Geographic Information Systems

AN INTRODUCTION

JEFFREY STAR
JOHN ESTES

GEOGRAPHIC INFORMATION SYSTEMS
An Introduction

JEFFREY STAR
JOHN ESTES

University of California, Santa Barbara

PRENTICE HALL, Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Star, Jeffrey.
Geographic information systems : an introduction / Jeffrey Star,
John Estes.
p. cm.
Bibliography: p.
Includes index.
ISBN 0-13-351123-5
1. Geography—Data processing. I. Estes, J. E. II. Title.
G70.2.S73 1990 89-33220
910'.285—dc20 CIP

Editorial/production supervision: Christina Ferrari
Cover design: Photo Plus Art
Manufacturing buyer: Paula Massenaro
Cover photograph: California Condor database. Photo by Jeffrey Star.
Color insert layout: Thomas Nery

A Paramount Communications Company
Englewood Cliffs, New Jersey 07632

Printed in the United States of America

20 19 18 17 16 15 14 13 12 11

ISBN 0-13-351123-5

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

To Toni and Claire

Preface

This book was written for several reasons. First, we hope it helps to fill a gap by providing an entry into the literature of geographic information systems. Up until very recently, most of the information about geographic information systems (GISs) has been in the fugitive, or *gray*, literature; the latter includes in-house technical reports, conference proceedings, non-refereed publications, and poorly distributed manuscripts describing efforts funded by federal and state agencies. Publications from professional conferences, such as Purdue University's "Machine Processing of Remotely Sensed Data" Symposia, the Harvard Computer Graphics Symposia, and the Urban and Regional Information Science Association meetings have been some of the best places to find current information about GISs, in terms of both the theoretical and the practical aspects of the field.

Special-purpose documents have also been useful sources, for those of us who know where to look. Briefing documents prepared by such agencies as the U.S. Forest Service and companies such as the Environmental Systems Research Institute, and surveys of the field by the American Farmland Trust, have been valuable resources in their respective narrow domains. Publications such as the American Society of Photogrammetry and Remote Sensing's *Photogrammetric Engineering and Remote Sensing* also have articles about GISs on a regular basis. Recently, journals have been started that emphasize spatial data processing. Two of these new journals are the *International Journal of Geographical Information Systems*, published quarterly by Taylor and Francis in the U.K., and *Aerial and Space Imaging, Remote Sensing and Integrated Geographical Systems*, published quarterly by Imaging Systems Publications in Sweden. Recent conferences have focused specifically on geographic information systems. In particular, the proceedings of GIS '87 (October, 1987; San Francisco) and IGIS '87 (November, 1987; Crystal City, Virginia) provide a summary of the state of the field, although primarily from an academician's point of view.

Information about geographic information systems may sometimes be found in one chapter of a book - see, for example, Short's *Landsat Tutorial Workbook* (1982), a classic volume which is unfortunately out of print, or the chapter edited by Marble in the second edition of the *Manual of Remote Sensing* (American Society of Photogrammetry and Remote Sensing, 1983). Recently, a monograph has been published that focuses completely on GIS's. Burrough's *Principles of Geographical Information Systems for Earth Resources Assessment* (1986) is an excellent resource in this rapidly emerging field. For information on a variety of aspects of GIS, see Marble, Calkins, and Peuquet (1984) - a self-published three ring binder of readings on GIS that has also become a well-known standard reference.

A second important function for this text is to serve as an introduction to the field, but at the upper-division undergraduate or beginning graduate level. Much of the material that follows has been used in short courses we have presented under the auspices of George Washington University and elsewhere. We hope it serves as a text for a one-semester course focusing on geographic information systems and their applications, as well as a resource for individual study. A student with some background in computer graphics, cartography, and geographic theory should find this text accessible.

A final reason for this book's existence is to raise the banner of *data integration* as a philosophical basis for modern GIS's. From one perspective, the essence of a geographic information system is its integration of the many different kinds of information we may be able to obtain about the spatial objects in an area. In other words, a user of a GIS should be able to work with many different types of data, bringing *all* the information relevant to a problem together in a consistent form, and should be able to bring to bear on the problem all the power of all the sophisticated analysis tools available. By the same token, we believe a geographic information system is the correct tool for integrating the different technologies that are used in gathering, analyzing, and assessing spatial data.

We begin with a chapter that introduces some of the jargon of geographic information systems. It does so by presenting a completely non-digital land-use planning problem: the design of a golf course. This is followed by a chapter on the history of geographic information systems, and by another that introduces the essential components of a GIS. Chapter 4 discusses data structures: how a GIS actually stores the kinds of data and information that are commonly used in systems.

In the next five chapters we explore the five essential parts of a GIS: data acquisition, preprocessing, database management, analysis and manipulation, and producing the final output. In these chapters we attempt to show the practical use of the different tools and techniques.

Chapter 10 examines the relationship between the technology of remote sensing and the philosophy of geographic information systems. From a theoretical point of view, the image processing systems used when working with remotely sensed data are geographic information systems, although with strengths in certain kinds of numerical operations and presentation functions. Based on our experience, we believe that remotely sensed data and GIS's are natural allies. In this chapter we have tried to highlight the interdependencies of these two areas, and indicate the potential each has for the other.

Chapter 11 briefly discusses some practical matters. Bringing a digital geographic information system into an organization requires changing the way the organization conducts its day-to-day operations. This affects an organization's management in several ways, including short-term disruptions and costs during the transition to a digital system, as well as changes to the long-term planning process. A GIS makes new demands on staff and staff development, and affects funding requirements in both the short and long term. Some of these are the same as the impacts of any other digital information system, and thus, there are some relevant guidelines and case studies. Problems specific to a GIS, in terms of estimating database volume, developing specifications, and managing the overall project, are discussed briefly.

In Chapter 12 we present a potpourri of applications, drawn from friends and colleagues who were able to contribute material to us. These applications cover a wide range, from the design and management of large facilities to long-range research on the global carbon cycle. The applications also touch on the wide range of hardware and capabilities now used to solve spatial data processing problems.

Finally, in the last chapter we gather our thoughts about the future. Shakespeare's Macbeth had perfect counsel about future events from three witches. We claim no such accuracy, and hope our prophecies are less disastrous. However, there are visible trends in the GIS industry, as well as interesting new developments coming out of the research laboratories. New data acquisition technologies are being developed and data continues to be less expensive to process and store, while our appetite for processing continues to grow. We report on some of these developments, both to describe the current state of the art and to show some of the directions we believe the state of the practice will take.

Several color plates are provided to help illustrate specific topics presented in the text. They are:

1. Aerial photograph: color infrared
2. Master planning
3. Proposed dam site placement
4. Waste site selection
5. African water resources
6. Raster/vector data for map update
7. California Condor database
8. Agricultural production modeling

For contributions of applications and figures, we thank John Jensen (University of South Carolina), Bruce Rado and Lawrie Jordan (ERDAS Inc., Atlanta, Georgia), S.J. Camarata (Environmental Systems Research Institute Inc., Redlands, California), and Heidi Schweikart (Intergraph Corp., Huntsville, Alabama).

We gratefully acknowledge the help and encouragement of many of our colleagues - in particular, Larry Carver, Frank Davis, Ellen Knapp, Terry Smith, Larry Tinney, and Waldo Tobler. Several of our students and staff - particularly Key Ho Bahk, Ken McGwire, Terri Morris, Jeff Sandberg, and Joe Scepan - have made valuable contributions to this effort. We particularly thank Hugh Calkins, Dave Cowen, and John Jensen for their careful reading of portions of the manuscript. All remaining errors, of course, are ours alone.

About the Authors

Jeffrey L. Star is a researcher and science manager in the Geography Department at the University of California, Santa Barbara. His Ph.D. is from the Scripps Institution of Oceanography of the University of California. As a lecturer he has taught many students at both undergraduate and graduate levels. In addition he is widely known for intensive workshops in GIS for agencies such as the Agency for International Development, the Defense Mapping Agency Aerospace Center, George Washington University, and the Stennis Space Center. He has served as a consultant and advisor to numerous organizations both public and private, including the United Nations, TRW, NASA, and the National Bureau of Standards.

John E. Estes is professor of Geography and Director of the Remote Sensing Research Unit, University of California, Santa Barbara. He received his Ph.D. from the University of California, Los Angeles. Included in a varied career are duties as Co-Director of the National Center for Geographic Information and Analysis, Visiting Senior Scientist at NASA Headquarters, and numerous roles as consultant to industry. He is a member of the National Academy of Science's Mapping Sciences Committee. His research has been funded by NASA, the U.S. Geological Survey, and the U.S. Department of Agriculture, among others.

Brief Contents

1 Introduction 1

2 Background and History 14

3 The Essential Elements of a GIS: An Overview 24

4 Data Structures 32

5 Data Acquisition 61

6 Preprocessing 76

7 Data Management 126

8 Manipulation and Analysis 143

9 Product Generation 174

10 Remote Sensing and GIS 191

11 Practical Matters 220

12 Applications 230

13 Looking Toward the Future 245

Suggestions for Exercises 252
Glossary and Abbreviations 261
Bibliography 283
Index 297

Contents

Preface xi

1 Introduction 1

1.1 Geography 1
1.2 Information Systems 2
1.3 A Manual Geographic Information System 4
1.4 Applications 8
1.5 Geographical Concepts 9
1.6 The Four Ms 12

2 Background and History 14

2.1 The Cartographic Process 15
2.2 Early History 17
2.3 The Modern Era 19

3 The Essential Elements of a GIS: An Overview 24

3.1 Functional Elements 24
3.2 Data in a GIS 28

4 Data Structures 32

4.1 Raster Data Structures 33
4.2 Vector Data Structures 48
4.3 Comparisons between Data Structures 57

5 **Data Acquisition** **61**

5.1 Introduction 61
5.2 Existing Datasets 64
5.3 Developing your own Data 70

6 **Preprocessing** **76**

6.1 Format Conversion 77
6.2 Data Reduction and Generalization 91
6.3 Error Detection and Editing 93
6.4 Merging 95
6.5 Edge Matching 96
6.6 Rectification and Registration 97
6.7 Interpolation 114
6.8 Photointerpretation 118

7 **Data Management** **126**

7.1 Basic Principles of Data Management 127
7.2 Efficiency 130
7.3 Conventional Database Management Systems 134
7.4 Spatial Database Management 139

8 **Manipulation and Analysis** **143**

8.1 Reclassification and Aggregation 144
8.2 Geometric Operations on Spatial Data 154
8.3 Centroid Determination 156
8.4 Data Structure Conversion 157
8.5 Spatial Operations 157
8.6 Measurement 164
8.7 Statistical Analysis 167
8.8 Modeling 169

9 Product Generation 174

9.1 Types of Output Products 174
9.2 Hardware Components 183

10 Remote Sensing and GIS 191

10.1 A Brief History 193
10.2 Remote Sensing Technology 195
10.3 Digital Processing of Remotely Sensed Data 209
10.4 Interfacing Remote Sensing and GIS's 213
10.5 Synergism 215

11 Practical Matters 220

11.1 A Case Study 220
11.2 Database Volume 223
11.3 Specifications 225
11.4 Project Management 228

12 Applications 230

12.1 Master Planning 230
12.2 Proposed Dam Site 232
12.3 Waste Site Selection 233
12.4 Irrigation and Water Resource Potential 234
12.5 Merging Raster and Vector Data for Map Update 235
12.6 Species Habitat Analysis 236
12.7 Municipal GIS 238
12.8 Agricultural Production Modeling 240
12.9 KBGIS: Artificial Intelligence and GIS 242

13 Looking Toward the Future 245

Suggestions for Exercises 252
Glossary and Abbreviations 261
Bibliography 283
Index 297

Chapter 1

Introduction

As in any other technical area, there can be a fair amount of technical vocabulary to learn before a student can be comfortable with the subject. Since this text is introductory by design, we will try to be consistent in our use of language. Much of the vocabulary of geographic information systems overlaps that of computer science and mathematics in general, and computer graphics applications in particular. We provide a glossary of technical terms at the end of this text, as a reference for the student.

1.1 Geography

Geography has been facetiously defined as that discipline which, when some use is found for it, is called something else. Slightly more serious scholars have defined geography as "what geographers do". The German philosopher Immanual Kant set geography in the context of the sciences by stating that knowledge could be subdivided into three general areas:

1. those disciplines that study particular objects or sets of objects and phenomena (such as biology, botany, forestry, and geology);

2. those disciplines that look at things through time (in particular, history); and

3. those disciplines that look at features within their spatial context (specifically, *geographic disciplines*).

In a more classical sense, the word **geography** may be defined in terms of its constituent parts: *geo* and *graphy*. *Geo* refers to the Earth, and *graphy* indicates a process of writing; thus *geography* (in this literal interpretation) means writing about the Earth.

Another definition of geography focuses on man's relationship with the land. In their writings, geographers deal with spatial relationships. A key tool in studying these spatial relationships is the map. Maps present a graphic portrait of spatial relationships and phenomena over the Earth, whether a small segment of it or the entire globe.

It is interesting that in a survey conducted to determine what factors influenced people to adopt the profession of geography, an early interest in maps rated at the top of the list. There are many skills that people possess to a greater or lesser degree. If a person speaks well, he or she possesses *fluency*. If a person understands writing well, he or she possesses *literacy*. If a person understands numbers and quantitative concepts well, he or she possesses (at least in Great Britain) *numeracy*. Similarly, there is a special skill in the analysis of spatial patterns in two and three dimensions. This skill can be referred to as *graphacy*. Although many individuals take this skill for granted, we all know those who have difficulty reading maps or interpreting aerial photographs. What these two activities have in common is the use of an essentially two dimensional view of geographic space, a view that helps the adept map-reader or photointerpreter to understand spatial relationships.

1.2 Information Systems

The function of an information system is to improve one's ability to make decisions. An **information system** is that chain of operations that takes us from planning the observation and collection of data, to storage and analysis of the data, to the use of the derived information in some decision-making process (Calkins and Tomlinson, 1977). This brings us to an important concept: a map is a kind of information system. A map is a collection of stored, analyzed data, and information derived from this collection is used in making decisions. To be useful, a map must be able to convey information in a clear, unambiguous fashion, to its intended users.

A **geographic information system** (GIS) is an information system that is designed to work with data referenced by spatial or geographic coordinates. In other words, a GIS is both a database system with specific capabilities for

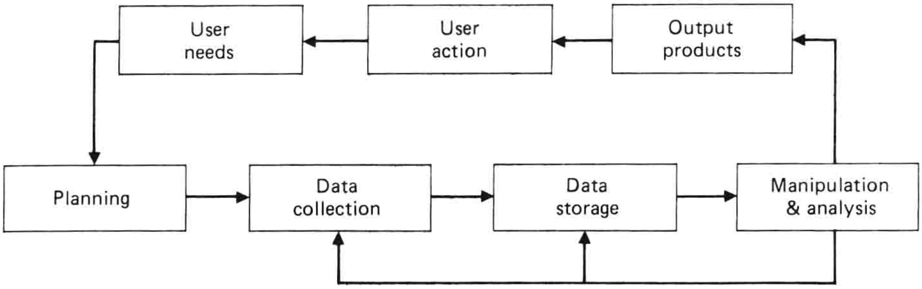

Figure 1.1 Simplified information system overview.

spatially-referenced data, as well a set of operations for working with the data (see Figure 1.1). In a sense, a GIS may be thought of as a higher-order map.

As we shall see later, a modern GIS also stores and manipulates non-spatial data. Just as we have maps designed for specific tasks and users (road maps, weather maps, vegetation maps, and so forth), we can have GISs designed for specific users. The better we are able to understand the range of needs of a user, the better we will be able to provide the correct data and tools to that user.

A geographic information system can, of course, be either **manual** (sometimes called analog) or **automated** (that is, based on a digital computer). Manual geographic information systems usually comprise several data elements including maps, sheets of transparent materials used as overlays, aerial and ground photographs, statistical reports and field survey reports. These sets of data are compiled and analyzed with such instruments as stereo viewers, transfer scopes of various kinds, and mechanical and electronic planimeters. Calkins and Tomlinson (1977) point out that manual techniques could provide the same information as computer-aided techniques, and that the same processing sequences may occur. While this may no longer be entirely true, manual GISs have played an extremely important role in resource management and planning activities. Furthermore, there are still applications where a manual GIS approach is entirely appropriate. Although this text focuses on the technology, instrumentation, and utilization of geographic information systems that are automated, it is still helpful to examine a manual GIS first.

1.3 A Manual Geographic Information System

To introduce some of the language of geographic information systems with a simple first example, let's examine an application of a simple manual GIS. This GIS arises during the early steps in developing a site for a golf course. We assume for this discussion that a specific site is already under consideration. A planner has sought out and gathered together a group of existing datasets for the site. This group might include a topographic map, a blue-line map of parcel boundaries from the local municipal planning agency, and an aerial photograph of the site (Figure 1.2). We refer to these three datasets - 2 maps and a photograph - as **data layers** or **data planes**.

The topographic map depicts several kinds of information. Elevation on the site is portrayed as a series of contour lines. These contour lines provide us with a limited amount of information about the shape of the terrain. Certain kinds of land cover are indicated by colors (often blue for water, green for vegetation) and textures or patterns (such as repeated patterns denoting wetlands). A number of kinds of man-made features are indicated, including structures and roadways, typically by lines and shapes printed in black. In many cases, the information on this map is five to fifteen years out of date, a common situation resulting from the rate of change of land cover in the area and the cycle of map updates. Each of these different kinds of information, which we may decide to store in various ways, is called a **theme**.

The map from the local planning agency provides us with additional and different kinds of information about the area. This map focuses principally on the infrastructure: legal descriptions of the proposed golf course property boundaries, existing and planned roadways, easements of different kinds, and the locations of existing and planned utilities such as potable water, electric and gas supplies, and the sanitary sewer system. The planning map is probably not at the same scale as the topographic map; the former is probably drawn at a larger scale than the latter. Furthermore, the two aren't necessarily based on the same map projection (see section 6.6.1). For a small area like our golf course, the approximate scale of the data is probably more important than the details of the map projection.

The aerial photograph is a rich source of data, particularly for an analyst with some background in image interpretation. A skilled interpreter may be able to detect patterns in soils, vegetation, topography, and drainage, based on the content of the photograph. Unfortunately, this photograph is