GIS Data Sources

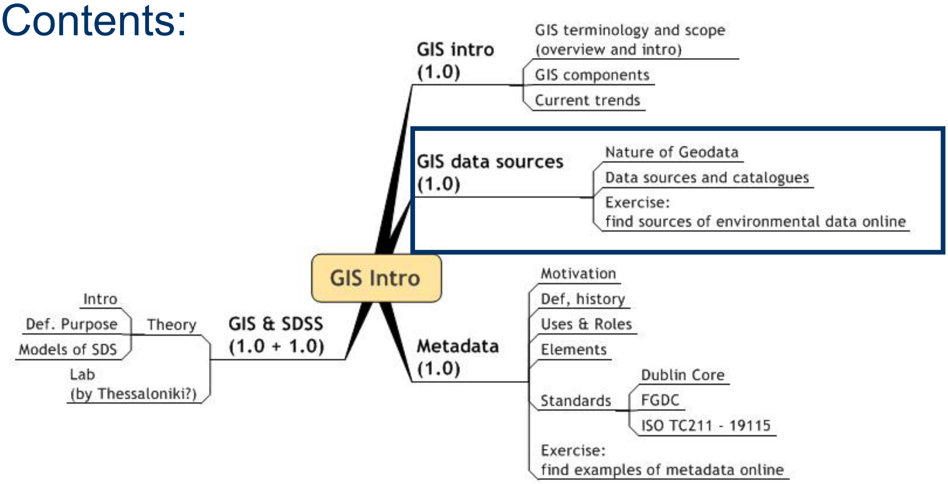
(2) Sources of Geodata

Dr. Adrijana Car

Centre for GeoInformatics – UNIGIS
Salzburg University

adrijana.car@sbg.ac.at

Contents:



This material begins by addressing a simple but very effective application of geographic data in decision making. This will demonstrate the nature of geographic data and also emphasize the power of geographic data.

About these materials

- These materials are part of the UNIGIS MSc Core Curriculum and were developed and compiled by **Wayne Forsythe, Stephen Swales & Adrijana Car**.
-
-

2. The Nature of Geographic Data

- Data are the most important part of a GIS.
- Their acquisition is normally the most costly part of a GIS.
- There are a wide range of data sources for GIS input including analogue and digital forms.

2.1 From Data to Geographic Data

What is the meaning of?

- Data
- Information
- Geographic Data

For example:
Consider the following raw numbers:

32 - 1
34 - 4
36 - 2
38 - 3
40 - 5

Do they have any meaning?

Data are facts about the real world that are largely useless unless placed in context where their meaning can be appreciated, then they become information, and are useful for a variety of purposes.

Let's look at the following example:

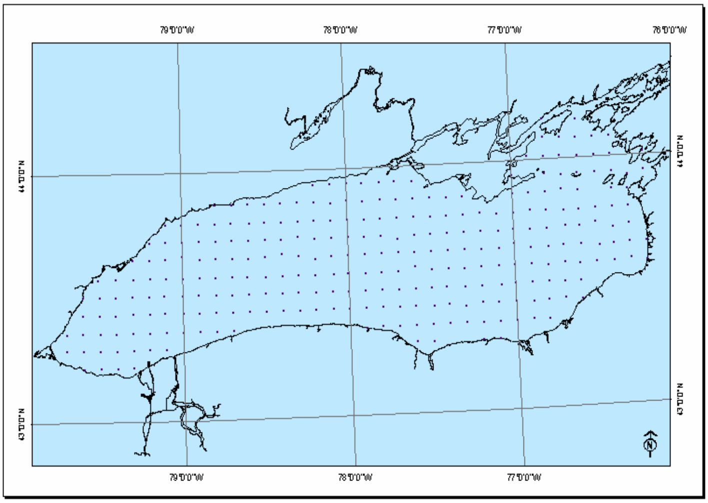
These raw numbers are largely meaningless, unless details are added with respect to what these data represent, such as the unit of measurement (i.e. the number of people with cars) and other context (e.g. addresses):

- 32 Yonge Street - 1 person with a car
- 34 Yonge Street - 4 people with cars
- 36 Yonge Street - 2 people with cars
- 38 Yonge Street - 3 people with cars
- 40 Yonge Street – 5 people with cars

Now the data are truly information and can be subject to analysis.

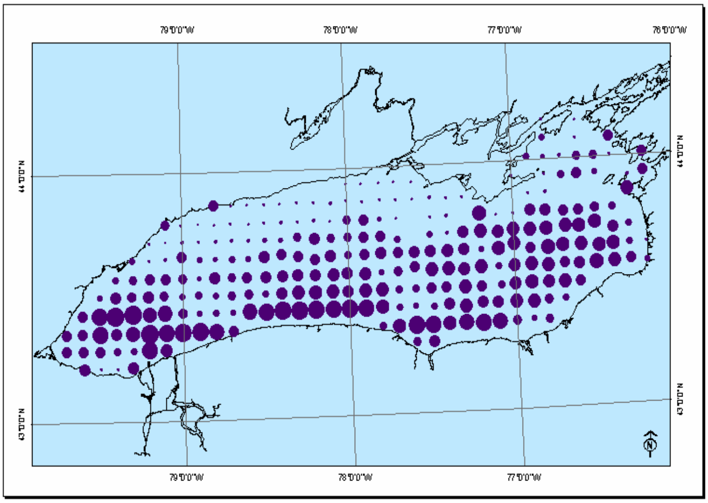
They are geographic data in that addresses are shown, but if we map them we can perhaps analyze them more effectively.

2.1 Example (1): What is Represented Here?



The map on the slide does not provide much information. If you check the coordinates on the map, you could determine that the area represented is Lake Ontario. The point data represent the sampling locations from a sediment contamination survey. More information however is needed for a better idea of what is being represented.

2.1 Example: Better Representation



The original map now has the data points represented by proportional circles.

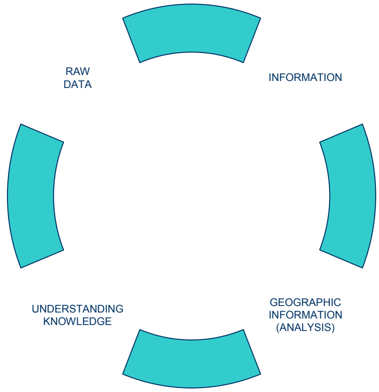
Can you observe any spatial relationship? The variable represented in this map is the level of Mercury contamination found in the sediment samples. After collecting and mapping the data, as GIS professionals, the next stage would be to describe the spatial pattern. This particular pattern shows higher concentration levels (larger circles) along the southern shoreline and in the more central (and deeper) parts of the lake. If you were to think of eating bottom-dwelling fish caught in the lake (not a good idea!), it would be wise to use this information to make an informed decision.

What are the chances that you could have picked up the spatial pattern of contamination from a mere table of Mercury levels?

In this case the benefit of additional information in the form of attribute data and spatial data is clear.

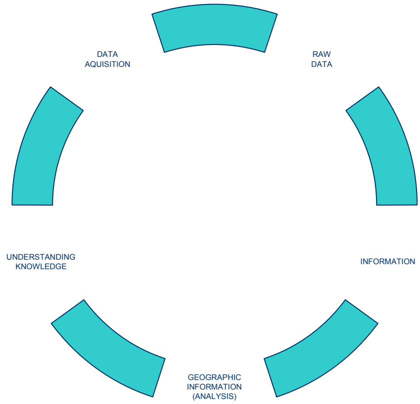
There are of course other spatial analysis techniques that allow you to perform further analyses on data such as those represented in the preceding maps. The anomalies in the map are also interesting, The northern shoreline of Lake Ontario is less populated…..perhaps this is one of the reasons for the lower contamination levels found in this vicinity.

2.1 Example: Geographic Data and Decision Making



In this application, the process of moving from raw data to information to geographic information should be apparent. By analyzing geographic data, an enhanced understanding of a problem was achieved, which facilitated decision making.

2.1 Types of Data



Of course data have to be acquired (or captured) before you can begin any analysis (see the diagram in the slide). In a sediment contamination study, you may collect the data yourself, in which case they would be primary data. Alternatively, you could have acquired the data from public records, in which case they would be secondary data.

The next slides briefly explore the importance of data in a GIS. Much of this material is concerned with the types of data available for a GIS, the means of data collection, and the methods we use to enter data into GIS.

2.2 The Importance of Data in GIS

- For many observers, data, i.e. geographic information, are the most important part of a GISystem.
  - A GIS cannot function without all its components.
  - Together with data there are additional requirements including software, hardware, personnel (or liveware), and an accommodating institutional context, but geographic data are at the heart of the GIS.

2.2 The Cost of Data Capture

	10 seats		100 seats	
	$	%	$	%
Hardware	30	3.4	250	8.6
Software	25	2.8	200	6.9
Data	400	44.7	450	15.5
Staff	440	49.1	2000	69.0
Total	895	100	2900	100

Breakdown of costs (in $1000s) for two typical client–server GIS implementations. Hardware costs include desktop clients and servers only (i.e. not network infrastructure). Data costs assume the purchase of a landbase and digitizing assets such as pipes and fittings (water utility), conductors and devices (electrical utility), or land and property parcels (local government). Staff costs assume that all core GIS staff will be full-time, but that users will be part-time.

(Source : Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W., 2001. Geographic Information Systems and Science. John Wiley & Sons, Ltd. p. 207. Licensed with permission from John Wiley & Sons, Ltd. Copyright (c) 2001 by John Wiley & Sons, Ltd.)

"Data capture costs can account for up to 85% of the cost of a GIS" (Longley et al., 2001).

2.2 Data in GIS

What role does data play and how do they influence the definition of GIS?

Consider the emphasis on data, or geographic information, in the following definitions of GIS:

- "A system for capturing, storing, checking, manipulating, analyzing and displaying data which are spatially referenced to the Earth" (Lord Chorley, 1987).
- "A powerful set of tools for storing and retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes" (Burrough, 1986).
- "Geographic Information System - A system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing and disseminating information about areas of the earth" (Dueker and Kjerne, 1989).

2.2.1 GIS Functions

GIS Function	Examples
Capturing	data entry, table creation, digitizing, geocoding.
Storing	attribute table, database management system.
Checking	checking a list of postal codes that are to be used to create a map of locations (a process known as geocoding).
Retrieving	calling up information directly from a map using an information tool and selecting information from a table.
Manipulating	changing projections; generalising (classifying) data into categories for mapping purpose
Analysing	descriptions of geographical patterns, measuring the association between variables.
Displaying	maps, graphs, tables, images.
Output	maps, graphs, tables and images.

It is the geographic nature of the data that distinguishes GIS from other information systems; it is a *geographic information* system. The key functions of a GIS involve *doing something with geographic data*.

Categorical		Scalar	
Nominal	**Ordinal**	**Interval**	**Ratio**
Presence/absence Counting Diff. in degree or quality Equality of category	Sequence Rel. position < , > , =	Differences Arbitrary Zero + , -	Ratio Real Zero * , /
Classification	**Ordering**	**Measuring**	
name of city type of land use name of highway	large city wettest soil primary highway	angle or bearing in degrees (cycle)	number of people, passengers distance

(see Laurini and Thompson, p.70)

Geographic attributes can be measured in different ways. The usual standards of measurement of data are *nominal, ordinal, interval, and ratio.* In GIS - in addition to nominal, ordinal, interval, and ratio - measurements such as *cyclic*.

The level of measurement is important because it affects data representation (e.g. in map types) and also what kind of analysis we can do with the data.

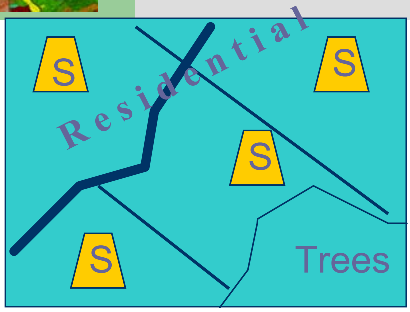
A *nominal* measurement is simply the category name of something such as maple tree, oak tree, or ash tree. No order is implied in these labels.

Ordinal data are categories that are ranked and therefore do imply order. We could, for example, rank the tree types from best to worst for construction purposes. The Richter scale of earthquakes and the Beaufort scale of wind strength are also ordinal scales of measurement.

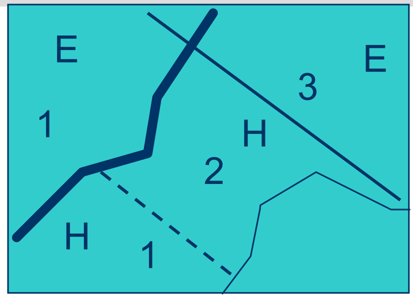
The *interval* level of data management is more precise because numbers are assigned to the observations being measured. An example would be the Celsius temperature scale. However zero on this scale is an arbitrary starting point and is not an observation of no temperature; a value of 30 degrees, for example, can not be said to be twice as warm as 15 degrees.

The highest level of measurement is *ratio* where a zero measurement is zero quantity of the observed phenomenon, and the observations can be subject to meaningful mathematical operations. For example, if one census tract has an average annual income of $100, 000 and another $50,000 we can say that the first has twice the income of the second.

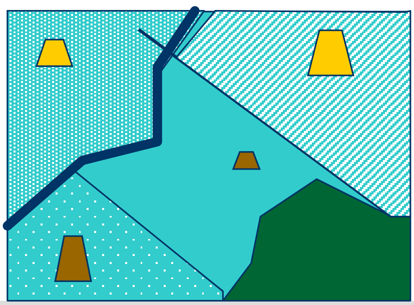
Chrisman (1997) has identified the special nature of some geographic data such as *cyclic*. For example, in compass directions, what is closest to 0, 359 degrees or 1 degree? If you are taking a circular journey through Hell and reach 180 degrees, you may just as well keep going! (Unless, of course, the degrees are in temperature!)



Schools, roads, land use



Type of school, Class of road, Housing type



Number of students
Volume of traffic
Density of housing

As mentioned earlier: The level of measurement is important because it affects data representation (e.g. in map types) and also what kind of analysis we can do with the data. See the above examples.

2.2.2 Review – Geodata Models & Representations

Reality

Conceptual models

Objects or Entities

Smooth, continuous spatial variation

Set of discrete objects, their attributes and relations

Continuous smooth fields

Data models

Sets of simpler objects (atomic entities), their attributes and relations (vector data models)

tessellation (raster data models, TINs)
continuous math. functions

Representations

2.2.3 GIS Data Sources

We distinguish between primary and secondary data sources:

– **Primary** Data
  - Primary geographic data sources are collected specifically for use in the GIS.

– **Secondary** Data
  - Secondary data sources already exist, often in analogue paper form.
  - They are transferred into the GIS for reuse.

- Examples of primary data sources are:
  - remote sensing data
  - GPS measurements
  - surveying, etc.

- Secondary data sources were originally created for a use other than GIS.
- Examples include:
  - topographic maps and
  - placename (toponymy) databases.

REAL WORLD >	SURVEY METHODS >	RAW DATA
	e.g. Remote sensing	e.g. Scanned images
	Ground survey	Photographs
	Interviews	Field documents
	Borehole logging	Census data
Source: Jones. 1997. *Geographical Information Systems and Computer Cartography*		

2.2.3 GIS Data Sources (2)

Classification of geographic data for data collection purposes		
	Raster	**Vector**
Primary	- Digital remote sensing images - Digital aerial photographs	- GPS measurements - Survey measurements
Secondary	- Scanned maps of photographs - Digital elevation models from maps	- Topographic maps - Toponymy (placename) databases
Source: Longley et al., 2001. *Geographic Information Systems and Science*, Chichester: Wiley.		

The type of primary and secondary data will vary according to the type of GIS data model being used: raster or vector, as summarized in the table.

2.3 Catalogues and Data Sources
2.3.1 Data Catalogues

- Data Catalogues
  - large, searchable databases
  - … are repositories used to obtain spatially referenced or geospatial data (Nebert, 2000).
  - Good organisation
    - essential for efficiency in finding the suitable materials
    - obtained through the use of keywords or nodes
      - often break up the information into separate disciplines such as Geography, Biology, Geology, and Water, or geographic areas
  - Can contain both vector and raster data
  - Query tools often provided

Data catalogues also tend to categorize based on areas of the world (usually on the scales of major cities, countries, and continents). Unfortunately, these scales often omit smaller regions and communities (USGS, 2003).

Both vector and raster data can be found in Data Catalogues (ESRI, 2004).

Query Tools are provided for the data catalogue users

• Query tools aid users in finding the correct information.

• They often employ simple Geographic Information Systems to help the user.

• Queries utilize the organization techniques employed by the individuals who own or operate the catalogue. For example they combine the discipline (eg. Geography), spatial requirements (eg. South America) and the content theme (eg. cultural and demographic).

• This type of query is utilized by such data catalogue organizations as the Geography Network (http://www.geographynetwork.com/) and GeoGratis (http://www.geogratis.ca) - (Nebert, 2003).

2.3.1 Query Interfaces for Data Cataolgues

Image Source:
http://www.geographynetwork.com

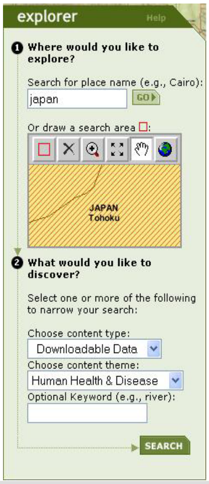
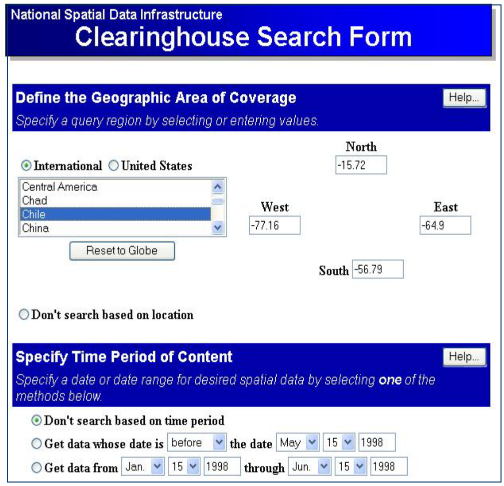


Image Source: http://edcw2ks15.cr.usgs.gov/servlet/FGDCServlet



Example query interfaces from:

- the Geography Network (left - http://www.geographynetwork.com/)
- and the National Spatial Data Infrastructure (right - http://edcw2ks15.cr.usgs.gov/servlet/FGDCServlet).

2.3.2 Obtaining Data

Costs or requirements are related to:

- Availability
- Distribution
- Registration

The methods for obtaining data include:

- Immediate download
- Ordered and sent on disk or tape
- Received from business partners of the data catalogue

The questions in this area include:

•Are the data available at no cost?

•Must they be purchased or must a distribution fee be paid?

•Are users required to register with the data catalogue?

Once the query has been performed and executed, the resulting data will be displayed. The above points outline what is required of the user before the data can be utilized. Although some data are free, other data can become quite expensive and therefore becomes unavailable to individual users (USGS, 2003)

The above methods for obtaining data are an important aspect as they will determine how quickly the data will become available and how the data will be applied (which software program etc) - (USGS, 2003).

2.3.2 Metadata Aspects

It is important to have metadata concerning data that you obtain from data catalogues

- can provide valuable information about the data being obtained
- should be viewed along with the other visual data.

The metadata information should include:

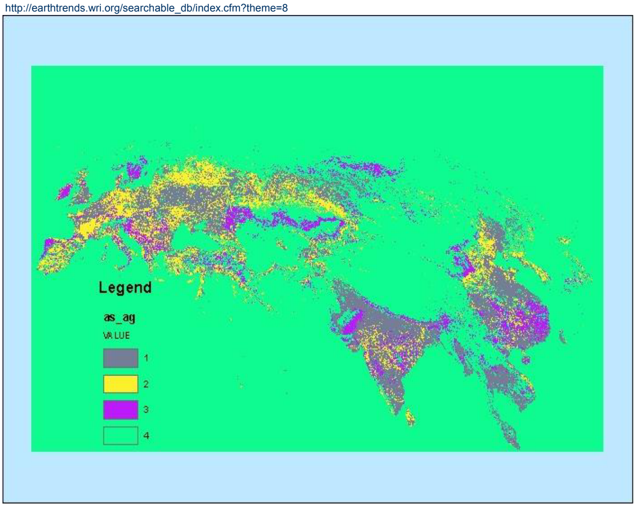
- Spatial data organization information
- Spatial reference information
- Attribute information
- Distribution information
- Metadata Reference information
- Information about which program the data are compatible with (NSDI, 2004).

Metadata can reveal:

- legend information
- file extension
- compression type
- and can give some indication as to the quality or reliability of the data (NSDI, 2004).

2.3.2 Example: Raw Image Data

http://earthtrends.wri.org/searchable_db/index.cfm?theme=8

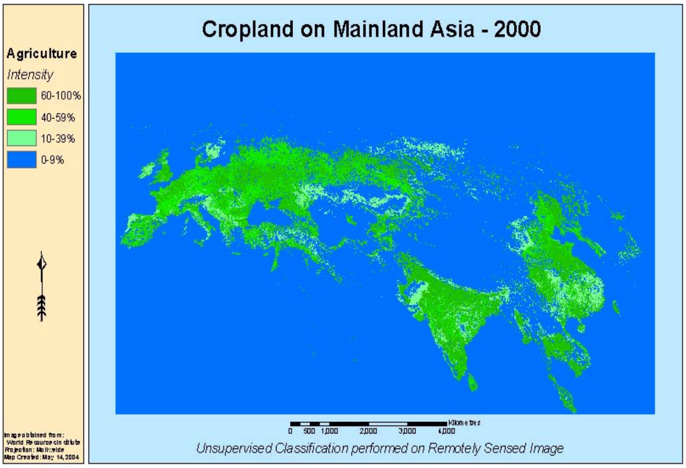


This and the next slide illustrate the importance of metadata.

On the above slide is the raw image as it comes when it is downloaded. It is impossible to know what the map is displaying.

2.3.2 Example: Interpreted Image Data

http://earthtrends.wri.org/searchable_db/index.cfm?theme=8



- The use of metadata allows for the user to make sense of the information which is contained in the image.
- This demonstration shows how important the metadata are when obtaining images from data catalogues.
- Without metadata, the image cannot not be interpreted. Metadata allows for the attributes to be identified (without metadata, the legend could not be determined). The projection can now be displayed and titles such as the source and date can be added accordingly.
- Raw image and metadata obtained from World Resources Institute (http://earthtrends.wri.org/) and outputs created in ArcMap 8.3.

2.4 Viewing Data

- Compatibility
- On-line applications
- Commercial software required

Image Source:
http://www.geographynetwork.com/apps/gateway.html

Data can come in many different forms which are only compatible with certain programs. Sometimes data catalogues provide the user with on-line applications. Commercial software is sometimes required to view data that have proprietary formats (USGS, 2003).

Online applications will often allow the user to zoom in and out of an image or map, select or identify certain features/landforms, and perform minimal levels of analysis (USGS, 2004). E.g. see Geography Network online application (http://www.geographynetwork.com/).

2.4.1 Viewing Data: Online Catalogue Mapping

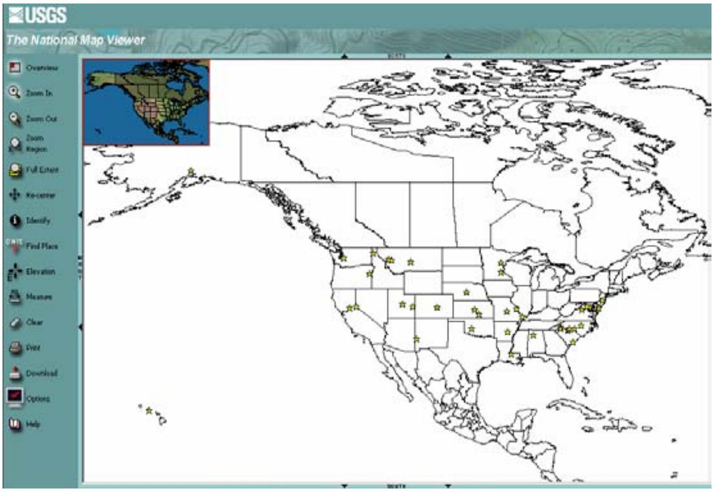


Image Source: http://nmviewogc.cr.usgs.gov/viewer.htm

Online applications are basically equivalent to very simple and user friendly GIS's (USGS, 2004). USGS National Map Viewer (http://nmviewogc.cr.usgs.gov/viewer.htm) is another such example.

2.4.2 Prominent Data Catalogues

Some prominent Data Catalogues include and can be located at:

- CAST – Centre for Advanced Spatial Technologies http://ww.cast.uark.edu/
- Geography Network (ESRI) http://www.geographynetwork.com/
- National Geospatial Data Clearinghouse (USGS is a node of this data catalogue) http://www.fgdc.gov/clearinghouse/clearinghouse.html
- World Resources Institute http://earthtrends.wri.org/
- Geogratis (Natural Resources Canada) http://geogratis.cgdi.gc.ca/clf/en

References

See the notes section for the full reference list

- ESRI. 2004. Geography Network – Frequently Asked Questions. Environmental Systems Research Institute Inc. Available on the World Wide Web: http://www.geographynetwork.com/faq_details.html#datatype
- GeoGratis (Natural Resources Canada). 2004. http://www.geogratis.ca and http://geogratis.cgdi.gc.ca/clf/en
- Nebert, D. 2000. Developing Spatial Data Infrastructures: The SDI Cookbook. Global Spatial Data Infrastructure Association. Available on the World Wide Web at: http://www.gsdi.org/pubs/cookbook/
- Nebert, D. 2003. U.S. National Spatial Data Infrastructure. U.S. Federal Geographic Data Committee. Available on the World Wide Web at: http://wwwlmu.jrc.it/ginie/sdi_ws/323,34,Douglas Nebert
- NSDI. 2004. Geospatial Metadata. National Spatial Data Infrastructure. Available on the World Wide Web at: http://www.fgdc.gov/publications/documents/metadata/metafact.pdf
- The Geography Network (ESRI). 2004. http://www.geographynetwork.com
- Aronoff, S. 1989, Geographic Information Systems: A Management Perspective, Ottawa: WDL Publications.
- Bernhardsen, T. 1999, Geographic Information Systems: An Introduction, New York: Wiley.
- Birkin, M, G. Clarke, M. Clarke and A. Wilson 1996, Intelligent GIS, GeoInformation International/Wiley.
- Business Geographics, Fort Collins, Co.: GIS World Inc.
- Castle, G., 1993, Profiting from a GIS, Fort Collins: GIS World Books.
- Chang, Kang-tsung, 2004, Introduction to Geographic Information Systems, New York: McGraw Hill.
- Chrisman, N. 2002, Exploring Geographic Information Systems, New York: Wiley.
- Clarke, K.C. 2003, Getting Started in Geographic Information Systems, New Jersey: Prentice Hall.
- Delaney, J., 1999, Geographical Information Systems: An Introduction, South Melbourne: Oxford.
- DeMers, M. N. 2000, Fundamentals of Geographic Information Systems, New York: Wiley.
- Dueker, K. J., and Kjerne, D. 1989, Multipurpose cadastre: terms and definitions, Falls Church, VA: ASPRS AND ACSM.
- ESRI, 1997, Getting to Know ArcView GIS, Cambridge: GeoInformation
- Grimshaw, D. J., 2000, Bringing Geographical Information Systems into Business, New York: Wiley.
- Harder, C. 1997, ArcView GIS Means Business, Redlands: ESRI.
- Heywood, I., S. Cornelius and S. Carver 2002, An Introduction to Geographical Information Systems, Harlow: Longman.
- Jensen, J. R., 2000, Remote Sensing of the Environment, Upper Saddle River, N.J.: Prentice Hall.
- Jones, C. 1997, Geographical Information Systems and Computer Cartography, Harlow: Longman.
- Kitchin, R. and N. J. Tate 2000, Conducting Research into Human Geography, Harlow: Prentice Hall.
- Laurini, R. and D. Thompson, 1992, Fundamentals of Spatial Information Systems, London: Academic Press.
- Longley, P. and G. Clarke (eds.) 1995, GIS for Business and Service Planning, Cambridge: GeoInformation.
- Longley, P., M.F. Goodchild, D. Maguire and D. Rhind 1999, Geographical Information Systems: Principles, Techniques, Management, Applications. New York: John Wiley.
- Longley, P. A., M .F. Goodchild, D. J. Maguire and D. W. Rhind 2001, Geographic Information Systems and Science, Chichester: Wiley.
- Maguire, D., M. Goodchild and D. Rhind, (eds.) 1991, Geographical Information Systems: Principles and Applications, New York: Wiley.
- Martin, D. 1998, "Geographical Information Systems and Spatial Analysis", in R. Flowerdew and D. Martin (eds.) Methods in Human Geography: A guide for students doing a research project. Harlow: Longman.
- Masser, I., and M. Blakemore, 1991, Handling Geographical Information, Essex: Longman.