Introduction to GIS

http://www.sli.unimelb.edu.au/gisweb/

Dr F. Escobar, Assoc Prof G. Hunter, Assoc Prof I. Bishop, Dr A. Zerger
Department of Geomatics, The University of Melbourne

Introduction

Definition of GIS

Like the field of geography, the term Geographic Information System (GIS) is hard to define. It represents the integration of many subject areas. Accordingly there us no absolutely agreed upon definition of a GIS (deMers, 1997). A broadly accepted definition of GIS is the one provided by the National Centre of Geographic Information and Analysis:

a GIS is a system of hardware, software and procedures to facilitate the management, manipulation, analysis, modelling, representation and display of georeferenced data to solve complex problems regarding planning and management of resources (NCGIA, 1990)

Geographic information systems have emerged in the last decade as an essential tool for urban and resource planning and management. Their capacity to store, retrieve, analyse, model and map large areas with huge volumes of spatial data has led to an extraordinary proliferation of applications. Geographic information systems are now used for land use planning, utilities management, ecosystems modelling, landscape assessment and planning, transportation and infrastructure planning, market analysis, visual impact analysis, facilities management, tax assessment, real estate analysis and many other applications.

Functions of GIS include: data entry, data display, data management, information retrieval and analysis.

A more comprehensive and easy way to define GIS is the one that looks at the disposition, in layers (Figure 1), of its data sets. "Group of maps of the same portion of the territory, where a given location has the same coordinates in all the maps included in the system". This way, it is possible to analyse its thematic and spatial characteristics to obtain a better knowledge of this zone.

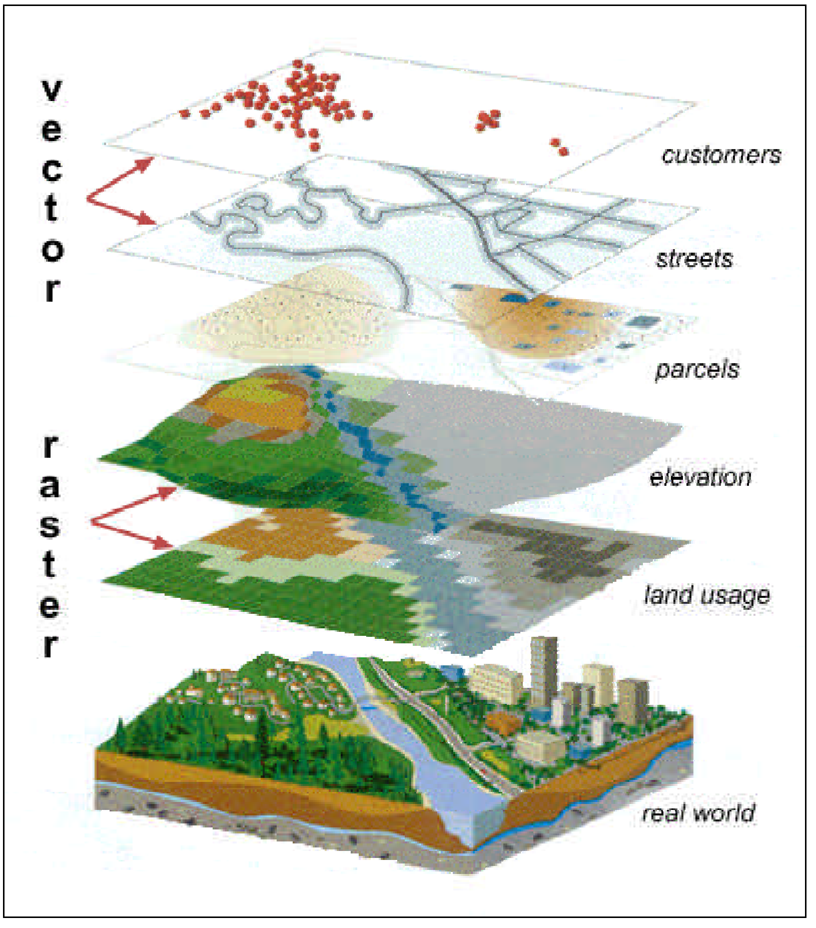


Figure. 1. The concept of layers (ESRI)

GIS applications

mapping locations: GIS can be used to map locations. GIS allows the creation of maps through automated mapping, data capture, and surveying analysis tools.

mapping quantities: People map quantities, like where the most and least are, to find places that meet their criteria and take action, or to see the relationships between places. This gives an additional level of information beyond simply mapping the locations of features.

mapping densities: While you can see concentrations by simply mapping the locations of features, in areas with many features it may be difficult to see which areas have a higher concentration than others. A density map lets you measure the number of features using a uniform areal unit, such as acres or square miles, so you can clearly see the distribution.

finding distances: GIS can be used to find out what's occurring within a set distance of a feature.

mapping and monitoring change: GIS can be used to map the change in an area to anticipate future conditions, decide on a course of action, or to evaluate the results of an action or policy.

Geospatial data

Geospatial data has both spatial and thematic components.

Conceptually, geographic data can be broken up in two elements: **observation or entity** and **attribute or variable**. GIS have to be able to manage both elements.

Spatial component: The observations have two aspects in its localisation: absolute localisation based in a coordinates system and topological relationship referred to other observations. Example: The Department of Geomatics is located at the particular coordinate X,Y, or, The Department is located between Grattan Street and Old Engineering Building. A GIS is able to manage both while computer assisted cartography packages only manage the absolute one.

Thematic component: The variables or attributes can be studied considering the thematic aspect (statistics), the locational aspect (spatial analysis) or both (GIS).

data for GIS applications

data for GIS applications includes:

- digitised and scanned data
- databases
- GPS field sampling of attributes
- remote sensing and aerial photography

digital representation of geospatial data

The advantages of digital versus analogue data are outlined in the table below:

digital	analogue
easy to update	whole map to be remade
easy and quick transfer (e.g. via internet)	slow transfer (e.g. via post)
storage space required is relatively small (digital devices)	large storage space required (e.g. traditional map libraries)
easy to maintain	paper maps disintegrate over time
easy automated analysis	difficult and inaccurate to analyse (e.g. to measure areas and distances)

Vector based GIS

general definitions

Vector is a data structure, used to store spatial data. Vector data is comprised of lines or arcs, defined by beginning and end points, which meet at nodes. The locations of these nodes and the topological structure are usually stored explicitly. Features are defined by their boundaries only and curved lines are represented as a series of connecting arcs. Vector storage involves the storage of explicit topology, which raises overheads, however it only stores those points which define a feature and all space outside these features is 'non-existent'.

A vector based GIS is defined by the vectorial representation of its geographic data. According with the characteristics of this data model, geographic objects are explicitly represented and, within the spatial characteristics, the thematic aspects are associated.

There are different ways of organising this double data base (spatial and thematic). Usually, vectorial systems are composed of two components: the one that manages spatial data and the one that manages thematic data. This is the named hybrid organisation system, as it links a relational data base for the attributes with a topological one for the spatial data. A key element in these kind of systems is the **identifier** of every object. This identifier is unique and different for each object and allows the system to connect both data bases.

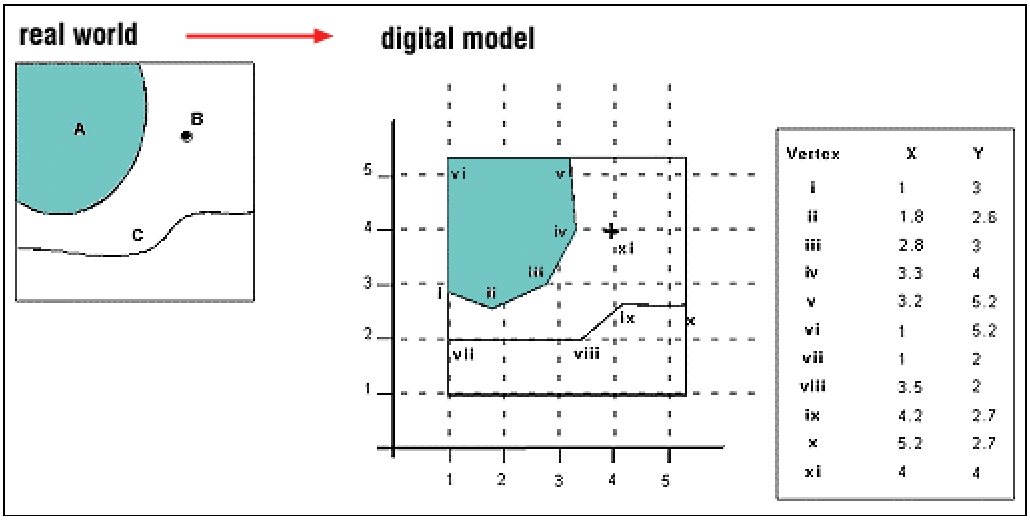


Figure 4.Vector representation

Vector representation of data

In the vector based model (figure 4), geospatial data is represented in the form of co-ordinates. In vector data, the basic units of spatial information are points, lines (arcs) and polygons. Each of these units is composed simply as a series of one or more co-ordinate points, for example, a line is a collection of related points, and a polygon is a collection of related lines.

co-ordinate

Pairs of numbers expressing horizontal distances along orthogonal axes, or triplets of numbers measuring horizontal and vertical distances, or n-numbers along n-axes expressing a precise location in n-dimensional space. Co-ordinates generally represent locations on the earth's surface relative to other locations.

point

A zero-dimensional abstraction of an object represented by a single X,Y co-ordinate. A point normally represents a geographic feature too small to be displayed as a line or area; for example, the location of a building location on a small-scale map, or the location of a service cover on a medium scale map.

line

A set of ordered co-ordinates that represent the shape of geographic features too narrow to be displayed as an area at the given scale (contours, street centrelines, or streams), or linear features with no area (county boundary lines). A lines is synonymous with an arc.

arc

An ARC/INFO term that is used synonymously with line.

polygon

A feature used to represent areas. A polygon is defined by the lines that make up its boundary and a point inside its boundary for identification. Polygons have attributes that describe the geographic feature they represent.

vector models

There are different models to store and manage vector information. Each of them has different advantages and disadvantages.

- list of coordinates "spaghetti" (figure 5)
- vertex dictionary (figure 6)
- Dual Independent Map Encoding (DIME) (figure 7)
- arc / node (figure 8)

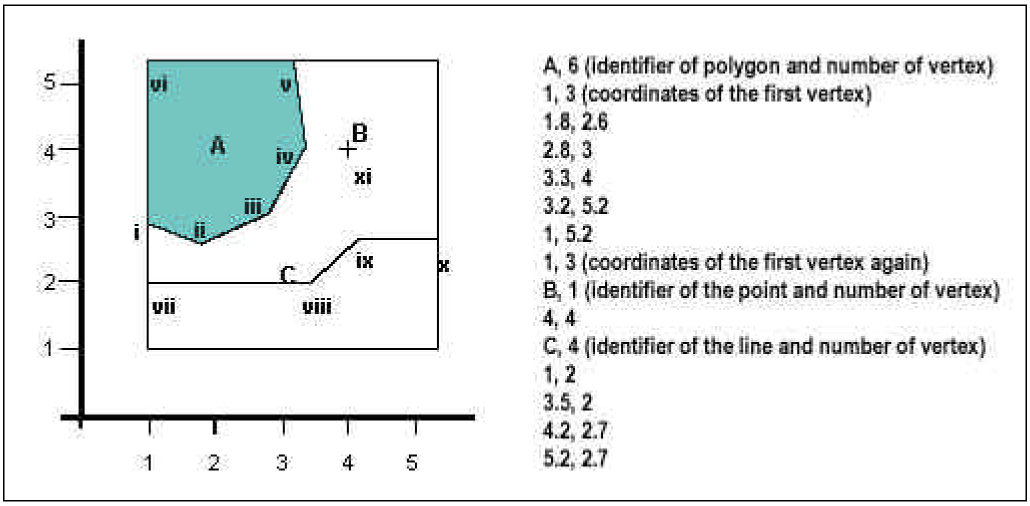
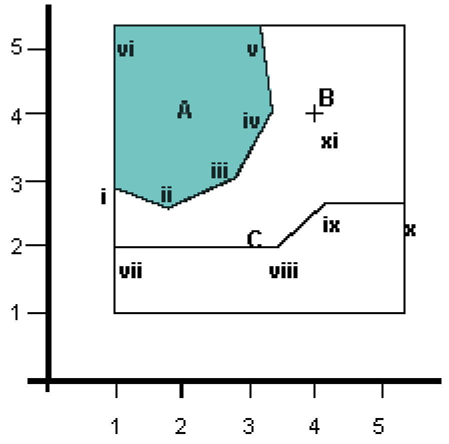


Figure 5. List of coordinates "spaghetti"

- simple
- easy to manage
- no topology
- lots of duplication, hence need for large storage space
- very often used in CAC (computer assisted cartography



file 1

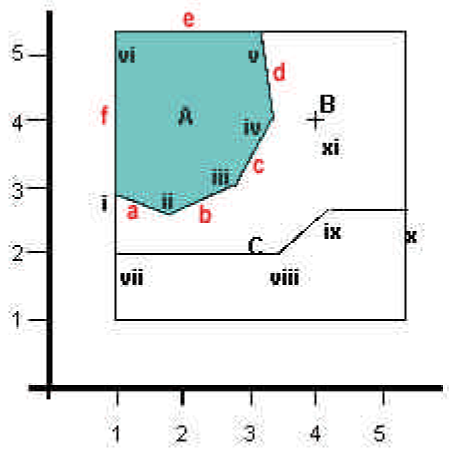
vertex	X	Y
i	1	3
ii	1.8	2.6
iii	2.8	3
iv	3.3	4
v	3.2	5.2
vi	1	5.2
vii	1	2
viii	3.5	2
ix	4.2	2.7
x	5.2	2.7
xi	4	4

file 2

polygon A: i, ii, iii, iv, v, vi
point B: xi
line C: vii, viii, ix, x

Figure 6. Vertex dictionary

?? no duplication, but still this model does not use topology



file 1

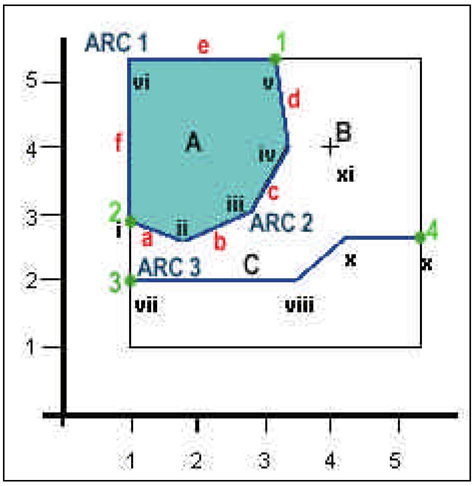
vertex	X	Y
i	1	3
ii	1.8	2.6
iii	2.8	3
iv	3.3	4
v	3.2	5.2
vi	1	5.2
vii	1	2
viii	3.5	2
ix	4.2	2.7
x	5.2	2.7
xi	4	4

segment	right polygon	left polygon	from vertex	to vertex
a	external	A	i	ii
b	external	A	ii	iii
c	external	A	iii	iv
d	external	A	iv	v
e	external	A	v	vi
f	external	A	vi	i
g	external	external	vii	viii
h	external	external	viii	ix
i	external	external	ix	x

polygon	segments
A	a, b, c, d, e, f

Figure 7. Dual Independent Map Encoding (DIME) format

- developed by US Bureau of the Census
- nodes (intersections of lines) are identified with codes
- assigns a directional code in the form of a "from node" and a "to node"
- both street addresses and UTM coordinates are explicitly defined for each link


ARC 1
e
1
vi
v
d
B
f
A
iv
xi
c
2
iii
ARC 2
i
a
ii
b
4
ARC 3
C
x
x
3
vii
viii
5
4
3
2
1
1
2
3
4
5

File 1. Coordinates of nodes and vertex for all the arcs			
ARC	**F_node**	**Vertex**	**T_node**
1	3.2, 5.2	1, 5.2	1,3
2	1,3	1.8,2.6 2.8,3 3.3,4	3.2, 5.2
3	1,2	3.5,2 4.2,2.7	5.2,2.7

File 2. Arcs topology				
ARC	**F_node**	**T_node**	**R_poly**	**L_poly**
1	1	2	External	A
2	2	1	A	External
3	3	4	External	External

File 3. Polygons topology	
Polygon	**Arcs**
A	1, 2

File 4. Nodes topology	
Node	**Arcs**
1	1,2
2	1,2
3	3
4	4
5	5

Figure 8. ARC / NODE structure or POLYVRT

data bases

The elements in a vector based GIS are then the DBMS (Data Base Management System) for the attributes and the system that manages the topological data. In some GIS packages, the DBMS is based in an existing software, i.e. dBASE.

entity-relation model

Three elements are considered in this approach: **(a) Entities** as the relevant objects for the data base. In a GIS, an entity is any fact that can be localised spatially. **(b) Attributes** or characteristics attached to the entities. Each attribute has a limited domain of possible values, i.e. the quality of a road can be *bad*, *average*, *good*, *very good*. **(c) Relations** or mechanisms that allow to relate entities. Some examples are: ‘located in’, ‘contained in’, ‘crossed with’, etc.

DBMS

The data bases used in GIS are most commonly relational. Nevertheless, Object Oriented data bases are progressively incorporated.

relational data bases

In a relational data base, data is stored in tables where rows represent the objects or entities and columns the attributes or variables. A data base is usually composed of several tables and the relations between them is possible through a common identifier that is unique for each entity. Most of the relational data bases in GIS present two variables with identifiers; one of them is unique and correlative, it could be numeric or alphabetic, and the second one might be repeated and helps to organise the attribute table.

The advantages of using this kind of data base are:

- The design is based in a methodology with heavy theoretical basis, which offers confidence in its capacity to evolve.
- It is very easy to implement it, specially in comparison with other models such as hierarchical, network, and object oriented.
- It is very flexible. New tables can be appended easily.
- Finally, many powerful DBMS using this approach contains query languages (like SQL) which makes easy to include this tool in a GIS. Thus, some commercialised GIS packages include a DBMS pre- existent.

object oriented data bases

Based on objects, it can be defined as an entity with a localisation represented by values and by a group of operations. Thus, the advantage in comparison with relational data bases is based on the inclusion, in the definition of an object, not only its attributes but also the methods or operations that act on this object. In addition, the objects belong to classes that can have their own variables and these classes can belong to super-classes.

Raster based GIS

raster representation of data

Raster is a method for the storage, processing and display of spatial data. Each area is divided into rows and columns, which form a regular grid structure. Each cell must be rectangular in shape, but not necessarily square. Each cell within this matrix contains location co-ordinates as well as an attribute value. The spatial location of each cell is implicitly contained within the ordering of the matrix, unlike a vector structure which stores topology explicitly. Areas containing the same attribute value are recognised as such, however, raster structures cannot identify the boundaries of such areas as polygons.

Raster data is an abstraction of the real world where spatial data is expressed as a matrix of cells or pixels (see figure 9), with spatial position implicit in the ordering of the pixels. With the raster data model, spatial data is not continuous but divided into discrete units. This makes raster data particularly suitable for certain types of spatial operation, for example overlays or area calculations.

Raster structures may lead to increased storage in certain situations, since they store each cell in the matrix regardless of whether it is a feature or simply 'empty' space.

grid size and resolution

A pixel is the contraction of the words picture element. Commonly used in remote sensing to describe each unit in an image. In raster GIS the pixel equivalent is usually referred to as a cell element or grid cell. Pixel/cell refers to the smallest unit of information available in an image or raster map. This is the smallest element of a display device that can be independently assigned attributes such as colour.

Pixel size and number of rows and columns:
"The size of the pixel must be half of the smallest distance to be represented" Star and Estes (1990)

raster data structures

exhaustive enumeration

(figure 9)
In this data structure every pixel is given a single value, hence there is no compression when many like values are encountered.

run-length encoding

(figure 10)
This is a raster image compression technique. If a raster contains groups of cells with identical values, run length encoding can compress storage. Instead of storing each cell, each component stores a value and a count of cells with that value. If there is only one cell the storage doubles, but for three or more cells there is a reduction. The longer and more frequent the consecutive values are, the greater the compression that will be achieved. This technique is particularly useful for encoding monochrome images or binary images (Chrisman, 1997).

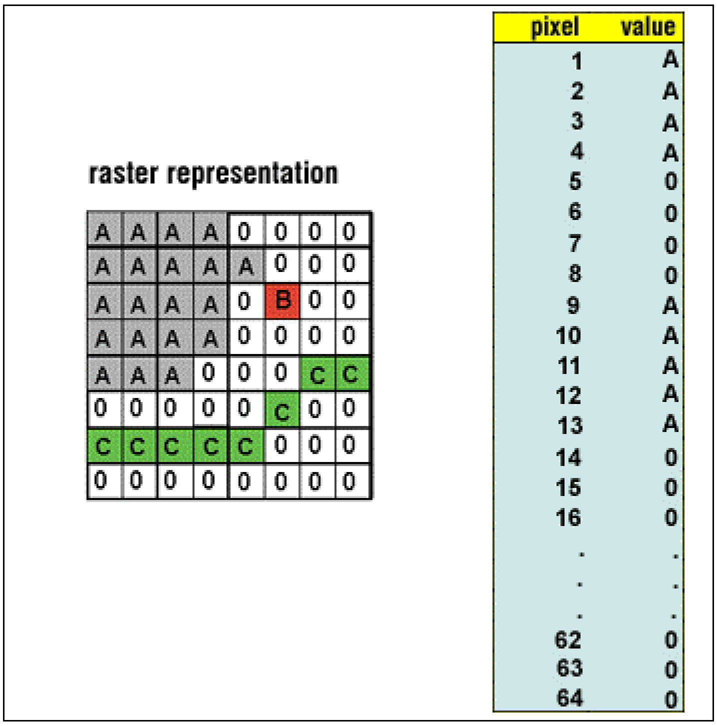


Figure 9. Exhaustive representation

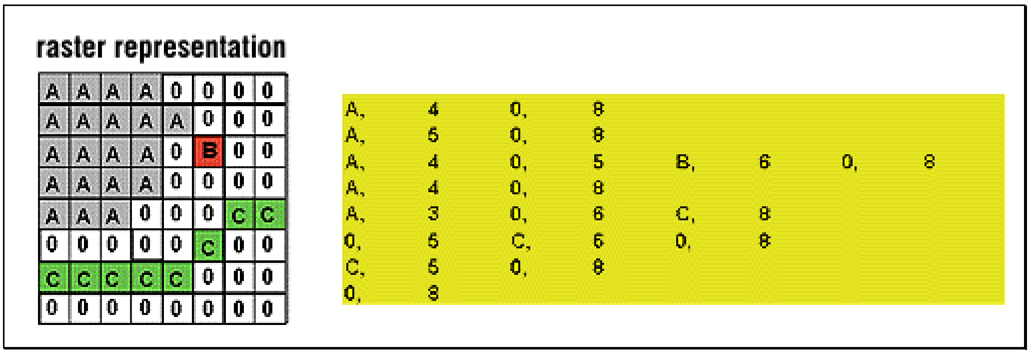


Figure 10. Run-length encoding

advantages/disadvantages of raster and vector data models

	raster	vector
precision in graphics	†	†
traditional cartography	†	†
data volume	†	†
topology	†	†
computation	†	†
update	†	†
continuous space	†	†
integration	†	†
discontinuous	†	†

data capture

Data capture for raster datasets can include:

- Remote Sensing
- Manual digitisation:
  - Points
  - Lines
  - Polygons
- Automatic digitisation
- Scanning

rasterisation of vector data

The process of converting vector data, which is a series of points, lines and polygons, into raster data, which is a series of cells each with a discrete value. This process is essentially easier than the reverse process, which is converting data from raster format to vector format.

raster to vector conversion

The process of converting an image made up of raster cells into one described by vector data. This may or may not involve the encoding of topology.

references

Berhardsen, T. (1992) *Geographic Information Systems*. Viak IT/Norwegian Mapping Authority, Arendal, Norway.

Berhardsen, T. (1996) *Geographic Information Systems*. Halsted Press.

Bernhardsen, T. (1999) *Geographic information systems : an introduction*. Wiley, New York.

Chrisman, N.R. (1997) Exploring Geographic Information Systems. John Wiley and Sons.

deMers, M.N. (1997) Fundamentals of Geographic Information Systems. John Wiley and Sons.

Huxhold, W.E. (1991) *An Introduction to Urban Information Systems*. New York, OUP.

Laurini, R. and Thompson, D. (1992) *Fundamentals of Spatial Information Systems*. London, Academy Press.

Maguire, D.J., Goodchild, M.F. and Rhind, D.W. (eds.) (1991) *Geographical Information Systems: Principles and Applications*. Avon, Longman Scientific and Technical.

Martin, D. (1991) *Geographical Information Systems and their Socioeconomic Applications*. London, Routledge.

Peuquet, D.J. and Marble, D.F. (eds.) (1990) *Introductory Readings in Geographic Information Systems*. London, Taylor and Francis.

Star, J. and Estes, J. (1990) *Geographical Information Systems: An Introduction*. Englewoods Cliffs, New Jersey, Prentice Hall.