data which was being digitized and plotted, allowing the user to select map features by values associated in relational tables. In the end, (with degree in hand), the project never got off the ground beyond theoretical considerations. The relational part was fairly well described, but the graphical part

was beyond hope. Neither had been implemented, and both were later examined and rejected.

In 1977 and 1978, a new direction for computer mapping at the Survey was forged in the form of a partnership between the Survey staff member (working with the RDBMS etc. leftovers) and Visiting REsearch Scientist Dr. Tho Trang Cao of the Bureau des Recherches Geologiques et Minieres (BRGM), the french equivalent of the Bureau of Mines and the Geological Survey. Dr. Cao had brought with him the basic ideas of topological mapping, graph theory and a broad knowledge of the map-making process for his 1-year visit. These ideas quickly ended consideration of the previous Survey work, and the two began working in concert on computer mapping.

In France, Dr. Cao had developed software used to support the digitization of maps; a simple set of digitizing rules (the syntax) to promote simple and accurate capture of features while preserving the topology; generation of the cartographic database (or set of special, interrelated files) which acts as a model for the map; basic interactive, graphical editing; automatic recognition of zone (polygon) boundaries; and proper selection of features for the production of color-separated materials for map publication.

Dr. Cao's ideas were used as the foundation for the system (also begun under the ARES program) which has steadily grown since that time (1978) at the Survey. The system of computer software (programs, libraries, dictionaries, and databases), initially known as the GIMMS (Geodata Interactive Management and Mapping System), is now known as the GIMMAP ("jim-map") for the Geodata Interactive Management Map Analysis and Production system.

The basic functions of GIMMAP were built in the first few years of work, and have been refined and redesigned more than once since that time (e.g. at the time when the data structures were completely revised). These basic functions cover all operations needed to process a single map from input to output, including:

#### **Basic Functions of Automated Cartography**

- \* accurate and precise data capture - digitization  
(conversion of graphical data to digital form)
- \* generation and support of cartographic database  
with topology and correct modelling of the map
- \* interactive editing of graphical data with addition  
and modification and basic attribute values
- \* automatic creation of polygon boundary topology with  
ID marks, default colors, and boundary extremes

- \* interactive coloring and linking of polygons
- \* interactive placement and editing of symbology
- \* selective retrieval by feature type and location for archival, transfer and map production

In 1978, the first product, a color proof of the simplified geology of the Lawrence East (USGS) quadrangle was produced after considerable modification to the software. The scribecoats were drawn on the Xynetics flatbed plotter and prints were made at the Geography department at KU.

In 1979, the first published map, the Geologic Map, Lawrence West, Kansas, became map M-14 for the Kansas Geological Survey. For it, the GIMMAP system created only the color separations (as represented by lines in scribecoats) for the geologic formations and for the associated color boxes in the legend.

Beginning in 1979, an investigation was undertaken into the possibility of developing a relational database management system to incorporate attribute data into the digital mapping program. Research was done and designs were formulated for generating various groupings of the graphical entities (map features) to which attributes might be attached. Plans were developed to provide linkage between elementary or grouped features and the entries in relational tables, and to support retrieval of map data based on the many functions which could be applied to the tables (relational calculus).

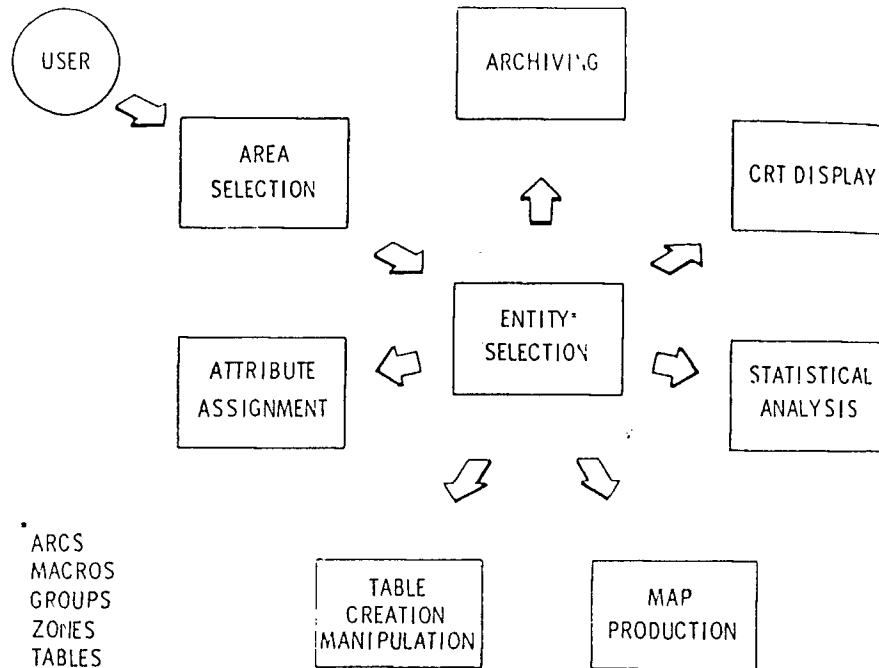
This was the beginning of the development of a Geographic Information System (GIS) at the Survey, according to the current definition (at that time the GIS definition included the GIMMAP system as it was then and now). The GIMMAP system has the graphical abilities of a GIS, but lacks many of the abilities to process attribute data and to select and manipulate map data based on qualitative and quantitative attributes associated with the (graphical) map features.

In 1979, a paper on the GIMMAP system was presented (C.Ross) at the fourth International Symposium on Cartography and Computing AUTO-CARTO IV in Reston, Virginia. Included in the paper was a statement of current developments which contained the following plan for attribute data:

"The third area of development is the construction of a relational data base system for cartographic attribute data ... providing the full capabilities of the relational calculus ... accessed with a graphical query language ... provide feature or attribute retrieval for

analysis or map production with selection by relation, attribute, attribute values, multi-relation manipulation, or statistical evaluation of graphical or attribute data"

In addition, an illustration was included in the paper which showed the overall structure of the GIMMAP system as projected with the complete ability to incorporate and utilize attribute data in conjunction with the graphical data:



By 1980, it became clear that the RDBMS and attribute handling project was not feasible at the Survey at that time because of the large requirements for (main or random-access) computer memory imposed by the RDBMS, the attribute handling activity, the linkage mechanisms between the attribute and graphical data, and the fact that both the at times the system must handle virtually all these components at a single time. At that time, our 32KB limit was being sorely tested by the graphic side software as it was. The notion of adding attributes etc. was laughable. Because of this conclusion and the press of other software development and needs for map production, the attribute/RDBMS subject was shelved, never to arise again.

Since 1980, the Survey has continued to develop the GIMMAP system as an automated cartography system, creating improved and extended graphical capabilities while ignoring the attribute side. The first production staff (students) was trained and put to work

in 1980 to build a state-wide, digital cartographic database known as the Kansas Cartographic Database or KCD. Using the USGS 7.5' topographic series maps as the most accurate source of map data in the state, the staff proceeded to capture data across the entire state as soon as coverage of the state by the source maps (USGS 7.5' quadrangle maps at 1:24000) became complete. To date, this coverage includes:

- \* state and county boundaries
- \* township and range lines
- \* section and township corners
- \* federal and state highways
- \* named streams and lakes exceeding 1/4 section
- \* principal railroad lines
- \* oil and gas fields

Current major projects to capture additional map data include:

- \* all streams on USGS 7.5' quad maps
- \* all lakes on USGS 7.5' quad maps
- \* surface geology on a county basis
- \* revised oil and gas information
- \* oil/gas pipelines

Additional map data exists in digital form for some features, but is not complete across the state:

- \* contour elevations
- \* surface geology
- \* sub-surface geology
- \* microearthquake and fault data
- \* oil/gas wells

Following the basic development period, the production staff (currently 1 permanent staff member and 3 students) has used GIMMAP and the KCD to produce maps for research and publication purposes, including one-of-a-kind maps plotted in color or black and white on the Xynetics flatbed plotter, scribed and plotted materials for publication maps, and color maps (suitable for sale) plotted directly on the Calcomp electrostatic, color plotter.

Development of GIMMAP software in this period has often been guided by the choice of maps to be produced. (Sometimes, potential map projects were selected because they required software which was to be developed next.) New kinds of maps often required new kinds of mapping techniques which must be incorporated into the software.

Many of the features developed in this period (listed below) have been developed as a result of a special map project:

### **Secondary Development in GIMMAP**

- \* projection, deprojection, reprojection
- \* selection and display for color terminal
- \* automatic point/line symbolization
- \* point/line/area color plotting
- \* geographic/legal reference system conversion
- \* geographic/cartesian clipping
- \* shading and cross-hatching of areas
- \* graduated point symbol sizes
- \* feature annotation/labelling and editing
- \* geographic gridding and tick marks
- \* map joining (concatenation)
- \* map project request and accounting
- \* direct database examination and editing

The GIMMAP system is used by the production staff to produce dozens of maps each year, with as many as fifty concurrent requests in the queue for map projects. Many maps are produced for in-house research, some for the controlled addition of new data for research or publication. The GIMMAP system has been employed to create many maps which have been published either as part of a technical report or journal article, or as a separate entity. Some examples of such maps produced at the Survey are listed:

### **Some Maps Produced by GIMMAP for Publication**

- \* Aeromagnetic Map of Kansas (KGS map M-16)
- \* Microearthquakes and Faults of Kansas
- \* Lithofacies Analysis of the Viola Limestone
- \* Historic Earthquakes, Microearthquakes, and Faults of the Kansas-Nebraska-Iowa-Missouri Region
- \* High Plains Aquifer, Kansas
- \* Geologic Map, Lawrence, Kansas (KGS map M-14)
- \* Water Quality and Basin Areas in Kansas
- \* The Geology of Allen County, Kansas
- \* Pre-Development (1950) Water-Table Elevation in GMD 4
- \* Generalized Geology, SSC Proposal
- \* Detailed Geology (North), SSC Proposal
- \* Detailed Geology (South), SSC Proposal
- \* Potential Radon Hazard Map
- \* Recent Drilling Trends in Kansas
- \* Structural Contour Map, SSC Proposal
- \* Location Map, SSC Proposal
- \* Oil and Gas Fields of Kansas (KGS map M-17)
- \* Saturated Thickness in G.M.D. No. 4

As early as 1980, recognition was made of the fact that the expanding digital database (the KCD) could be used not only as a source of information for map production, but also as a valuable source of direct information. That is, the collection of map information is valuable in digital form to anyone with knowledge and the computing ability to use it directly for map-making and analysis. However, it was not until 1982 that digital data began flowing from the Survey to both state (KDOT, KU...) and federal (USGS...) agencies, to companies in Kansas, and to entities in foreign countries. Today, there exists an expanding interest in the cartographic data we have and in the data we will produce in the future.

In 1988, there remains a good deal of work to be done to extend the graphic capabilities of the GIMMAP system (as planned) and to modernize the entire system in accordance with the changed needs and resources which are now available. There is a need for publication-quality map data and a growing market for products that can come from it. There is still a need for production of one-time-only, research maps for internal use, and an external need for maps produced for sale to outside individuals. These are all a part of the justification for the original development of GIMMAP.

Today, we are at the crossroads where internal development of a computer cartography system called GIMMAP meets the inevitable acquisition of a commercial software system (ARC/INFO) which combines the graphical and attribute capabilities in a single GIS package, obtained at a price that seems to defy all reason.

On one side, we have a system which has been built over a long period that comes with experience, knowledge and a proven ability to perform the desired functions and to grow with needs. We can proceed to extend and modernize the graphic side while adding the ability to incorporate and use attribute data as a true GIS. We can put the system on the new microcomputers for many existing and new operations. We can continue to support internal research and development, from which can reap many rewards (see below). One of the principal reasons GIMMAP was built was to seek the frontier in computer mapping - to do the research.

The Survey is, after all, a research organization, and there are many good reasons to support internal development of software over commercial acquisition:

#### **Advantages of In-House Software**

1. Designed for local or external needs
2. Can be modified locally as needed
3. Can be corrected quickly, at will
4. Can be expanded locally for new applications
5. Documentation to desired levels, easily changed

6. Introductory education locally at minimal cost
7. System oracle (problem-solver) is in-house
8. Can be done at lower cost over the long term
9. May distribute for \$\$\$, glory, or ...
10. May be designed to take advantage of local hardware
11. Outside software may cost for each unit of hardware
12. Transfers to/from outside may be tailor-made
13. Today, we buy research in computer science Tomorrow, we may buy geologic research. Shall we dismiss our staff, and hire agents to buy the geologic knowledge that Kansans need?
14. Money spent on commercial software usually goes out of the state, while money for in-house software stays in the state.

Some are concerned by the time required for development of in-house software. Given advantages such as those above, some time could be tolerated. But, how long will it really take for those who use ARC/INFO to actually do what they only dream of now? It is not unusual that an ARC/INFO installation has not gained the ability to perform desired tasks even a year after the acquisition. With such a complex system, a trained staff will be needed (as predicted) for researchers who have tasks other than ARC/INFO to perform.

It is important to note that unlike virtually any other successful mapping system, the GIMMAP system has been developed with a very small staff (on average only 1 1/2 members, including students) on very low state salaries. And, in the initial years, while the basic system we being developed, the research staff was also required to double as the production staff. Having a small staff (still in existence) has not prevented the development of a very useful system in a reasonably short time.

In 1987, it was suggested that the Survey obtain a relational database management system to help researchers manage the many sets of data residing in the Survey. It was also noted that acquisition of an RDBMS could make possible the addition of attribute handling in the GIMMAP system on the current "super" minicomputer. At that time, the researcher involved in the 1979-80 investigation into potential development of these features suggested that the RDBMS be written in-house by himself and a second staff member. The latter scoffed at the idea at the time, but recently concluded that the job could be done in-house and would probably have produced at least a prototype system by now. No new database management system of any kind has been acquired.

The alternative to the above development option is not really an option. It is a choice in how ARC/INFO will be used at the Survey. The Survey will get ARC/INFO and must decide how and where it is to be used or not used. Many questions arise.

Will ARC/INFO operate in parallel with GIMMAP as forecast?

Will a staff of (ARC/INFO) operators replace the staff of researchers developing original work?

Will individuals learn to use ARC/INFO or will a trained staff be required to operate it for everyone?

Can GIMMAP be used for some phases of mapping while ARC/INFO is used for the rest?

Can the two co-exist side-by-side with overlapping functions?

Can GIS development occur in the GIMMAP system even though ARC/INFO is present?

These and many other questions face us at these crossroads. Perhaps the easy move would be to join the bandwagon and opt for participation in a statewide GIS network. But such a move is contrary to the research mandate which governs the Survey, and a network of GIS capabilities would serve our needs less well than a computing network might have in the past ten years.

Certainly, the Survey has more than the required hardware and the advantage of relative isolation with the ability to dedicate an appropriate staff to the task at hand. This gives the Survey the ability to do the job - to create a successful GIS from the existing GIMMAP system which is tailored to the needs of the Survey, in a reasonable period of time.