3 GEOGRAPHIC INFORMATION

Bengt Rystedt, Sweden

3.1 Introduction

With Geographic Information we mean information that has a geographic location. The location must be given in a mathematical form that can be used in a computer. The most useful one is longitude and latitude. The location is described more in the next chapter. An easy way to describe how geographic information is handled in a computer is to think in layers (see Figure 3.1), where the landscape is seen in different layers.

Then you can continue with the topographic layers with one layer each for administrative areas, roads, lakes, rivers and so on. Thematic data describing geology, land use and vegetation may be other layers. In Figure 3.1, you see the principle of a digital landscape model based on different layers. This idea to organize geographic data was first introduced in Canada in the 1960s when the *Canada Land Inventory* was built as a fundament for all kind of spatial planning and for the management of national resources.

The layers give the geographical dimensions but these must also have attribute data, which are stored in relational tables. An area in the layer and its attribute data are connected with a unique identity normally called the identification number. A great step forward in handling geographic information and making geographical analysis in a computer was taken when Jack Dangermond found that geometry could be handled in

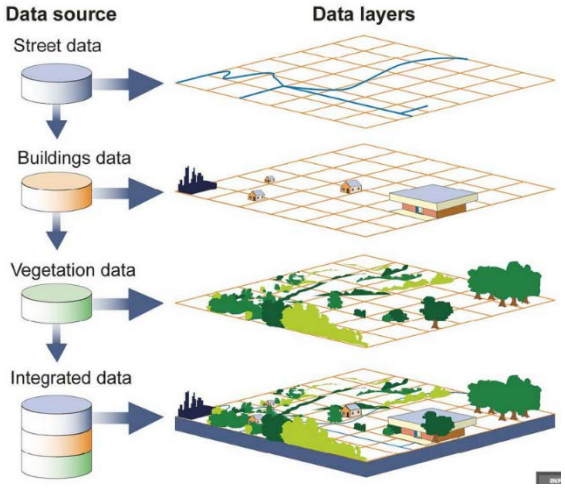

Figure 3.1. The figure shows the principle of a digital landscape model. Each layer contains both location data and a set of attribute data stored in the data source.
Source:
http://education.nationalgeographic.com/education/photo/new-gis/.

one database and the attribute data in another database simultaneously. He called the system ARC/INFO – ARC for geometry and INFO for attribute data in a relational database. Later on, many other systems arrived.

3.2 Data modelling

Before geographic information can be used both for analysis and mapping a geographic data model must be built. One is shown in Figure 3.1, where a beginning of a data model is shown by different layers. The next thing is to define all *objects* that shall be included. The objects are built up by the element´s *points, lines* and *areas.*

The most important part of a geographic data model is its *topology,* which tells how the different elements fit together to form networks and area structures. In a network such as a road system the points are called *nodes* and the topology tells which roads are connected to the node as shown in Figure 3.2.

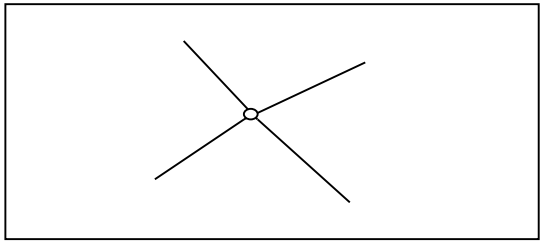

Figure 3.2 shows the principle of a road structure with a node in the middle and four connected roads. The nodes and the roads must have an identity (e.g., an identity number that is easy to find in a database) and may also have attributes.

In an area structure, each area has several neighbour areas. Following a borderline in a direction you always found one area to the left and one area to the right. When the topology for an area structure is calculated each line is given twice, one for each direction, having one area to the left and one area to the right. That may seem un-necessary but it is necessary for getting a system that can be used for geographical analysis in a Geographic Information System (GIS). Figure 3.3 shows an area structure for a municipality with two parishes.

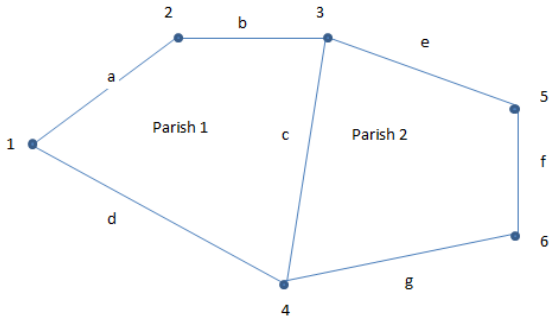

Figure 3.3 shows a municipality with two parishes. Following the boundaries clockwise for each parish, we can see that the boundaries "c" has two directions while the outer boundary only has one direction.

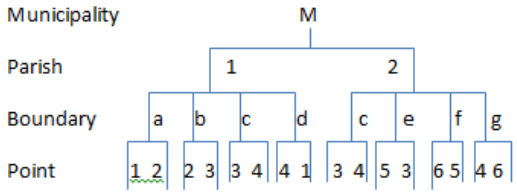

Figure 3.4 shows a hierarchical data structure for the two parishes shown in Figure 3.3. It is also shown that boundary "c" is represented twice and that all points are represented twice but the points "3 and 4" are represented 4 times.

A full delimitation of administrative areas may be *nation, county, municipality, parish and land parcel.* That also means that all these areas can have all these areas as boundaries and must be given in the database (e.g., in a hierarchical data structure as shown in Figure 3.4). The Figure also shows that lines and points will be registered several times in the database and that the size of the database will grow faster than in a linear way.

We have now shown a data structure and hinted that the data shall be stored in a database. A most common organization for a database is the relational database. That means that data are handled in tables and that relations show the connection between the tables. A relational database for the example shown above is given in the following table. The number of columns is defined on how many it will be. The coordinates are just an approximation.

Table 3.1 shows the tables in a relational database. The X and Y coordinates are just a guess.

Municipality	**Parish 1**	**Parish 2**
Municipality name	Parish name	Parish name

Area	**Line**	**Line**	**Line**	**Line**
Parish 1	a	b	c	d
Parish 2	c	e	f	g

Boundary	Point	Point	Parish 1	Parish 2
a	1	2	1	
b	2	3	1	
c	3	4	1	2
d	4	1	1	
e	3	5	2	
f	5	6	2	
g	6	4	2	

Point	**X-coord**	**Y-coord**	**Line**	**Line**	**Line**
1	80	229	a	d	
2	221	121	a	b	
3	375	119	b	c	e
4	372	295	c	d	g
5	517	127	e	f	
6	544	228	f	g	

3.3 Finding the Coordinates in a Database

The tables described above are organized in order of the identity of each item. Each table is stored as a file in the database so it will be rather easy to find an object. It is more difficult with the coordinates. The x-coordinate is defined along the distance from the equator to the pole (the North or the South). The y-coordinate gives the distance in East-West from the meridian that has been chosen and which projection that has been adopted (more details are given in Chapter 9). It is obvious that the coordinates cannot be organized in a table. The problem has been solved by organizing quad-trees. First, we divide the area in four squares and then all the four squares into four squares so we have 16 squares, and so on until we have only one pair of coordinates in each square. We use the binary system to give the squares an identity. After the first division, we get the numbers 00, 01, 10 and 11. By using quad-trees it is easy to find the coordinates in a database by just clicking the screen. An example of a quad-tree is not given here. If you want to know more on quad-trees, Worboys and Duckham (2004) is recommended.

3.4 Information modelling

A geographic database must be based on the real world and be specified on a request analysis. As an example we may look at a system for the management of a fibre-optic cable in a neighbourhood. That will include objects like real properties, property owners (or users), location of the cable, management agreements and costs. The request analysis must be discussed with the future user of the system. It is also important to document the whole process. The working steps and the documentation are given in Figure 3.5.

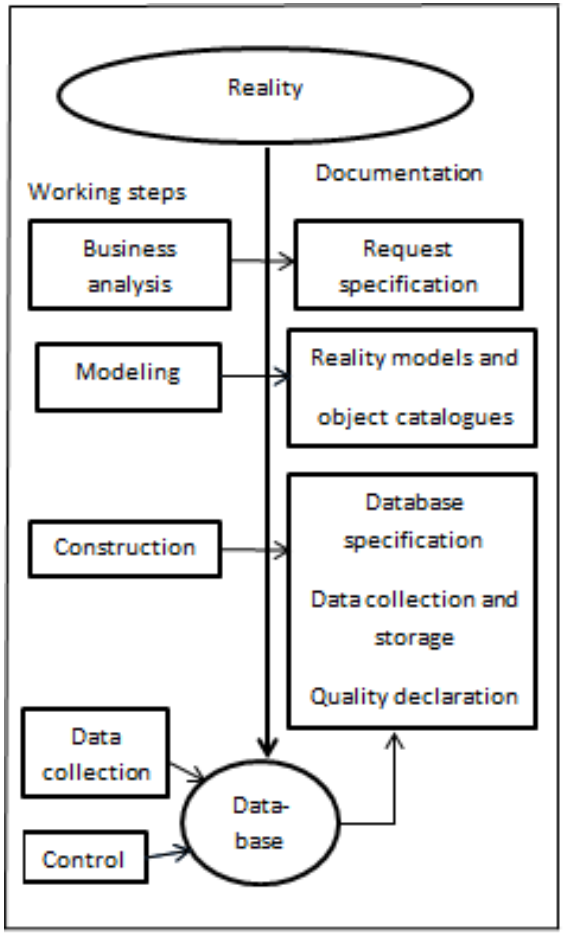

Figure 3.5 shows the working steps in information modelling to the left and the document that should be presented to the right.

Information modelling has been intensively studied during the last years. An ISO Technical Committee (ISO TC 211) has worked out and published many standards and these have been used by all producers of geographic data. For our example with the fibre cable, some object types were mentioned. These shall be documented in the object catalogue.

3.5 Metadata and Quality

We will not go through all standards for geographic information but only a short description of metadata and quality information. Metadata gives a summary of what kind of data is included in the database and gives a summary of the data that may be useful for a thought application.

Metadata (data about data) give a description of the database and may include:

- The name of the database.
- Management organization.
- Geographic area covered.
- A list of the object catalogue.
- Coordinate system.
- Rules for downloading and applications.
- Costs.

Quality data are also a kind of metadata and may include:

- *Origin* giving the basic sources for the data, how data have been collected, and which organisation is responsible.

- *Location accuracy* giving the specified accuracy of the coordinates (both planar and elevation).

- *Actuality* giving the actuality of the data and information of planned updating.

- *Completeness* telling if all objects are included or not, classification correctness, and if the topology is complete (e.g. are the roads a complete network).

The same kind of quality measurement can be given for attribute information included in the objects. Quality may also include information on a quality control system. Altogether, the quality information will tell a user if the database can be used for an intended application.

3.6 Data Collection

A geographic database contains both geometry and attributes. Furthermore, geometry can either be in vector format or in raster format. Vector format is more natural and closer to the geometry we know from school. Raster data are small squares called *pixels* and give a not so detailed representation of the geometry.

Geographic data may be collected in many ways. The highest quality is given by field measurement. Digitizing is more common, when aerial photos and maps are digitized. Land surveyors create land administration systems, for which they measure land parcels and the results of planning applications like locations of houses, roads and bridges. Covered lines for el, telephone and sewage must also be measured. The municipality stores the location of these lines, and an entrepreneur who will dig in the ground has to ask for permission to dig and will receive a map where the lines are located and a permit to dig. However, a map with all the cables is not public since it could be used by criminals for finding spots to hurt vital interests in the municipality. In large cities, there are tunnels filled with cables of different kinds.

Other sources for geographic data are aerial and satellite photos. These are used in agriculture and forestry for measuring land use and vegetation. Google Earth gives a good idea on the possibilities. However, use of images with high resolution may be restricted by military forces or for private reasons. With the increasing high resolution we may see too much. It is free to look at the images but is it not allowed to collect defence items and to further transform the data without a written decision given by the authority who has to be told in accordance with the national law.

For processing geographic data we must have a Geographical Information System (GIS) that can handle geographic information in an efficient way (see Chapter 15). The result of geographic data processing may be shown as maps completed with tabular data like in an Atlas System (See Chapter 7). When we calculate shortest route between two places, we will get a map showing the shortest route and tabular data showing the distance between all turns.

Geographic data can also be collected by using GPS and a handheld device for registration of the data. When back at home we can download the data to the computer and when satisfied transfer the data to the exercise system or to the open street (www.openstreetmap.org), making data public and possible to use for everybody. More details on open street maps will be given in Chapter 16.

References

Worboys, M.F. and Duckham, M., 2004: *GIS: A Computing Perspective,* Second edition. London: CRC Press. ISBN 0-415-28375-2.

The text in this chapter is based on a guideline to databases, published in 1994 by Lantmäteriet, the National Land Survey of Sweden.