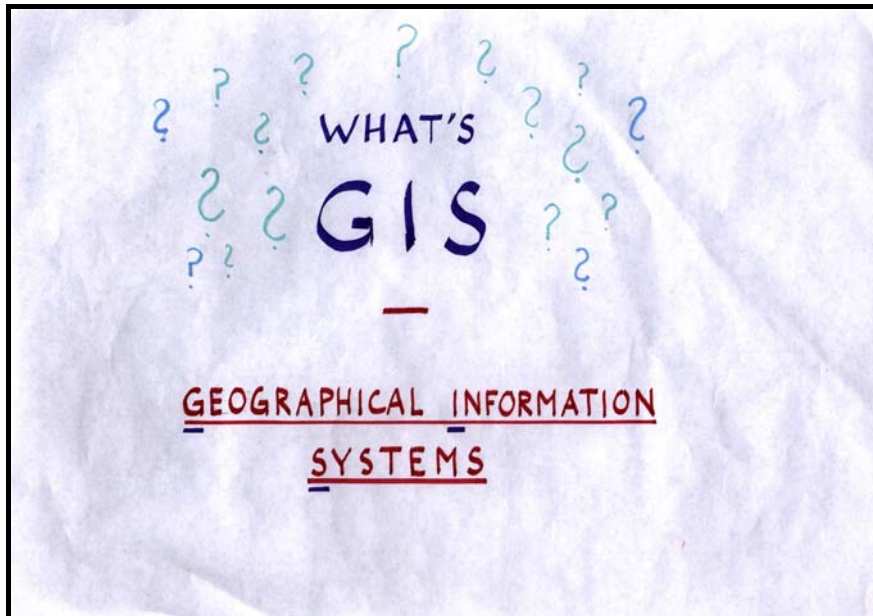




LUND
UNIVERSITY

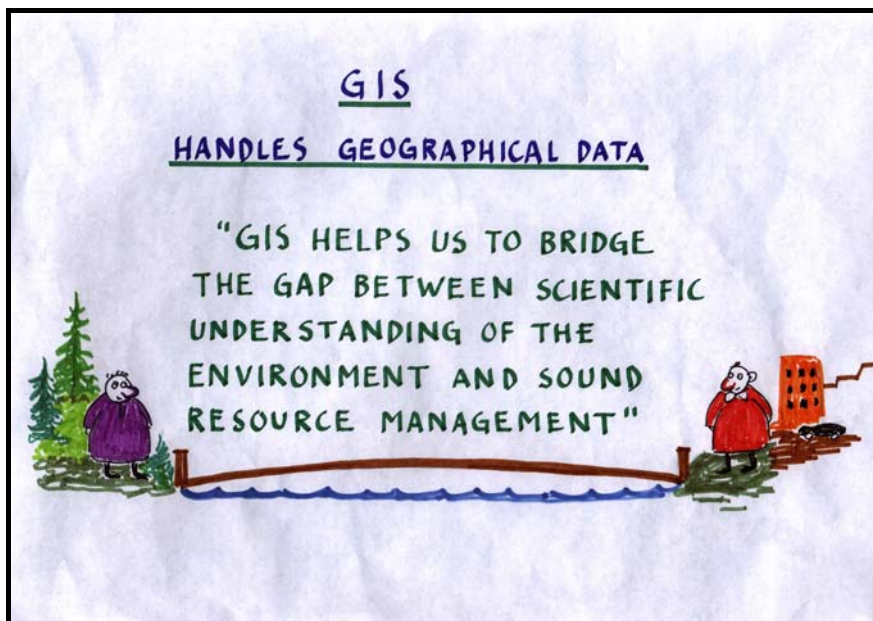


INTRODUCTION TO GIS



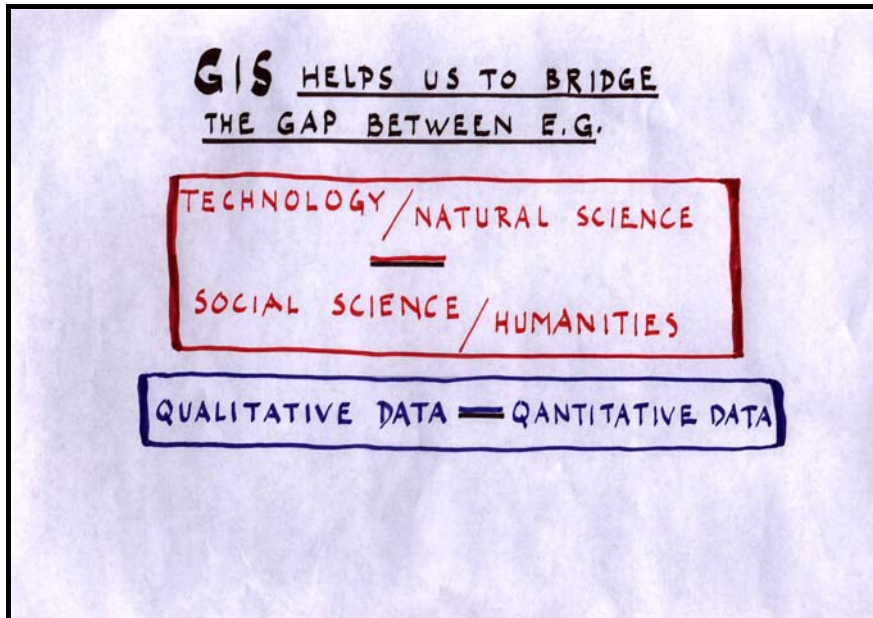
Slide 1

Welcome to this introduction to the Geographical Information System (GIS). This text and slides will guide you through an overview of GIS and introduce you to basic concepts. The first issue is to define what GIS is.



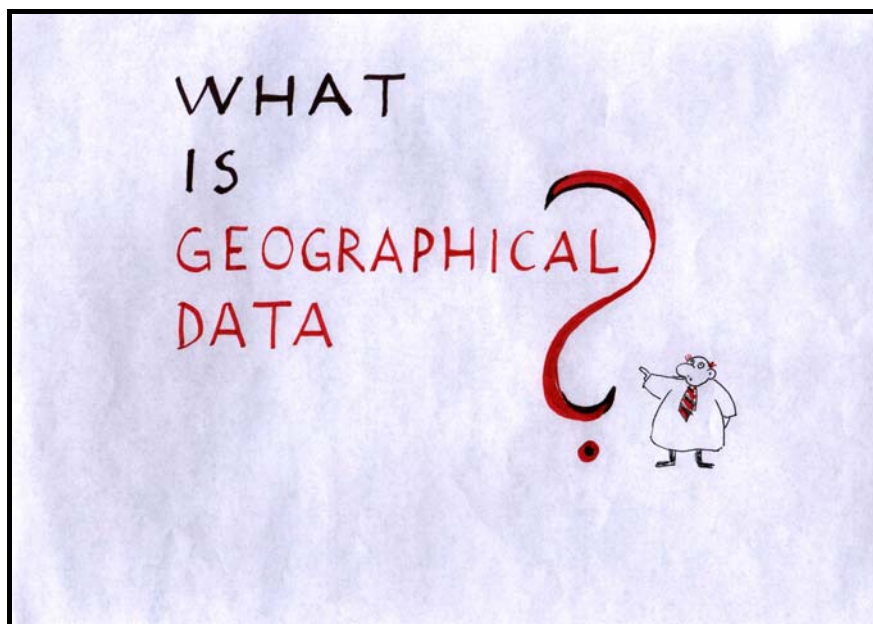
Slide 2

The name implies that GIS is a system that handles geographical information or geographical data. As such, it is a tool that helps bridging the gap between different disciplines, e.g. scientific understanding of the environment and sound resource management. It also provides the tools for using different types of information together to solve different problems. But the most important feature with the GIS is that it handles information in a digital format. In the past, such information was mainly in paper map format but now computers are used to handle it instead. From the computerised databases, information can be retrieved quickly and easily, as both maps and/or tables.



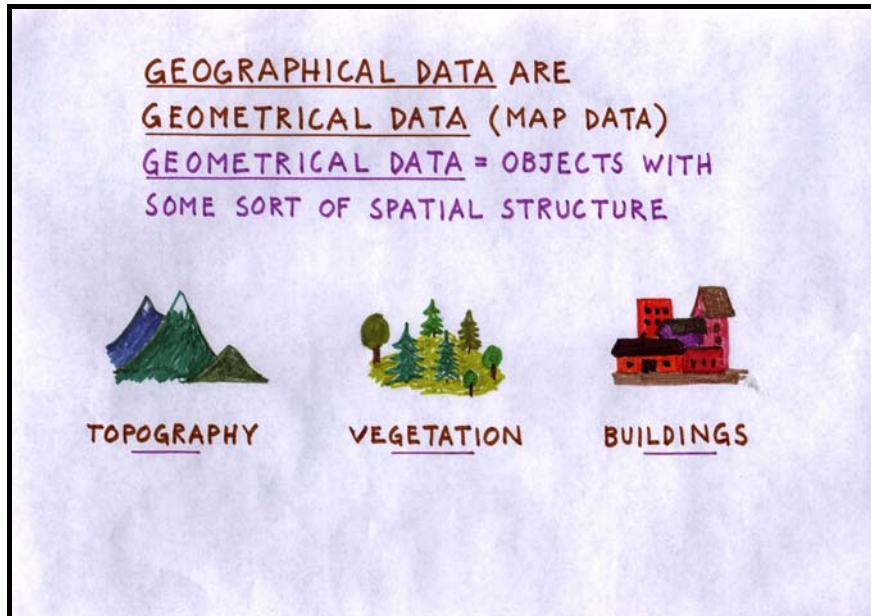
Slide 3

Another important feature with GIS is that it provides tools to bridge between technology and natural sciences / social sciences / human sciences, since it can be used to combine quantitative and qualitative data types.



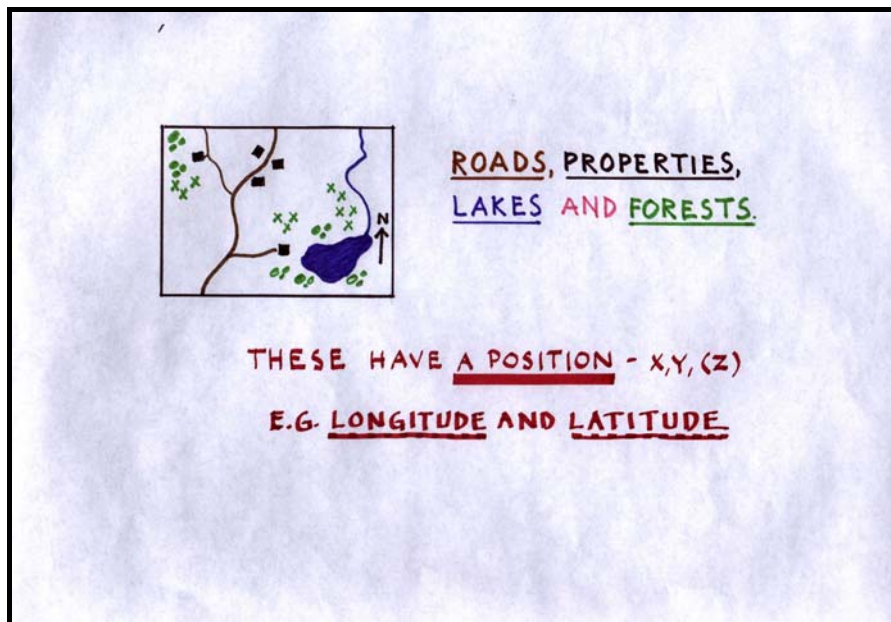
Slide 4

Before it is possible to understand the GIS concept, it is necessary to define what geographical data is, since GIS operates by using this type of data. The definition of geographical data is that it is data that can be allocated a location, most often described by geographical co-ordinates. In plain English, this means that anything that can be put on a map is geographical data. In the case of GIS, the map data is stored in digital format.



Slide 5


Geographical data is partly map data, often referred to as geometrical data, that is, objects with a spatial structure. Examples of geometrical data are vegetation, buildings, topography, roads, wells, cultivated areas, rainfall, etc. Almost all objects in our surroundings can be connected to a specific location and are possible to handle in a GIS. Even dynamic objects, e.g. objects that are moving, are possible to include in a GIS. GIS is also very efficient for handling information about change and development by comparing maps from different time periods.




Slide 6

This is a simple map showing a road, houses, a lake and a river, which all are geographical objects. The basic requirement for displaying these objects is that it must be possible to connect them to a position. Geographical positions are described by an x co-ordinate and a y co-ordinate that indicate the position in a two-dimensional co-ordinate system. Many maps use co-ordinates given in longitude and latitude to describe geographical position.

GEOGRAPHICAL DATA CAN BE REPRESENTED
AS THREE DIFFERENT TYPES OF
GEOMETRICAL OBJECTS IN A GIS

 **1** POINT OBJECTS (E.G. DRILL HOLES,
SAMPLE POINTS, WELLS ETC)

 **2** LINES (ROADS, WATERCOURSES AND
BORDERS)

 **3** POLYGONS (LAND USE, LAKES AND
PROPERTIES)

Slide 7

Almost every object in our surroundings can be represented on a map using one of three different basic geometrical object types. The simplest type is the point, which is indefinitely small since it has no spatial extension. Points can be used to represent objects like a well, a tree, a soil sample pit, etc. The second object type is the line. Linear objects could be, for example, a road, a river or a powerline. The third type of object is the polygon or surface, which covers an area. In map-making, the selection of geometrical object type used for representing different features is very important. Depending on the scale, the object types can also be dynamic. For example, on the map of Africa, the city of Nairobi would be represented as a point while on a map of central Kenya, it would be presented as a polygon.

SOME GEOGRAPHICAL DATA HAVE NO
SPECIFIC SPATIAL DELIMITATION
AND ARE THEREFORE BEST
REPRESENTED AS A CONTINUOUS
SURFACE

EXAMPLES OF SUCH DATA:

* PRECIPITATION

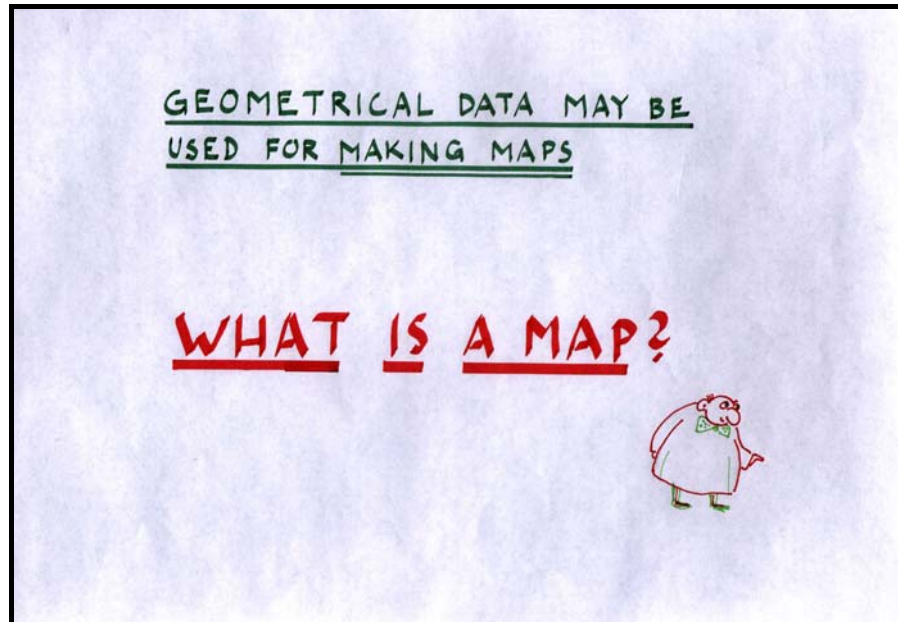
* TEMPERATURE

* TOPOGRAPHY



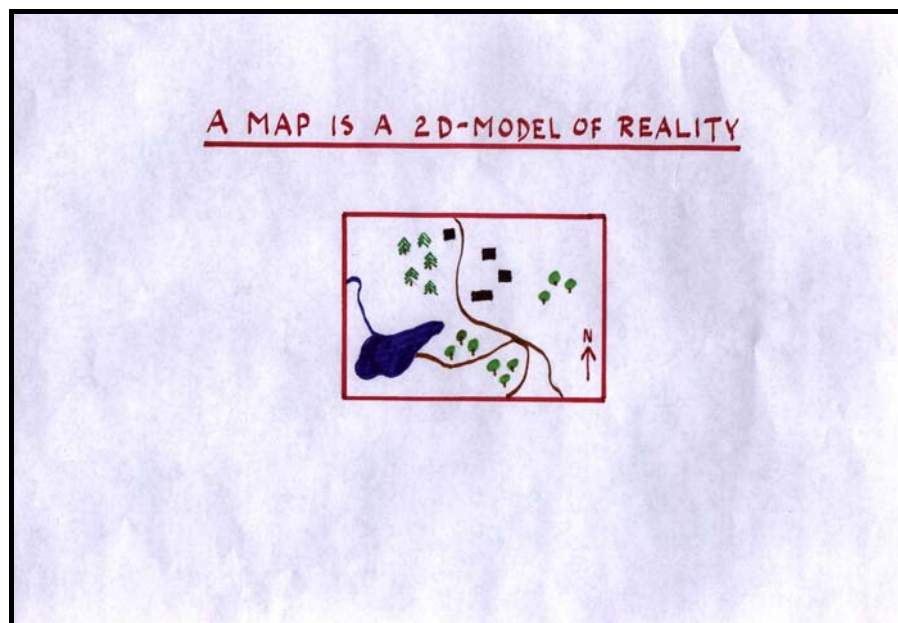
Slide 8

Points, lines and polygons can be used to represent most objects in our environment, which have a discrete and defined spatial delimitation. However, some geographical data have no defined spatial delimitation, and represent instead events that vary continuously over space, e.g. topography, temperature, precipitation and distance. These continuous surfaces must be represented differently on a map.



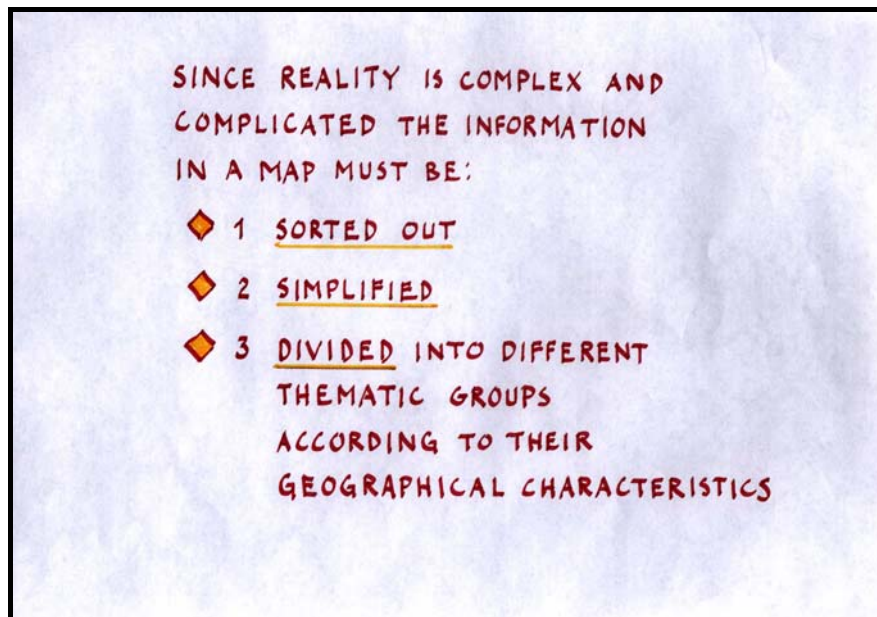
Slide 9

A map, by definition, is a generalised description of the reality. In a way it could be regarded as a static geographical database printed on paper, where, in most cases, points, lines and polygons are used to represent different objects.



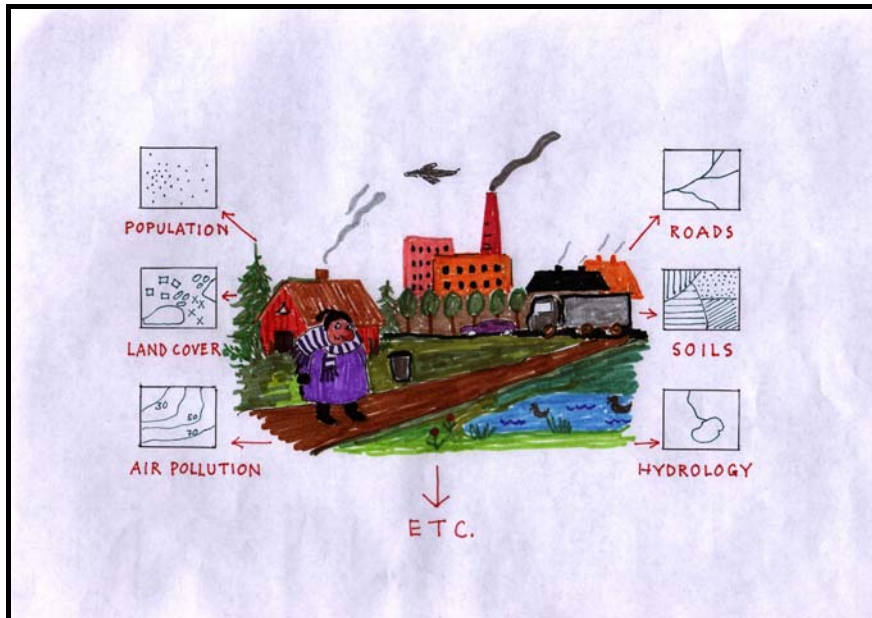
Slide 10

The real world is three-dimensional but often it could be represented on two-dimensional maps, making the map a two-dimensional model of the reality. The third dimension, e.g. topography, can also be represented on a two-dimensional map by using iso-lines or different colours for different elevations.



Slide 11

In order to depict a complex real world in an interpretable way, maps must be generalised. Generalisation is done in different ways. The objects to be represented may be sorted or selected so that only the most important objects will be represented on the map. What to include on the map is a decision made by the map producer. On a printed paper-map, it is not possible to change this once the map has been printed, while on a digital GIS-map, the user can control what objects to display by selecting different parts of the digital database. Objects may also be simplified, which means that objects that are geometrically complex will be simplified to a simpler geometrical form. A shoreline of a lake that is very complex would be simplified to a series of straight lines, and small bends in a river would be straightened out, etc. Very often, the reality is divided into different thematic map layers to decrease the number of objects presented in a single layer, and this is also a form of generalisation. This is done by separating roads from rivers, vegetation from agriculture, population from water supply, education from income, etc., into different thematic map layers.



Slide 12

This is an example of how the complex reality is divided into different map layers, representing different thematic aspects of the reality. The number of information layers that can be derived from an area is virtually indefinite, since almost everything can be allocated to a certain position. The different thematic layers are combined to a map containing only the objects of interest for a particular professional or that are to be analysed together for solving a specific problem. This means that the GIS-map is dynamic and that the user controls what to display or use in the analysis. This is a major advantage of GIS compared to traditional paper maps.

GEOGRAPHICAL DATA ARE ALSO DATA THAT DESCRIBE THE GEOMETRIC OBJECTS, E.G.

- 👤 THE SPEED LIMIT OF A ROAD
- 👤 THE OWNER OF A PROPERTY
- 👤 THE AGE OF A FOREST
- 👤 THE SIZE OF A HOUSE


SUCH DATA ARE CALLED:
ATTRIBUTE DATA

Slide 13

Geographical data is not only geometrical data but is also composed of what is called attribute data. Attribute data is data that provides information about the geometrical objects. A line object representing a road could have attributes describing the speed limit of that road, whether it is paved or not, number of vehicles passing per day, etc. The owner of a property, the age of a forest, the size of a house, are all attribute data about different geometrical objects. The attributes describe the characteristics of the map objects and are linked to these objects in the GIS. The complete geographical database is composed of both geometrical data (map objects) and attribute data.

ATTRIBUTE DATA CAN BE:

- *1 TABLES
- *2 TEXTS
- *3 IMAGES
- *4 SOUNDS
- *5 FILMS
- *6 ETC.



Slide 14

Traditionally, only tables have been considered as attribute data, but with today's computer technology, any type of digital information can be considered as attribute data. Tables, text, images, audio, video are all examples of attribute data that can be linked to the geometrical objects.



Slide 15

The following images exemplify different types of attributes that can be linked to geometrical objects. (in this case, the map of the Nordic countries, Denmark, Sweden, Norway and Finland). This data was downloaded from the Internet and combined with attribute data also downloaded from other Internet sites.



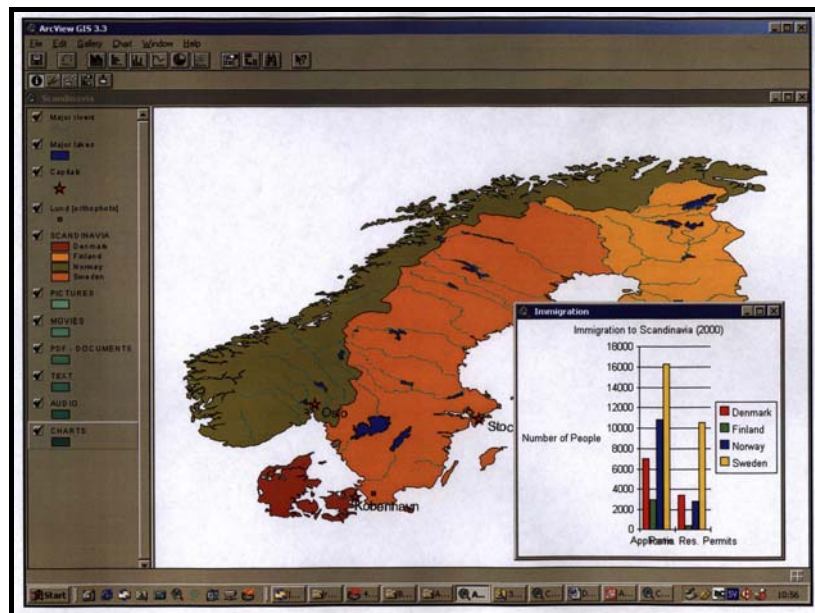
Slide 16

This is an ASCII text document linked to Denmark, containing a general description of the country. When pointing to the Denmark polygon, this text is displayed in a new window on the computer screen.



Slide 17

This example is an Internet *.pdf document containing text information about Finland that is displayed when the Finland polygon is selected.



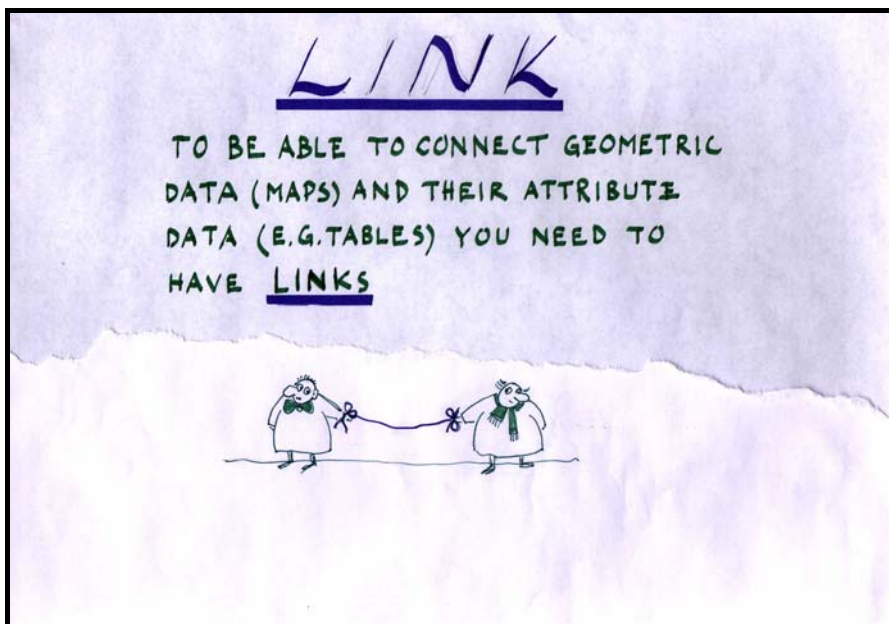
Slide 18

This example is a bar chart showing statistical information about the Nordic countries. This is maybe the classical way of using GIS, to present analysis done on tabular data.



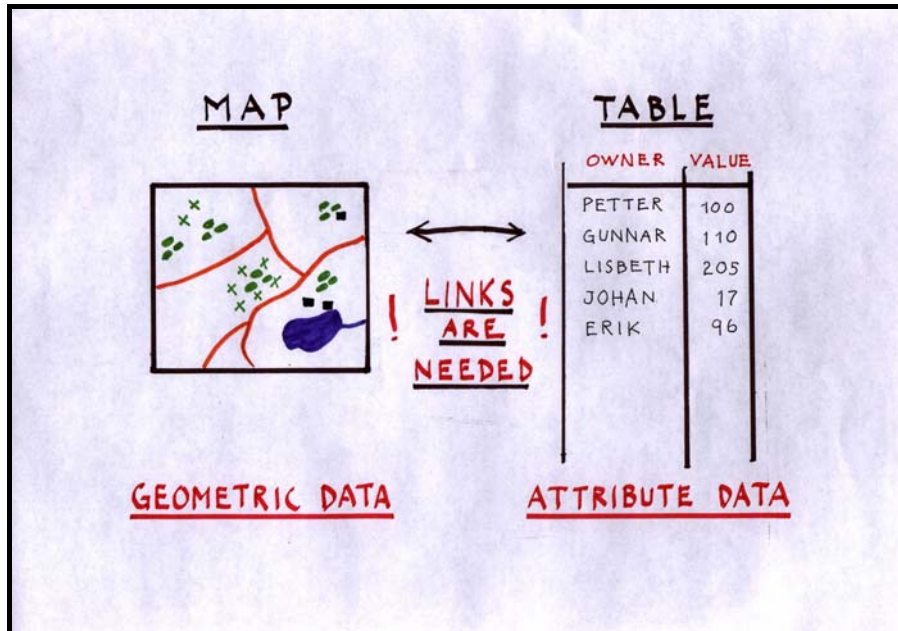
Slide 19

This example is an image showing the Swedish midsummer festivities. All these examples (slides 16-19) illustrate that the possibilities to link different types of attribute data to the map data is practically unlimited.



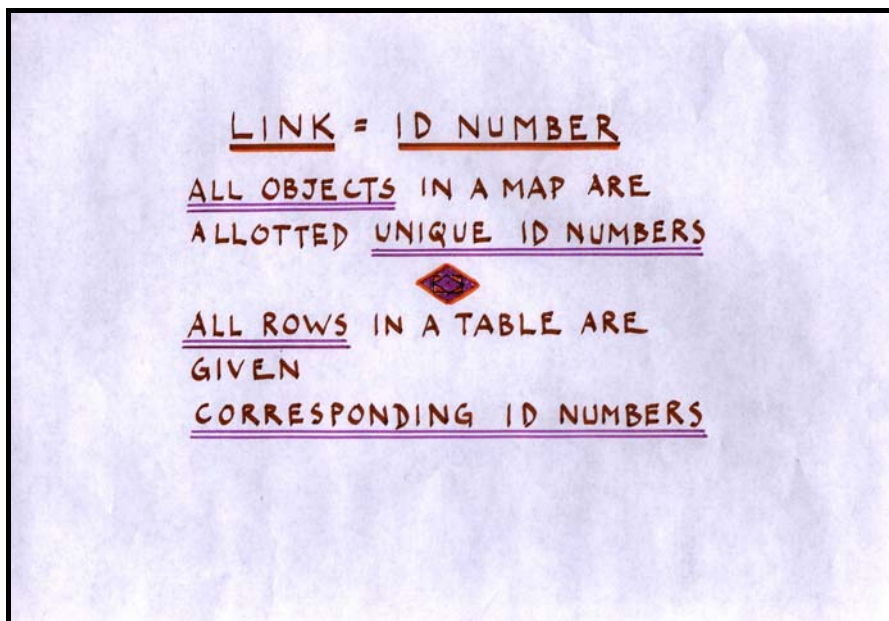
Slide 20

To be able to connect geometrical data (map objects) and the attribute data (for examples, tables), we need a link indicating which objects belong to each other. The link is used by the computer to recognise what attribute data is connected to what map object. This link is the very essence of GIS since without it the system will not operate properly.



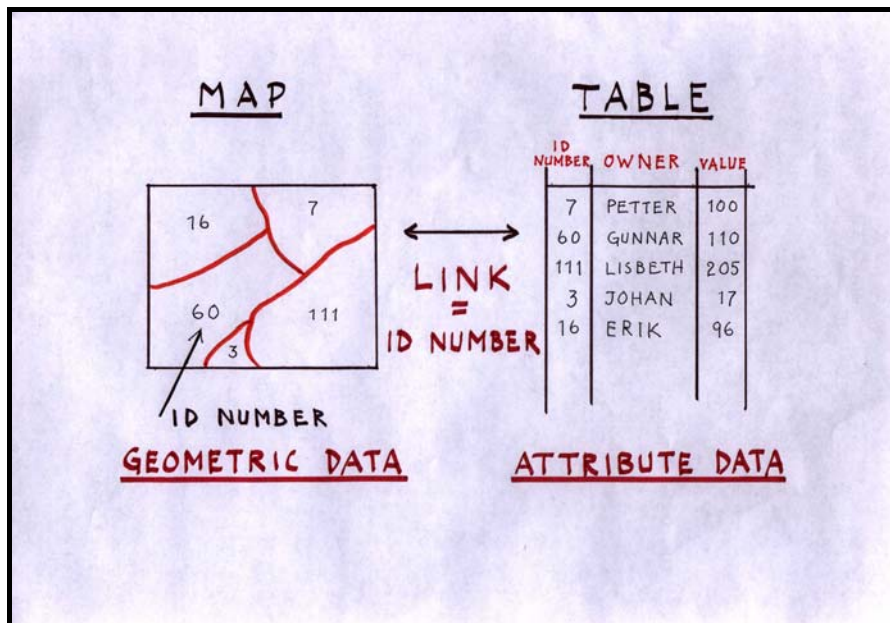
Slide 21

Geometrical data containing a number of map objects and a table containing attribute data describing the map objects. Without a link between them, it is impossible to know which row in the table describes which object on the map. A link that unites them is necessary.



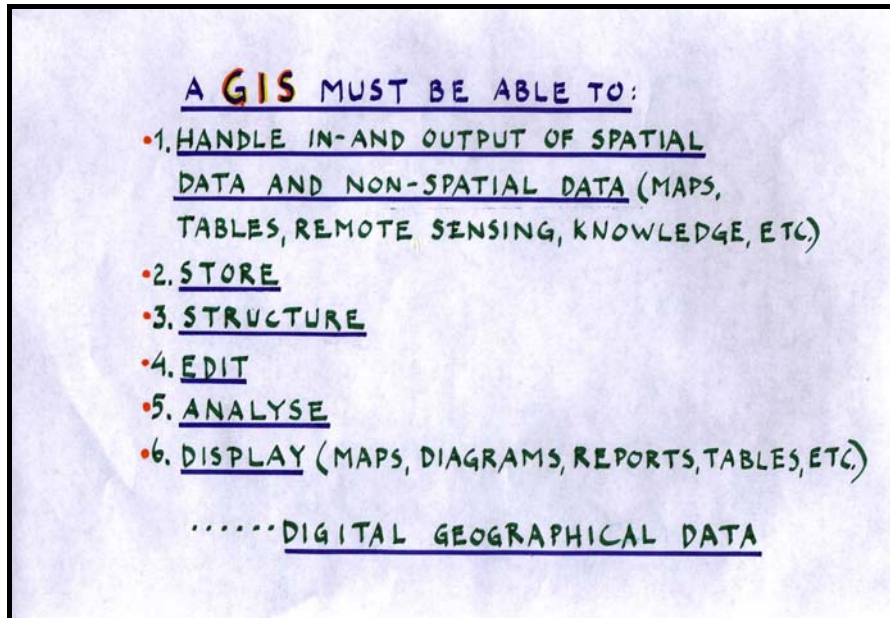
Slide 22

The link between map and attribute data is established using UNIQUE id-numbers. All objects on the map have been assigned unique id-numbers and the corresponding row in the table has been given corresponding unique id-numbers. If the Id-numbers are not unique, the system will not be able to distinguish between two rows in the table or between two objects on the map.



Slide 23

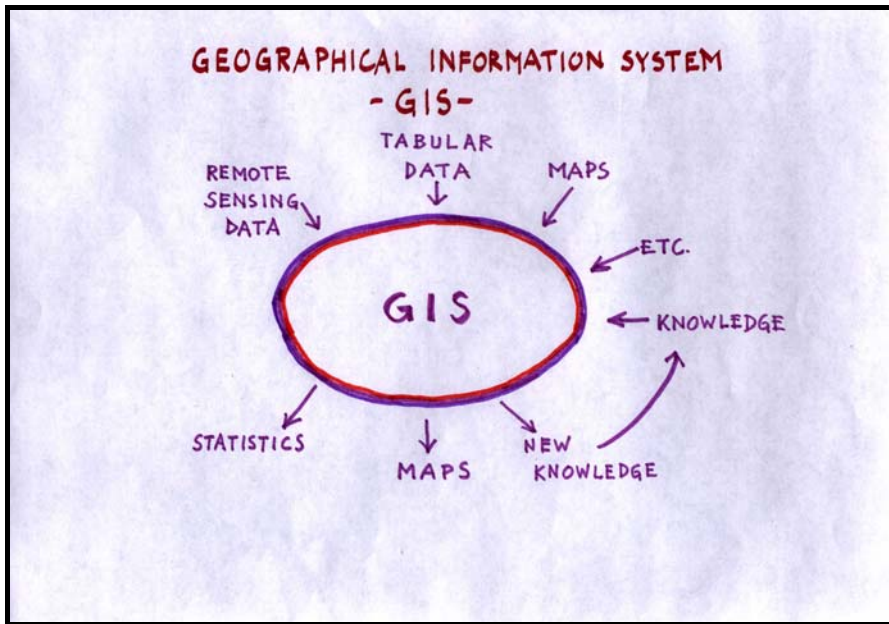
Five properties (polygon objects) on the map have been assigned unique id-numbers, in this case 3, 7, 16, 60 and 111. The id-number itself is not interesting, it could be any number and the objects do not need to be numbered sequentially or according to any system. The only important thing is that no two objects have the same id-number. Many GIS software do not display the id-numbers on screen or when printing maps. Similarly, the tabular data has been allocated id-numbers that corresponds to the numbers of the map objects. When the user is pointing at an object in the map, the computer will read the id-number of that object and then move to the table and find the corresponding table entry and present the table information stored in that row of the table on the screen. This is a very common and basic GIS operation – to search in the map data and get the result from the attribute table. The opposite is also possible, if a row in the table is selected the computer will recognise the id-number and move to the geometrical database to search for the map object with the corresponding id-number and flash this on the map. As a general GIS user, most people will work with GIS in this way – to retrieve information. A more advanced user may also add or change the information, particularly the information in the attribute table. It is however important to understand that if the id-number of an object or table row is changed, the link between map and attribute may be broken and the system will fail to operate properly. To summarise – the link is the essence of GIS since it is the link between geometrical data and attribute data that permits us to organise a GIS and to perform all kinds of operations on a GIS database.



Slide 24

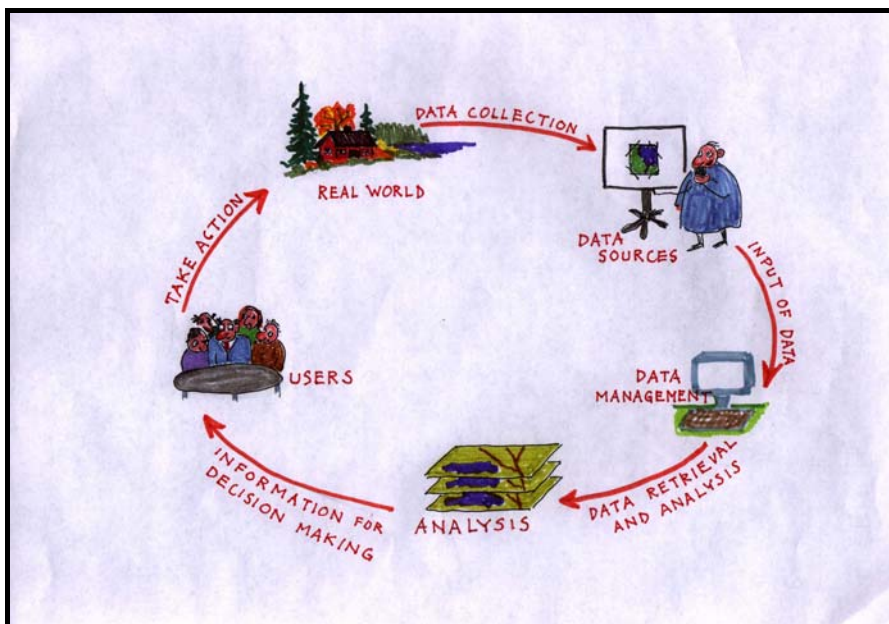
A complete GIS should be able to:

- Handle input and output of geometrical data and attribute data. This involves converting between analogue (paper map or paper table) data and digital format and between different digital formats.
- Store the information in compatible formats usable in GIS applications
- Structure the information to ensure data access at maximum performance, and maximum security for the data holdings
- Edit the information to reflect changes in the real world, updating the information
- Analyse – the GIS should provide tools for analysing the stored data
- Display – tools for visualising the information in the database



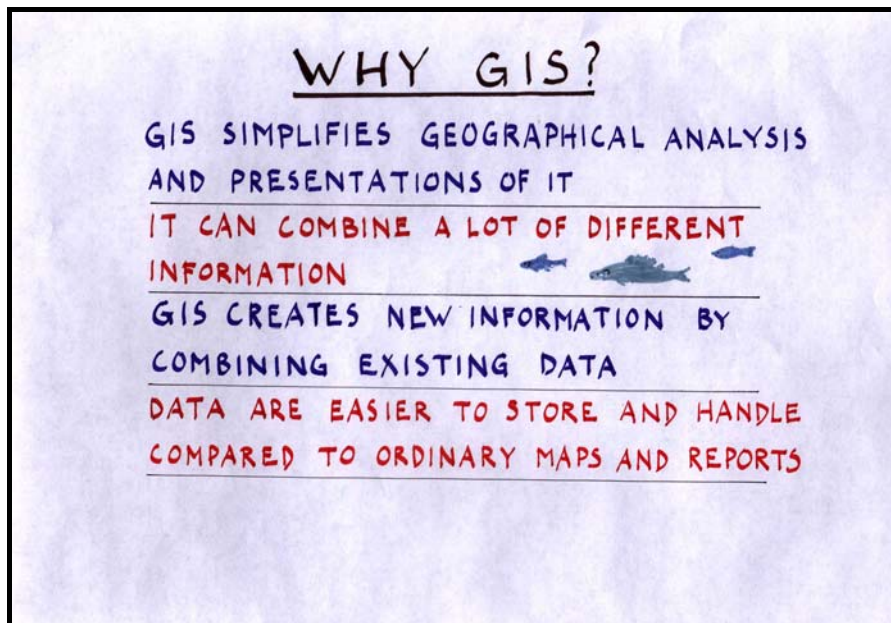
Slide 25

This slide does not really consider what is happening inside the GIS (it is rather a black box model of GIS). In the GIS, input data (which could be in the form of maps, tables, satellite images and aerial photographs, etc) is analysed and the GIS analysis produces new statistics or tables, new maps, and most importantly, new knowledge. Knowledge is naturally also part of the input, since it is necessary to have an understanding of the problems and issues that GIS are used to analyse. The GIS can handle very complex problems and knowledge is necessary to understand these problems. New knowledge generated in the GIS will increase the understanding, which will enable even more complex analyses.



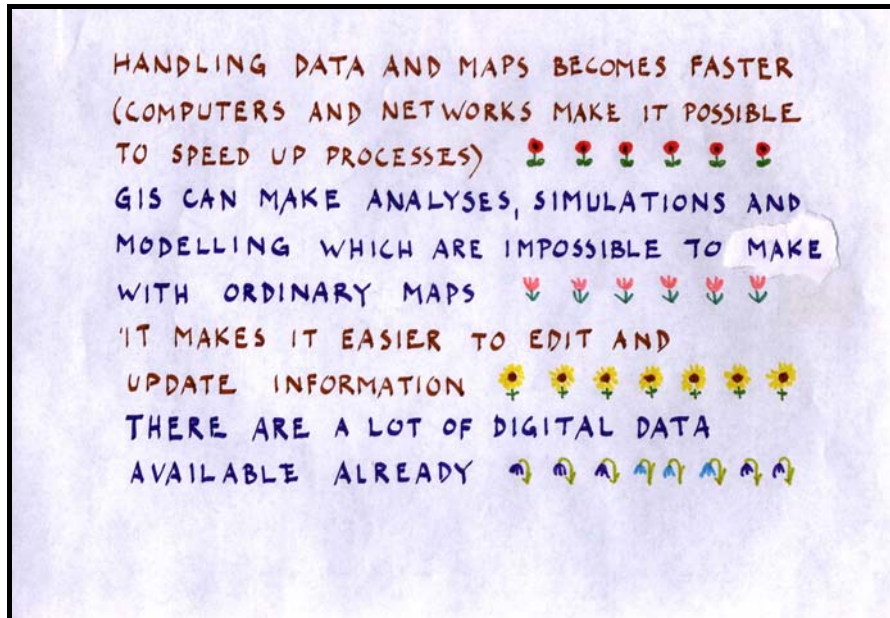
Slide 26

The use of GIS can be illustrated by the “GIS cycle”. Starting with data collection in the real world, data is digitised and structured in a computer database. Data from different sources may be combined and analysed together to yield results that are used for decision-making. The decisions will eventually have an effect on the real world, making new data collection from the real world necessary to update the database. Thus, the GIS-“cycle” will never end as long as there are changes in the real world.



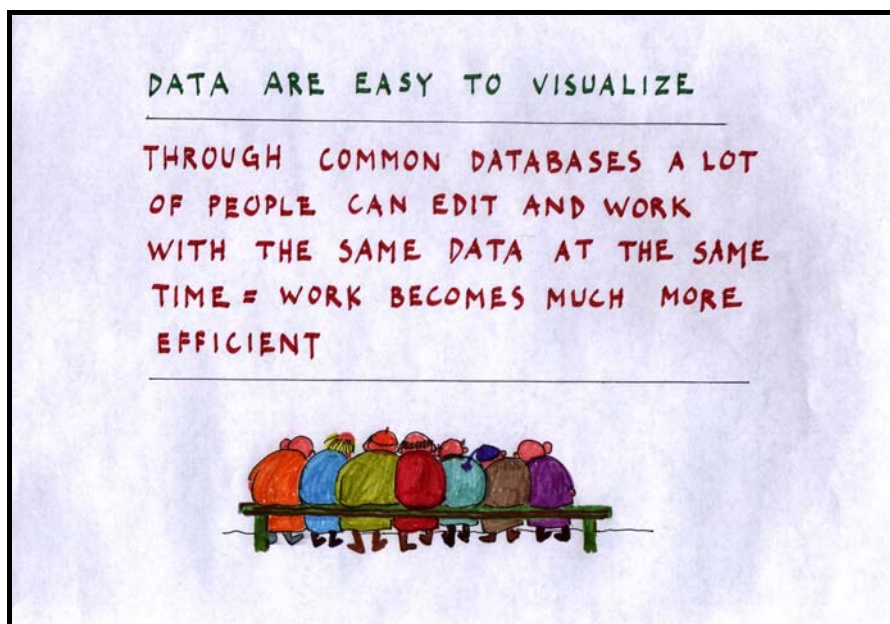
Slide 27

There are several reasons for using GIS instead of traditional manual techniques. The GIS will simplify the analysis and presentation of geographical information because the results from any analysis will also be in map format. When working with paper maps, results from analysis often had to be redrawn on new maps, a step that is now omitted in the analysis – the presentation process. Data from different sources can easily be combined in a GIS provided that it is stored in a compatible format and a corresponding geographical reference system. By combining information, it is possible to create new knowledge about the field of study or area of interest. Traditionally all map information was handled manually, and the process of co-analysis was both complicated and tedious. Now, the GIS can combine any type of information types as long as the databases are well organised and stored in a compatible format. GIS is also a tool that makes storing and handling of data more efficient, since all data is supposed to be accessed from any computer connected to the common database. This means that data is stored only at one place (thus conserving computer space) and handled more efficiently since it is shared. Via the Internet, a user in one country could connect to a database in another country and manipulate the data if permission is granted.



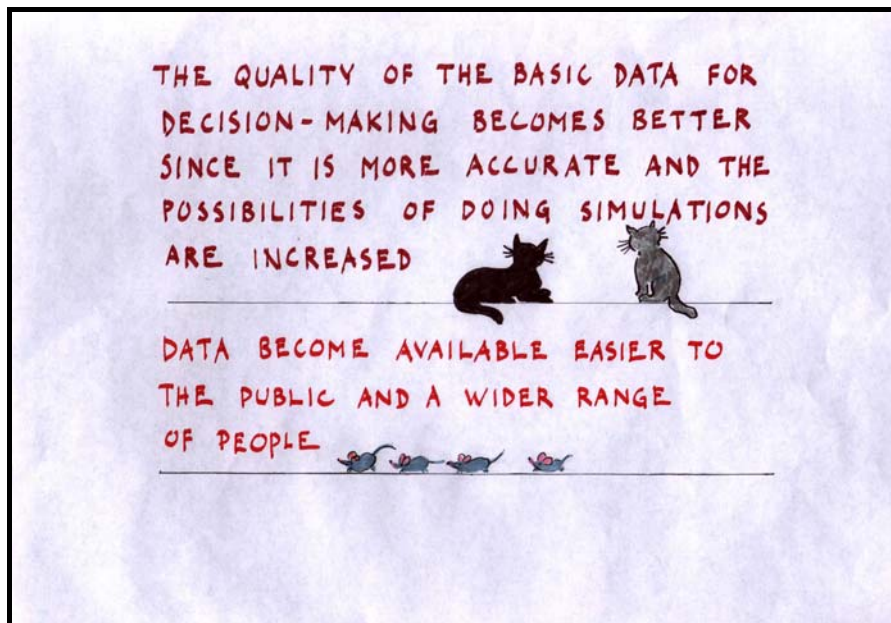
Slide 28

The GIS speeds up handling of data and maps since everything can be accessed from the users computer, reducing time for search and retrieval of information. Advanced GIS analysis, such as simulation, modelling and creation of scenarios, used to be very complicated when working with paper maps describing different spatially related problems. In a computerised environment, modelling of even very complex situations is possible due to the fact that data is stored in compatible databases, which can be handled and combined in different ways. At the same time, editing and updating is done more efficiency since there is no need for reprinting or redistributing material, as everybody can access the updated data directly in a computer network.



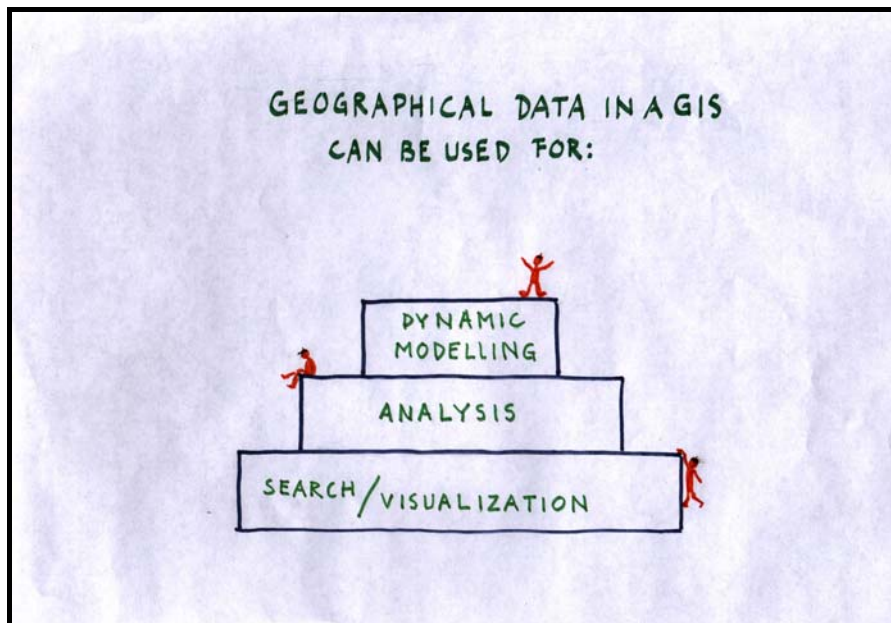
Slide 29

GIS also provides powerful tools for visualising data. GIS software contain tools for professional map-making that makes the communication with decision-makers and other “end users” more efficient. The GIS concept contains a strong component for data sharing, most often in the form of common databases shared by many users in a client-server network (LAN). Independent of the physical location, it is possible for all concerned parties to access and work with the same database at the same time.



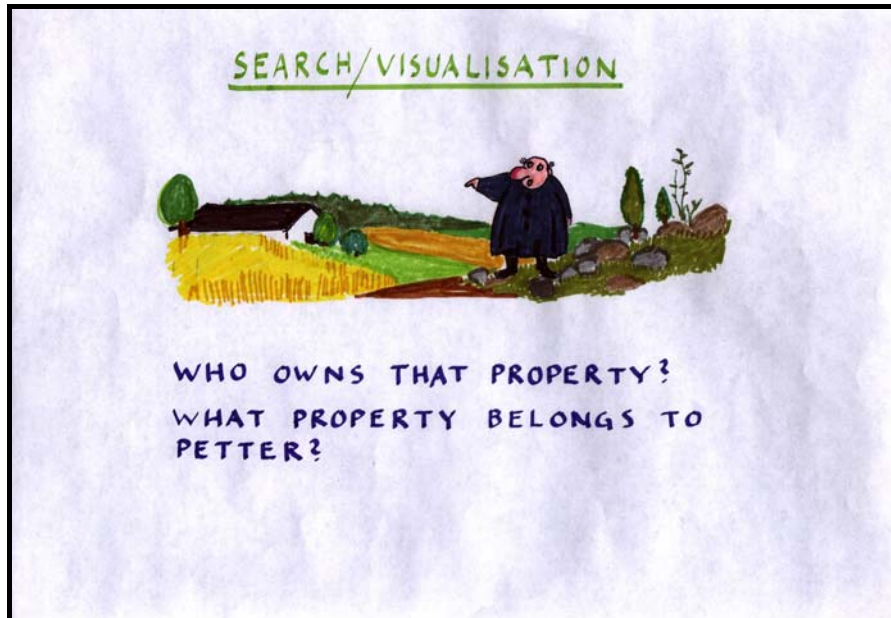
Slide 30

In a way, the quality of the basic data used increases since the use of data becomes more efficient with a digital handling system. However, the original quality of the input data will not increase just because a GIS is used. This is very important to consider – digital technology does not increase the input data quality, but the process, as seen in the GIS-cycle, will have higher quality yielding and higher quality in the analysis and decision-making processes. The GIS also make data available to more people since data dissemination becomes more efficient using digital techniques, e.g. the Internet.



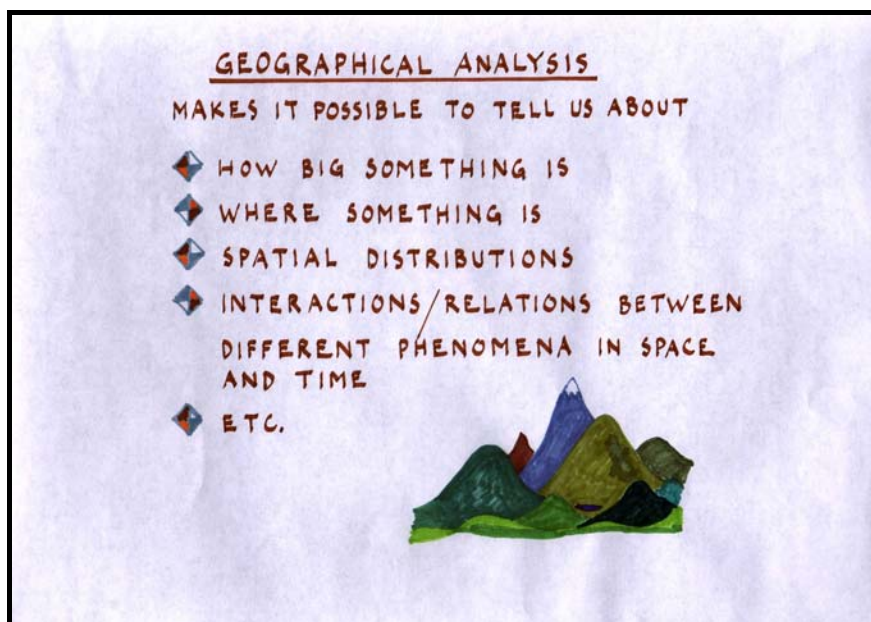
Slide 31

GIS users can be divided into three different categories, depending on their needs and knowledge. The first category, at the base of the pyramid are the people that are using GIS purely as a tool for collecting information from different geographical databases and visualise this information. Within this category, also users working with routine updating are found. This category could do with rather simple GIS software or will work with adapted applications based on more complex software and will not need more than basic understanding of the GIS concept. It is widely believed that this is the bulk part of present and future GIS users. The second category will work with analysis and more advanced visualisation of analysis results, involving some parts of database construction and planning as well. This category will need a professional training in subject/problem specific field combined with a more extensive understanding of GIS. Students attending this course will have sufficient knowledge about GIS to be able to perform such analysis. The third “level” of users is the advanced users that are completely fluent in GIS and can handle all aspects of it. These users are working with advanced analysis and are also the ones responsible for database design and maintenance. They will also be responsible for the construction of applications that can be used by the users at lower levels.



Slide 32

This is an example of the most basic use of GIS. To be able to answer the questions: “Who owns that property?” or “What property belongs to Petter?” illustrates how a user can point to objects in a map database and get the answer from the linked attribute table or select a row in an attribute database and see the location of the corresponding object in the map database. It is believed that many GIS application do not have to be more complicated than this basic “search and visualise” example.



Slide 33


Questions that can be answered by GIS could also be related to object characteristics, such as computing the area of a polygon and the length of a line. Location of a specific object, location of many objects (describing the spatial distribution of a specific phenomena, e.g. location of all polluting industries in a district) is another very common GIS-question. In more advanced analysis, connections, spatial interactions and relationships in time and space are typical examples of GIS applications.

ANALYSIS GIVES US THE POSSIBILITY
TO STUDY HOW GEOGRAPHIC
PLACEMENT
DIFFUSION
FORM
RELATE TO OTHER GEOGRAPHICAL
OBJECTS AND AFFECT PROCESSES
AND FLUXES

Slide 34

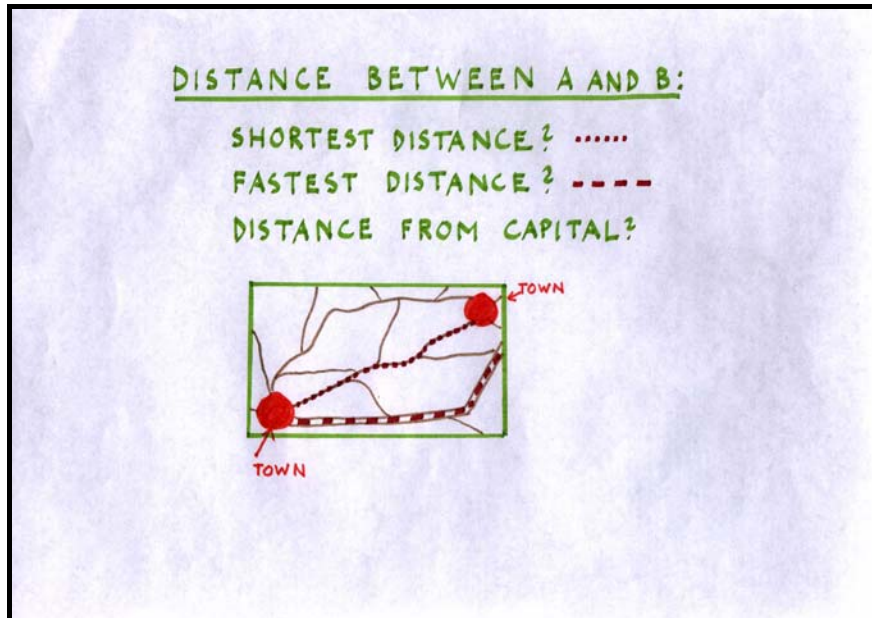
To summarise, GIS analysis answers questions involving geographic placement, diffusion and form, and how these relate between different objects. An example could be placement of polluting industries that influence the distribution of air pollution while wind direction and velocity will affect the form or shape of the cloud of fumes emanating from the individual industries.

EXAMPLES OF QUESTIONS THAT
CAN BE ANSWERED THROUGH
GEOGRAPHICAL ANALYSIS:



Slide 35

The following slides will give some examples of the use of GIS in different applications.



Slide 36

Distances between two points; in this case, two cities, on the map. This could be done in several ways. The simplest way is to measure the “way the crow flies”, that is the shortest distance between the two cities regardless of any connecting road network. The second option is to measure the shortest distance along the road network and a third option to measure the fastest “distance” between them, provided that you have a mean of transportation that allow you to travel at maximum speed permitted on the road network. This type of analysis is called network analysis and is frequently used to solve problems related to travel and transportation. Another type of distance is from a specified location to all other locations, as in the question “How far is it from the bus station to all individual house in a part of a city?”

RELATIONS BETWEEN OBJECTS:

- * HOW MANY FIELDS ARE SITUATED NEXT TO A CERTAIN WATERCOURSE?
- * HOW LARGE ARE THESE FIELDS?
- * NUTRIENT LEAKAGE?



Slide 37

Another example shows questions about relationships between different objects on a map. “How many fields are situated next to a certain watercourse?” is a relevant question when studying nutrient leakage from cultivated areas. Next operation would be to calculate the area of all fields adjacent to the watercourse and then by using attribute information describing fertilisation and crop type, additional information about topography, rain fall, etc, it would be possible to model or estimate nutrient leakage from these fields. To determine which fields that are close to the watercourse and to compute their area is simple. GIS operations while modelling or estimating the nutrient leakage are far more complex and it is easy to realise that the user will have to have profound knowledge about the nutrient leakage process as well as knowledge in GIS-modelling in order to be able to achieve this task.

BUFFER ZONES:

👤 HOW MANY SCHOOLS ARE SITUATED WITHIN FIVE MILES FROM A NUCLEAR PLANT?

👤 HOW MANY CHILDREN LIVE WITHIN ½ KM. FROM A CERTAIN SCHOOL?

Slide 38

To create buffer zones around objects is of great interest for many applications. In this case the distance from a nuclear power plant or the distance from a school are used as examples. Both the school and the nuclear power could be regarded as point objects but buffers could also be generated from line and polygon objects, like roads or lakes. The GIS software will generate the buffer zone according to the users specifications and visualise the buffer zone on the map. It is then possible to select all objects inside the buffer zone and analyse their characteristics.

RELATIONS BETWEEN DIFFERENT DATA LAYERS

CROPS CULTIVATED WITHIN A CERTAIN AREA

CORRELATION BETWEEN INCOME AND EDUCATION

DIFFERENT CROPS

INCOME

EDUCATION

RELATION?

HIGH

MEDIUM

LOW

Slide 39

Spatial relationship between different data layers is another issue of great concern in many applications. The first example is a cultivated area with information about the cultivated crops and the owners of each property. To explain the choice of crops by different owners, characteristics like education and income may be valid and combined to the crop data. In this case all the information is stored in the same geographical database. In the second example, data on education level and income is retrieved from different sources, e.g. Ministry of

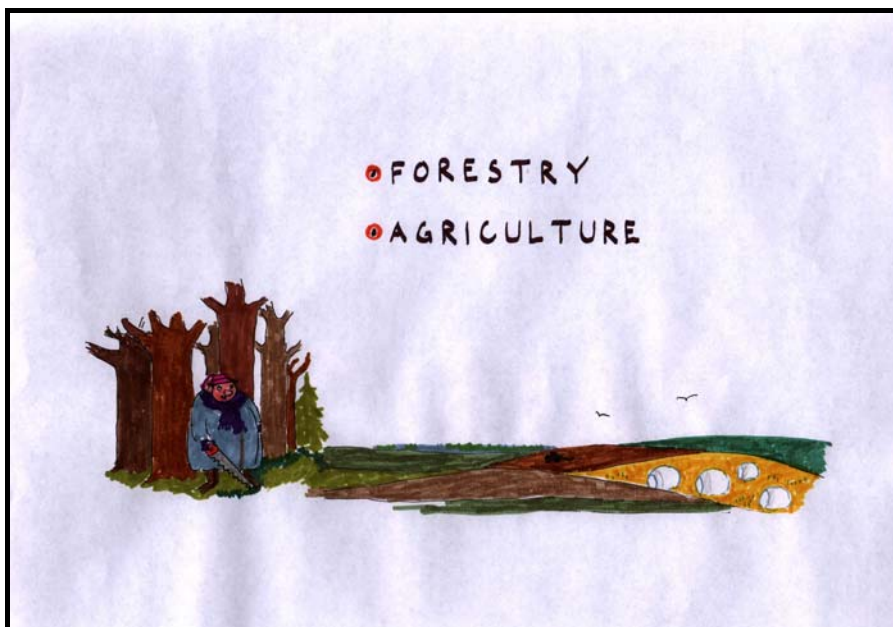
Education and Ministry of Finance. Data from the two sources are combined to see how strong the spatial relation or correlation is between the two.



Slide 40

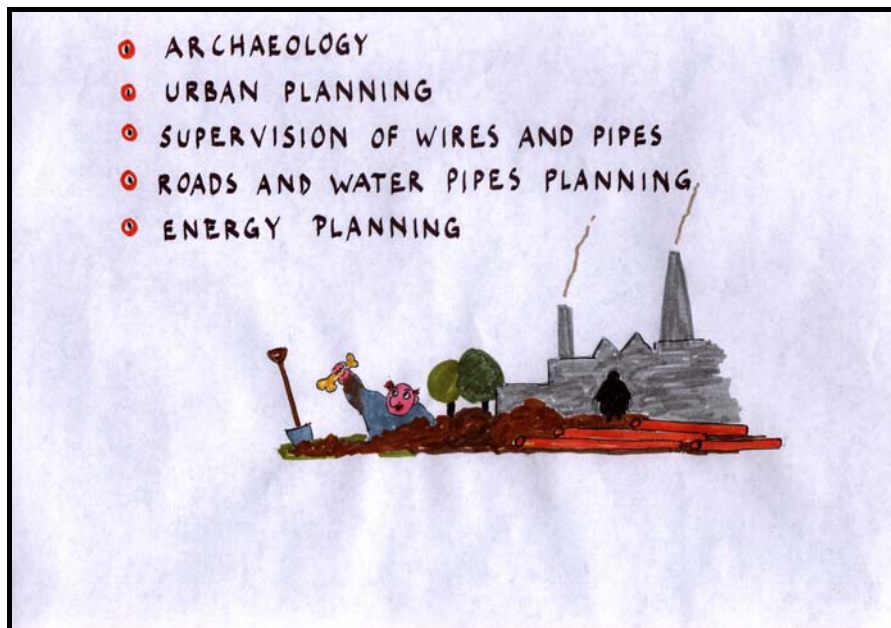
Examples of disciplines where GIS can be useful:

Within social planning and social studies, the GIS capabilities for combining qualitative data with quantitative data opens new possibilities to understand and explain processes in the society around us. Environment analysis and environment supervision are also disciplines that traditionally handle lots of geographical information and consequently benefits enormously from all increase in efficiency in data handling, analysis and presentation. GIS is often used to handle all information and perform risk and consequence analysis, particularly in environmental impact assessments.



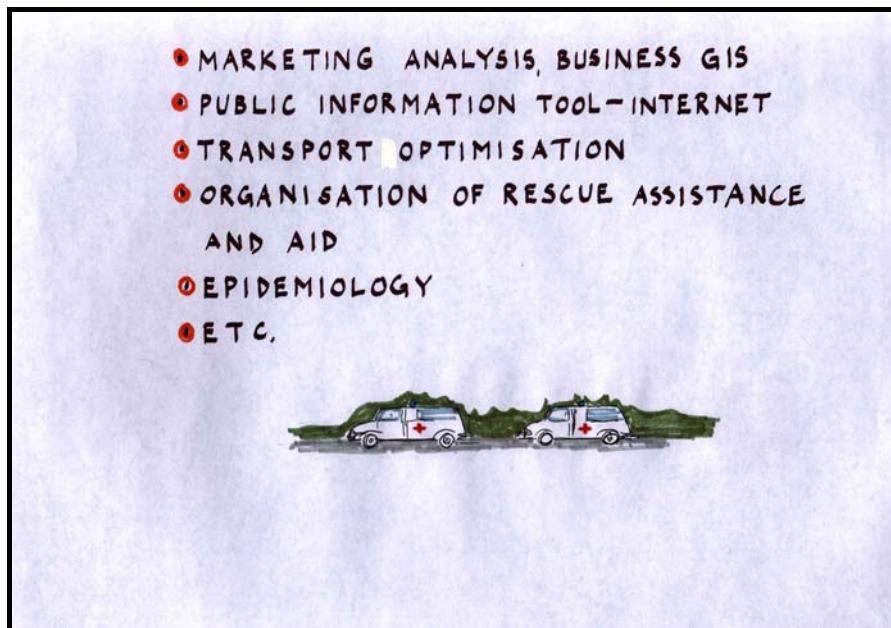
Slide 41

Both the disciplines of forestry and agriculture traditionally work with map data combined with attribute data. Land suitability applications, crops-nutrient interactions, economical analysis and issues related to sustainability and environment are examples of areas where both foresters and agronomists have been using GIS.



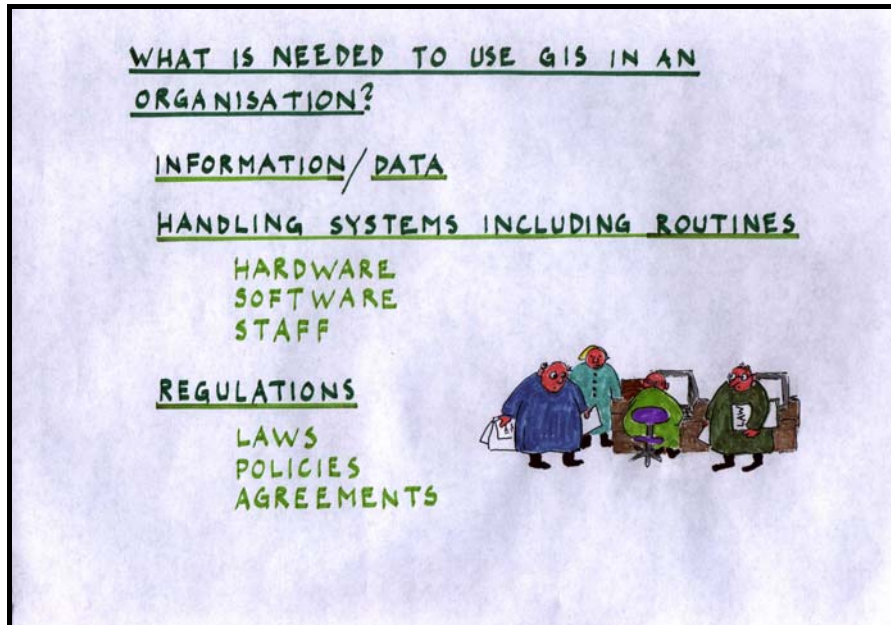
Slide 42

Archaeologists have been using GIS for managing data and particularly to get an overview of spatial relations in their data holdings, when combining archaeological data with data from other sources, such as soil fertility, vegetation development, etc. In all kinds of physical planning, the tendency for using GIS as a replacement for traditional manual methods becomes stronger and stronger. Evidently GIS helps planners at national, regional, local and detailed level make better plans and provide decision-makers with a better base for sound management and decision-making. This is valid for all fields within the planning process, be it urban planning, communication, sewage, energy and others.



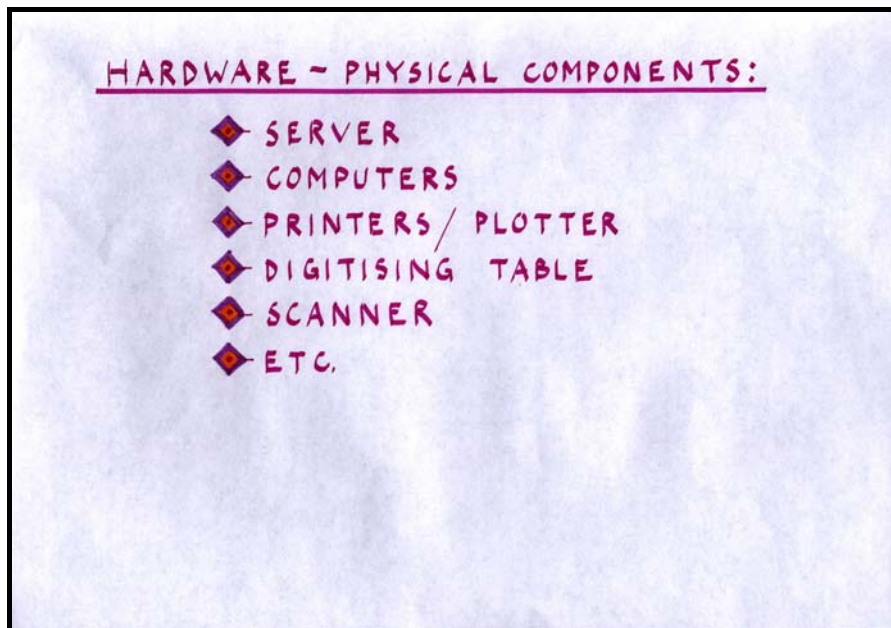
Slide 43

Business GIS is a common name often applied to commercial GIS applications. In marketing investigations, industry and supermarket establishment, banking, etc., GIS is often used for planning and decision-making. A very important application of GIS is in the field of public service and information dissemination to the public. There are many examples how governments and NGO's use GIS Internet applications to communicate with the public. A very important field of application is also transportation logistics, where important amounts of money can be saved if transport cost are minimised. Efficient transportation is also a means for reducing pollution from traffic, decreasing the effect of global warming and generally reducing environmental degradation. Risk analysis and organisation of rescue assistance are also fields where GIS is used to a different extent in different countries. In some countries, all rescue units are equipped with GIS support containing both map and attribute data to optimise assistance to the public. The health sector is also an evident beneficiary of GIS. Spread of diseases and the reasons behind epidemic outbursts of different diseases is not only a concern public health services, but for international aid organisations, etc., as well. Also animal diseases and the veterinaries are using GIS in similar applications.



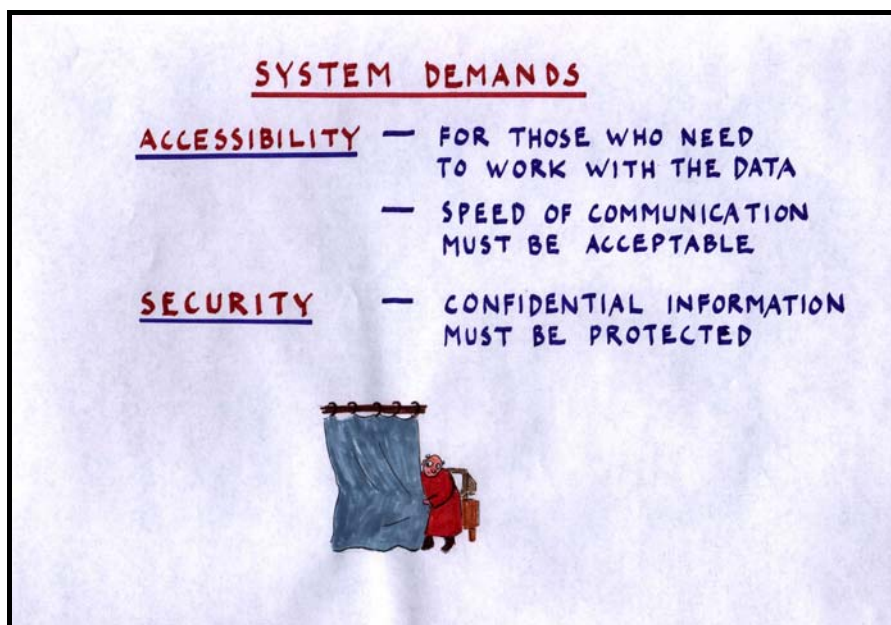
Slide 44

A GIS is composed of several different components and all of them must be present in order for an organisation to operate a full-fledged GIS. Obviously, data and information is crucial since with no information available, there is no need for an information system, geographical or other. This means that the organisation must assure that information is available, by acquiring it from other organisations or by producing it in-house or by a combination of the two. The actual handling system is composed of three components, hardware, software and staff. Neither of these is any major problem today, but often the staff training is the most important issue. Concerning the staff, it is of particular importance to assure that staff training does not stop after an initial introduction to GIS. In order to develop GIS applications, the staff should have opportunities to continuously follow developments of new technology and applications in the field of GIS. Regulations are important since all data handling is happening within the framework of different laws and regulations that control accessibility, copyright protection, ethical considerations, etc.



Slide 45

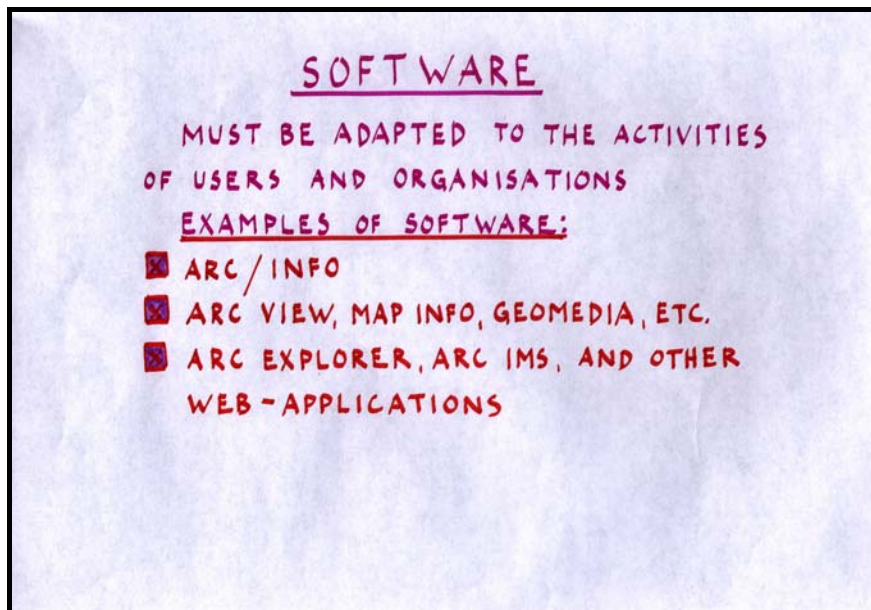
The hardware system needed to operate a GIS depends very much on the extent of GIS use and the complexity of the organisation. In the simplest case, a computer, GIS software and a geographical database is all that is needed. But in most cases both data and software are shared on a server. For presentations, a printer or a plotter is needed, digitising tables and scanners are needed to convert analogue map data and tables to digital databases, etc. It is important to adapt the investments in hardware to the needs of the organisation. All system components need to be connected in a Local Area Network (LAN) if maximum efficiency should be reached in a complex organisation.



Slide 46

Using a LAN involves demands on the system regarding accessibility. Databases should be accessible via a network to those who need access and the access times should be reasonable for the system to be appreciated by the users. In the concept of accessibility, should also be included measures to prevent unwanted use of the information. Only staff with security clearing or appropriate mandates to work with sensitive information (such as information of

very personal nature, information that has important economical significance, etc.) should be able to access that information.



Slide 47

In the past, the question “what software should we use” used to be a very big issue when starting with GIS but today’s software is more or less equivalent when it comes to functionality, independent of what supplier you choose. However, the software often represents a considerable share of the investment costs when introducing GIS in an organisation. To optimise the investment, it is important to consider the needs, the user category and the functionality, since it is rather useless to provide a person with very limited knowledge in GIS an advanced software and vice versa. Another important issue to consider when selecting software is the possibilities to get technical support from national suppliers rather than being dependent on support from organisations overseas.



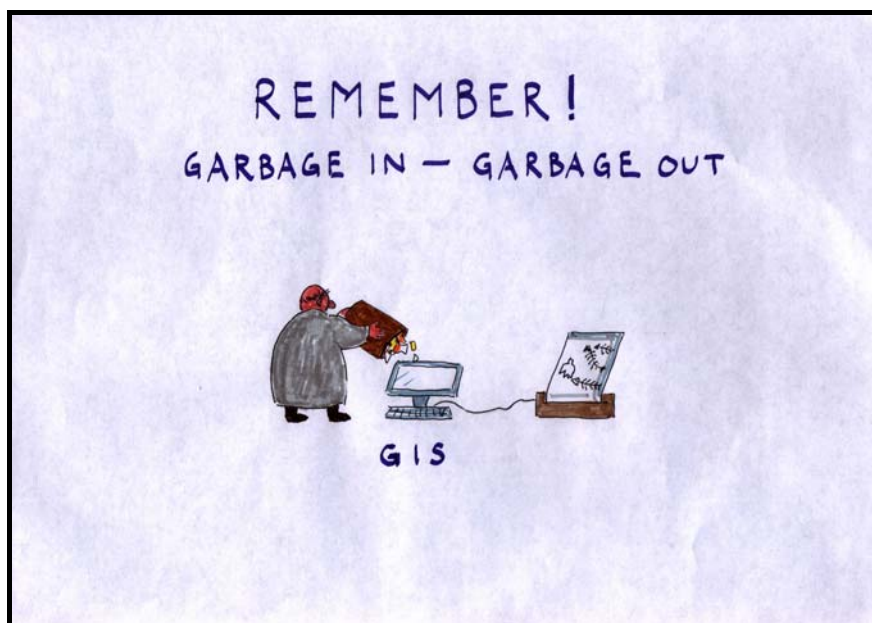
Slide 48

Regulations concerning the use of digital databases could be divided in different categories. There are laws that will inflict penalties if broken and there are agreements that are kept because all participants have reached an agreement in a specific issue. Copyright is an example of a law that protects the owner of a database. This means that the use of the database is restricted; data falling under the copyright law cannot be distributed commercially without the permission of the owner. National and international security will also impact on the use of certain types of information. Failing to follow these regulations will probably lead to prosecution and punishment. Standardisation and data harmonisation are ways to guarantee that the database user/producer communities use the same language when talking about map objects and attribute data, and should also provide for means of quality control, etc. Ethical considerations are agreements that insure e.g. the integrity of individuals. These may or may not be reflected in local laws, if not they are agreements that are kept for the benefit of everybody.



Slide 49

The raw data and the information stored in the databases must fulfil certain demands in order to be usable. A major concern is to assure a reliable data supply. This means that the information must be from a reliable source and the supplier must have a reliable rate of supply. Updates and changes in the database must be delivered in due time for the user to be able to use the information. Data quality must be sufficient for the purposes for which it is intended. If the quality is not guaranteed at a certain level, the database is useless. The user must have access to quality statements and have means to verify data quality. Data that is used should be relevant for the application and accessible.



Slide 50

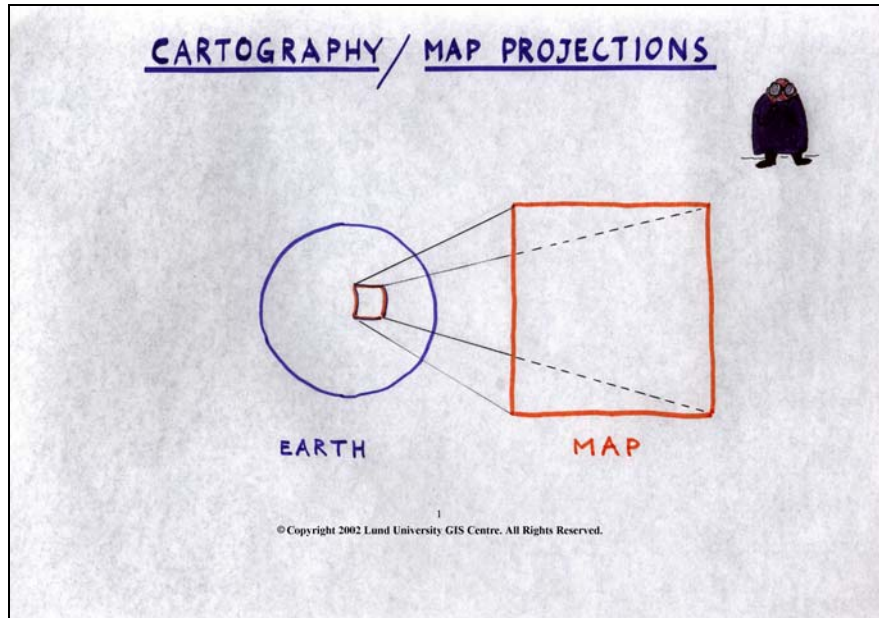
Regarding data, the basic rule is that GIS, how sophisticated it may be, cannot change the fact that the system is never better than its weakest link. If a GIS is fed by garbage data, the output from analysis and other operations will also be garbage!



LUND
UNIVERSITY

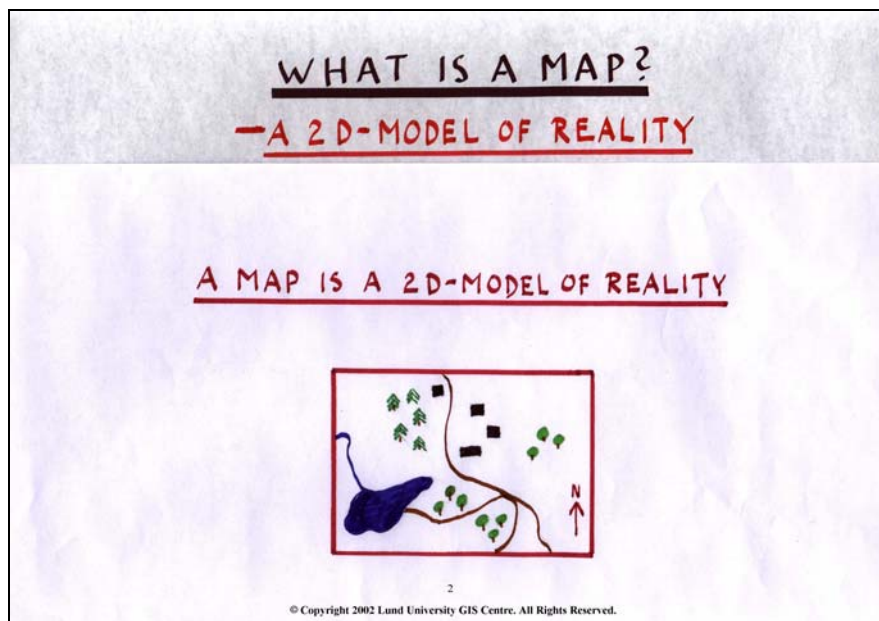


CARTOGRAPHY AND MAP PROJECTIONS



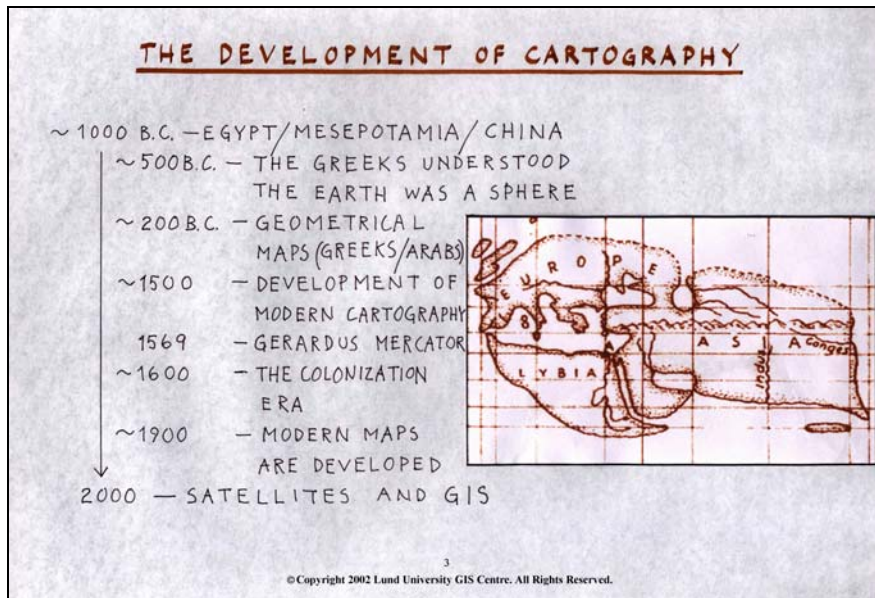
Slide 1

This lecture is about cartography and map projections. The basic problem with cartography and projections is that we have something that is round (the Earth) and we want to describe it as a flat surface (a map). This is a very complicated and mathematical subject, so we will just describe it briefly and try to keep it as simple as possible.



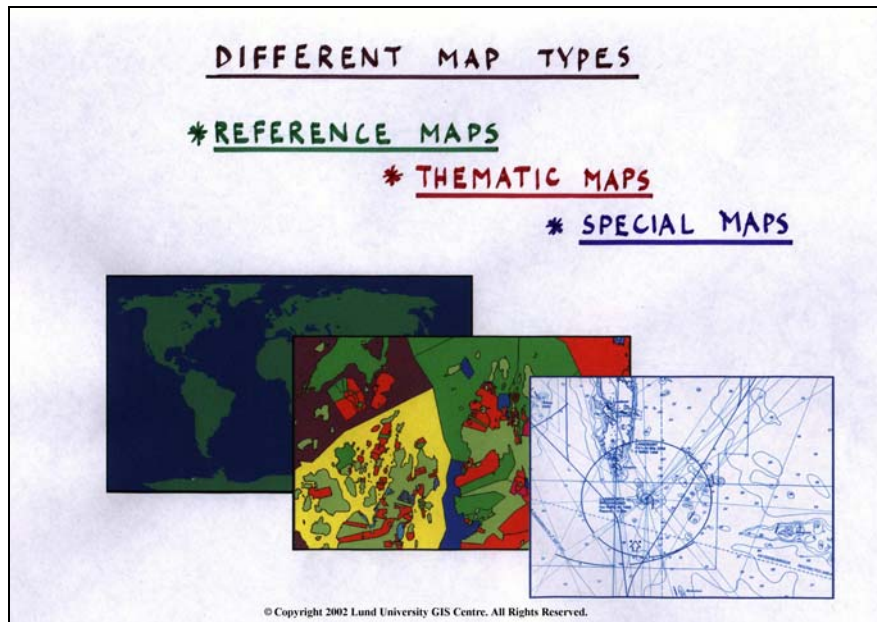
Slide 2

As we mentioned in the introduction chapter, a map is a two-dimensional model of reality and the problem is to create this two-dimensional model out of our three-dimensional reality.



Slide 3

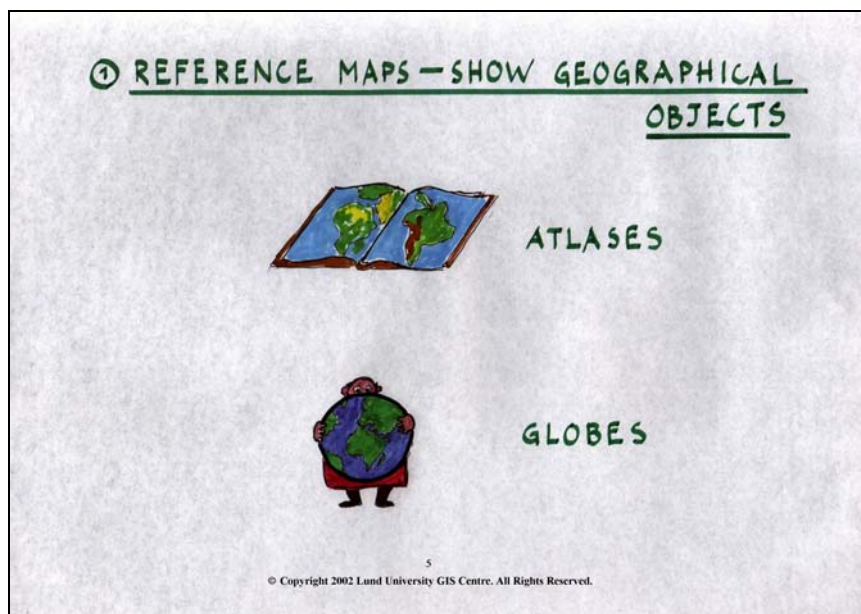
Cartography has interested mankind for a long time. The ability to describe the environment and visualise distances and routes has been essential for communication and travelling for people all over the world. The semi modern cartography started many thousand years ago in Egypt, Mesopotamia and China, where they began to produce maps. At around the time of 500 years B.C, the Greeks understood that the earth was round and this was an essential development for cartography since people earlier had been describing the earth as a flat surface. At around 200 B.C, the Greeks started to create more accurate maps of North Africa and in Europe, around the Mediterranean. They even tried to estimate the length of the equator, which they managed to do quite well. This is amazing considering the tools they had at this time. During the Middle Ages, the field of cartography in Europe suffered from a setback caused by the conservative church, which did not allow the field and studies of cartography to develop. It was not until the renaissance began, in the 16th century, that the modern cartography really started to develop. One very important person in the development of cartography was Gerardus Mercator, who lived in the 16th century. Then cartography progressed fast from the 16th to 19th century, but the really modern cartography didn't develop until about a hundred years ago. Now days we are quite skilled in this area, and can do very complicated, as well as simple, maps. We are often using GIS and satellite images in map making, and we get better and better instruments making it able to describe the earth with higher and higher accuracy.



Slide 4

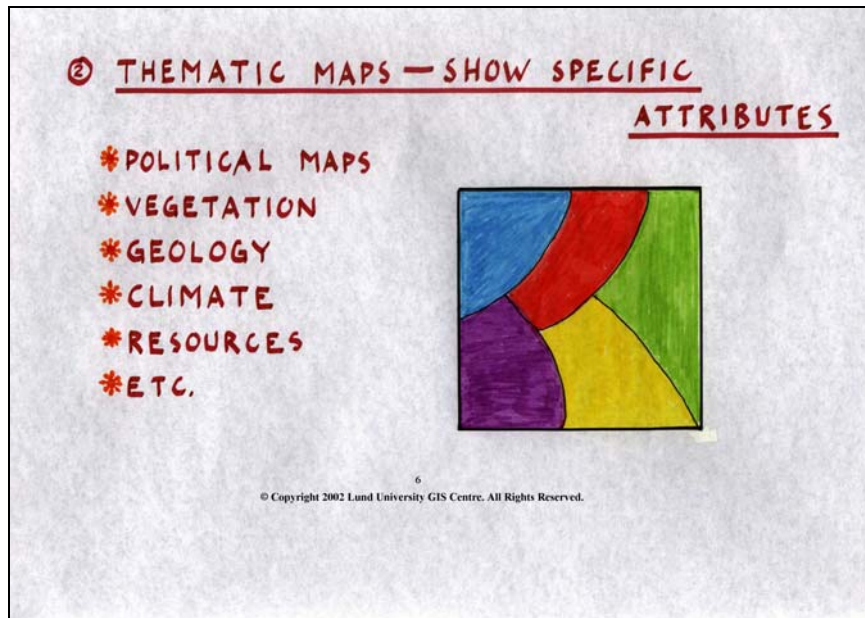
Before going into the subject of how to make a map, we would like to point out that there are different kinds of maps. For example:

- Reference maps, which are more generalised maps.
- *Thematic maps, which are maps showing a particular field of interest.*
- *Special maps, which can be described as maps covering a special field of cartography.*



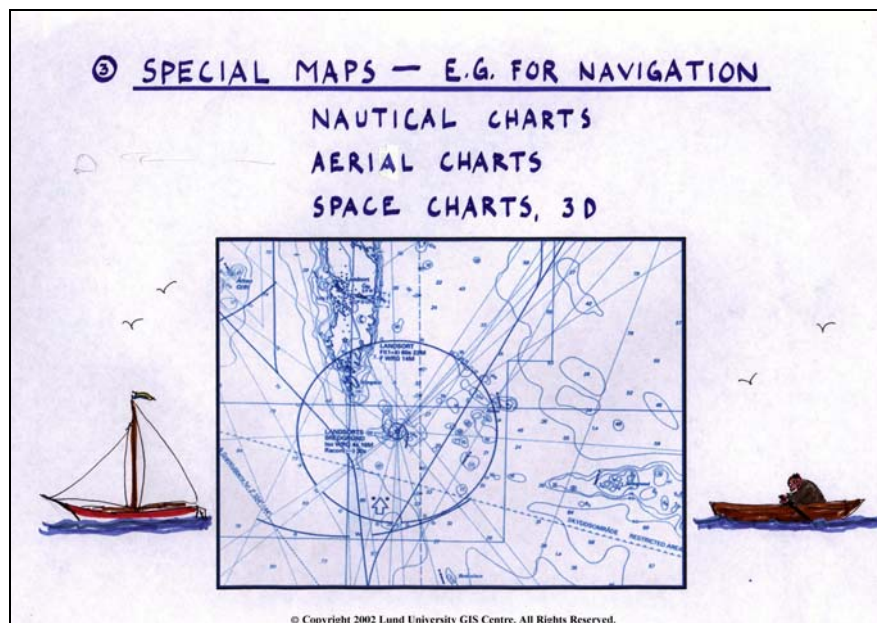
Slide 5

Reference maps give us a broad understanding of the environment around us. Atlases and globes are typical reference maps, which show a lot of geographical objects at the same time.



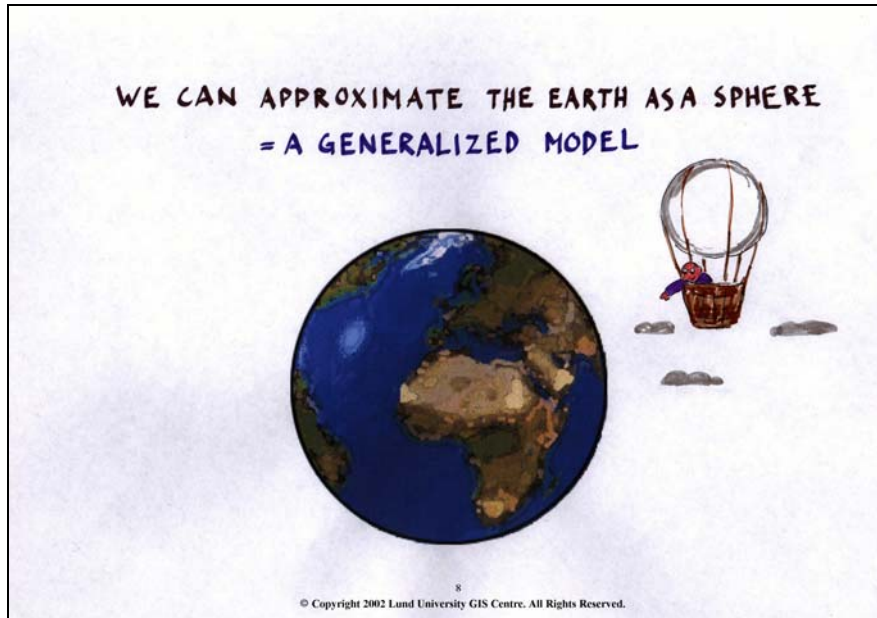
Slide 6

Thematic maps are used to visualise specific features. These could be anything of interest, e.g. political borders, population, vegetation, and sanitation. These maps are often visualised in a local or regional scale. In the slide above, the different colours might, for example, show us different sorts of vegetation. Yellow could be agricultural land, brown the forest and green some sort of grassland. The thematic maps are for sure the most common map type that we are working with in the field of GIS.



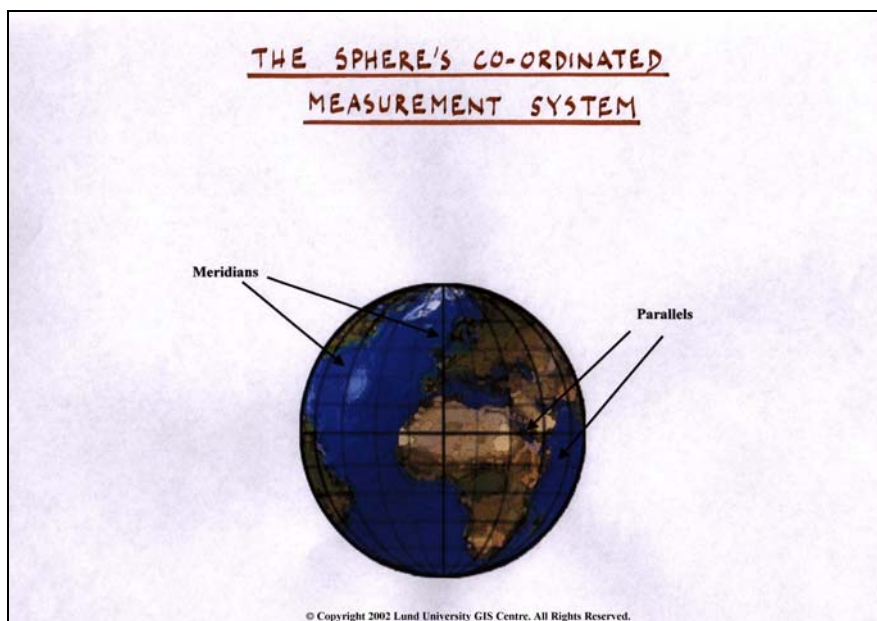
Slide 7

Special maps are commonly used for navigation, for example, nautical charts, aerial charts and for navigation in space in three dimensions. As mentioned above, these maps are often quite specialised. We are not going to go further into these during this course.



Slide 8

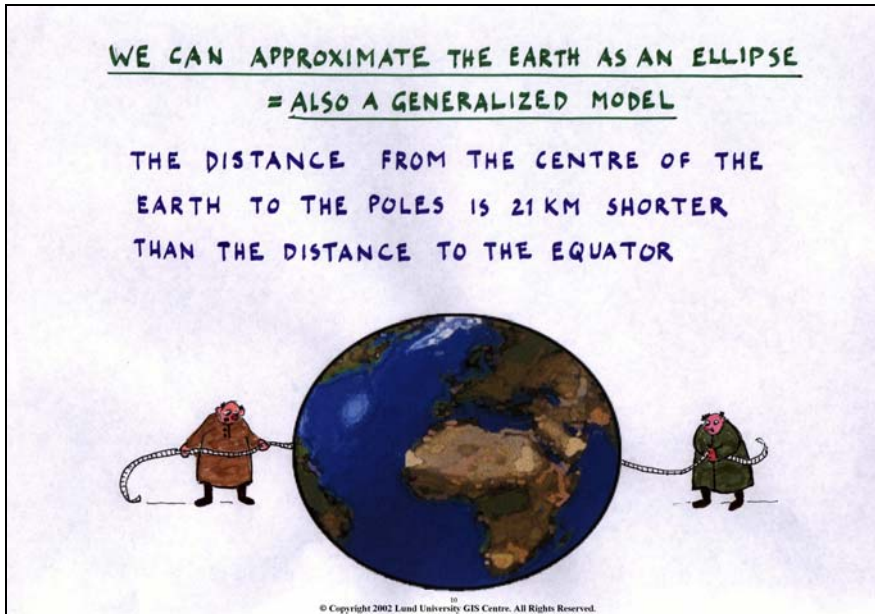
We can approximate the earth as a sphere. Note that this is only a generalised model. The earth is not really round but it's nearly round. When we want to make a map, we have to approximate the earth into some sort of generalised shape, and one of the possibilities is to approximate it into a sphere.



Slide 9

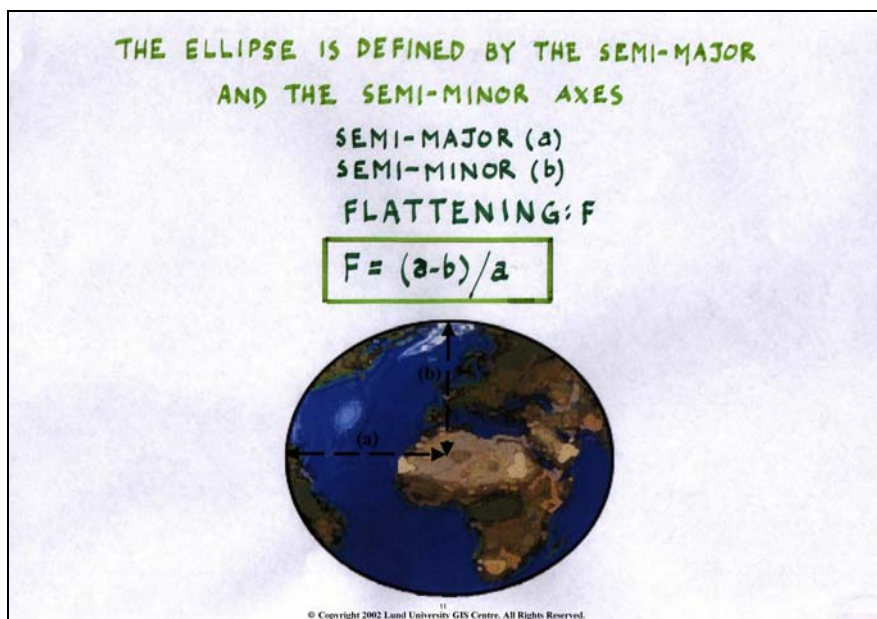
To locate something on this sphere, we need some kind of co-ordinate system, preferably with an x- and a y-axis, perpendicular to each other. However, this is impossible since the earth is round. This means that the co-ordinate system that we use on the sphere could never be right angled. Instead it's based on meridians and

parallels. The largest parallel is the one around the equator, while the others are decreasing in length the further north or south they are. The meridians on the other hand, will always meet at the poles, which means that the distance between the meridians will be quite big at the equator and increasingly narrow the further they move towards the poles.



Slide 10

The shape of the earth is more commonly approximated as an ellipse. This is also a generalised model since the Earth is neither a sphere nor an ellipse. Even if the ellipse approximation is truer than the sphere, it is not perfect. The true Earth is irregular. The ellipse form means that the distance between the two poles, in the north-south direction, will not be the same as the distance along the equator, in the east-west direction. Actually, the distance from the centre of the Earth to the poles is 21 km shorter than the distance to the equator. This irregular form is due to the rotation of the Earth. The gravitation flattens it at the poles.



Slide 11

The flattening of the earth at the poles is measured by using the lengths of the semi-major and the semi-minor axes. The semi-major axis is the longest one and stretches from the earth's centre to the equator and the semi-minor is the shorter axis and stretches from the earth's centre to the poles. To calculate the flattening, you use

the difference between the semi-major and the semi-minor axis and then divide it with the distance of the semi-major one. Several different ellipses, with different flattening can be used. This means that two map are not equal if they do not use the same ellipse approximation.

DIFFERENT ELLIPSOIDS

SINCE THE SURFACE OF THE EARTH IS UNEVEN,
DIFFERENT ELLIPSES ARE MORE OR LESS IN ACCORDANCE
WITH DIFFERENT PARTS OF THE SURFACE OF THE WORLD

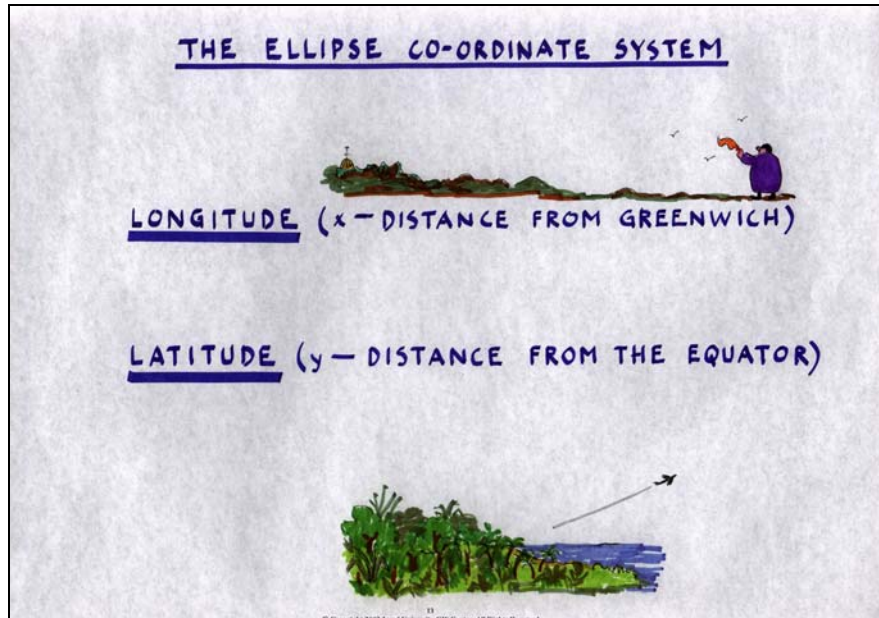
EXAMPLES:

◆ EVEREST 1830	— SOUTH ASIA
◆ CLARKE 1866	— USA
◆ BESSEL 1841	— SCANDINAVIA
◆ KRASOVSKY	— RUSSIA, EAST EUROPE
◆ WGS 84	— USED FOR GPS-SYSTEMS

12
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

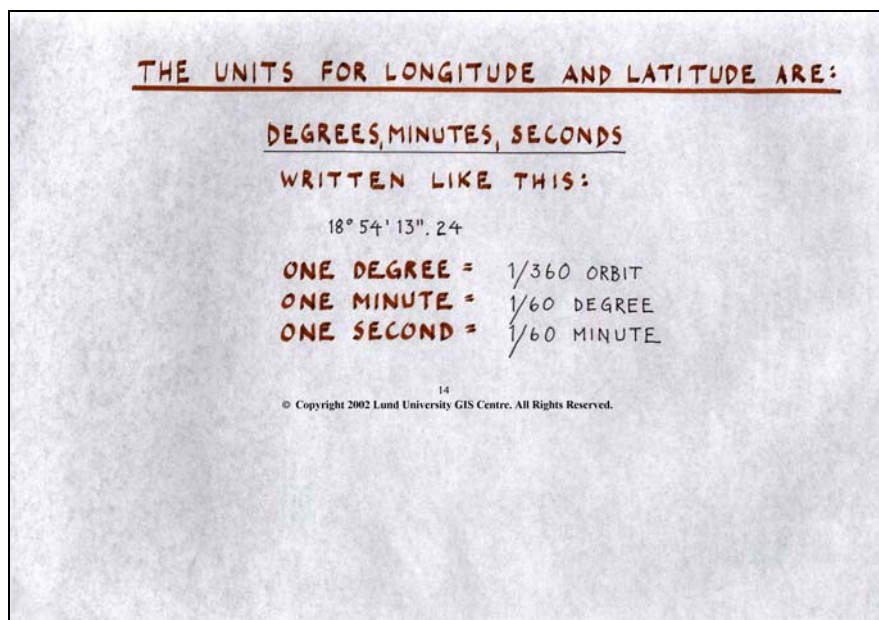
Slide 12.

These different ellipse approximations, with different grades of flattening, result in geometrically different maps. An ellipsoid is a three-dimensional body. Since the surface of the Earth is uneven, different ellipses are more or less in accordance with different parts of the Earth's surface. Different cartographers have suggested different ellipsoids that are better at certain places at the Earth's surface. It's very important to remember that in different parts of the world, we use different sorts of ellipsoids. One ellipsoid with a certain flattening suits perhaps best in China, while it might cause large errors in, for example, South America. This is the reason why we have different ellipsoids. In the slide above you find a few examples of different ellipsoids and in what regions they are preferably used. However, many more ellipsoids are used. Most of the different ellipsoids are named after the cartographer who created them. One exception is WGS 84, which was created in 1984 in an attempt to create some sort of standard ellipsoid. It is used in the GPS-system that we will go further into in another lecture.



Slide 13

The co-ordinate system for the ellipsoid is quite similar the one for the sphere. To determine a position on the ellipsoid we use the terms longitude and latitude. Longitudes are measuring the position along the “x dimension” and they stretches between the north and south poles. The distances for the longitudes are measured from Greenwich, which is a small town outside London. Not so long ago, different countries had different origins for their longitudes, which complicated the comparison of maps, but later on it was internationally agreed that the position of Greenwich should be the origin of the longitudes, which has eased exchange of geographical data between different countries considerably. The Latitudes are measuring the distance from the equator, so they can either be positive on the north hemisphere or negative on the south side of the equator.



Slide 14

How do we measure longitudes and latitudes? Normally we use degrees, minutes and seconds. One turn around the Earth is 360° . This means that one degree is $1/360$ turn of the full orbit. The minutes are calculated using the same principal as you are estimating time. If you divide an hour by 60 you'll get a minute and if you divide a degree by 60 you get a longitude/latitude minute. When calculating seconds, you once again use the same principal. These quantities are expressed using the $^\circ$ sign for degrees, the $'$ sign for minutes and the $''$ sign for seconds.

SOMETIMES DECIMAL DEGREES ARE USED

$$18^{\circ} 15' 13''.7 = 18.2538^{\circ}$$

$$15/60 = 0.25$$


-15 MINUTES EQUALS 0.25 DEGREES

$$13.7/60 = 0.228$$

-13.7 SECONDS EQUALS 0.228 MINUTES

$$0.228/60 = 0.0038$$

-0.228 MINUTES EQUALS 0.0038 DEGREES

$$0.25 + 0.0038 = 0.2538$$


15
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

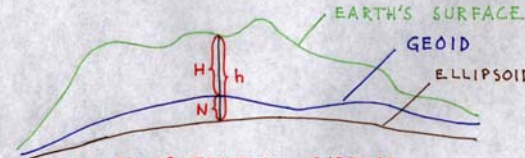
Slide 15

This is an example of how to recalculate longitude and latitude positions from degrees, minutes and seconds to decimal degrees, which is much easier for computerised technologies to use. Normally, in GIS, we are using decimal degrees.

OFTEN WE REFER ELEVATION TO THE GEOID

THE GEOID'S SURFACE COINCIDES WITH THE MEAN SEA-LEVEL AND ITS HYPOTHETICAL EXTENSION IN UNDER THE CONTINENTS

THE GEOID'S SURFACE IS PERPENDICULAR TO THE DIRECTION OF THE FORCE OF GRAVITY



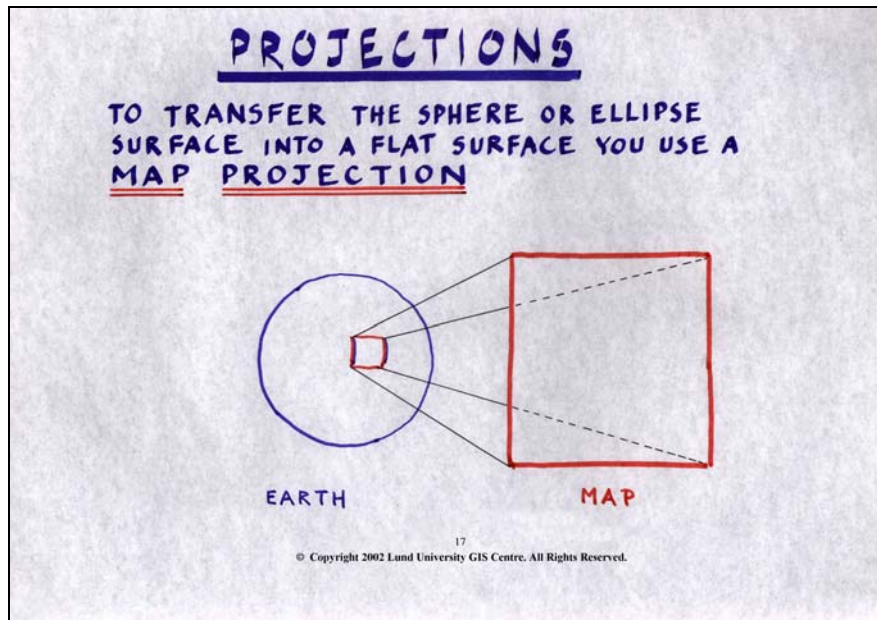
AN EQUIPOTENTIAL SURFACE
H= HEIGHT (ABOVE SEA-LEVEL)
h= ELLIPSOID HEIGHT
N= GEOID HEIGHT

16
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 16

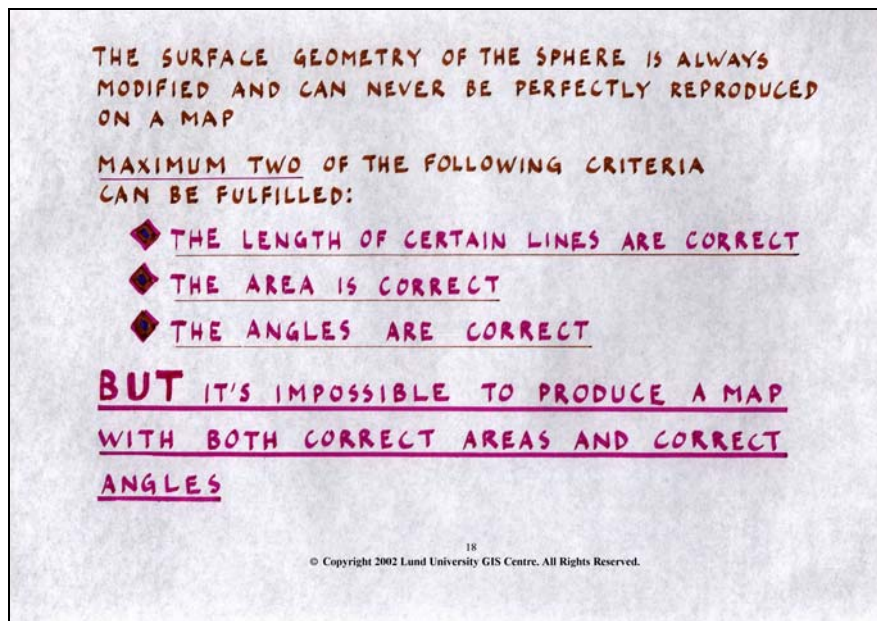
Apart from the sphere and the ellipsoid, we can also refer our maps to the geoid. This is a surface that coincides with the mean sea level and its imagined extension under the land surface. The geoid is more or less the true surface of the earth. This approximation is hence more accurate than the approximation of the Earth as a sphere or ellipsoid since it also takes the gravity, which is different in different parts of the world, into account. It should be noted that, depending on the surface we are referring to, we get “different” elevation values at the same point. We can measure elevations as the height above sea level (H), which is the same

distance as the elevation above the geoid, as the ellipsoid height (h) or as the geoid height (N). It is more common to refer the elevation as the height above sea level (H).



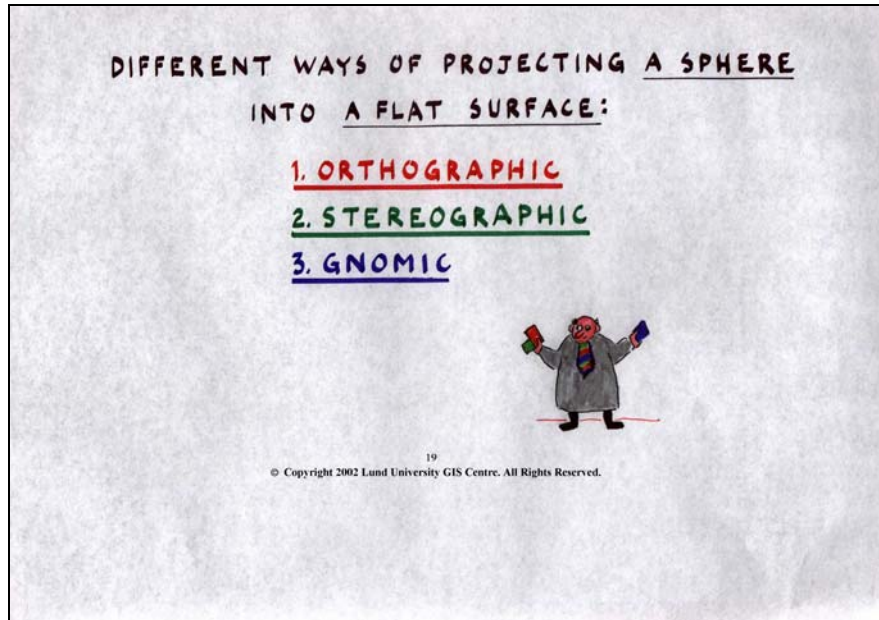
Side 17

When we have solved the problem of transferring the earth's shape into something that we can describe more mathematically and scientifically, we are still facing the problem of how to project our almost round surface into a flat map. To solve this problem we use different map projections.



Slide 18

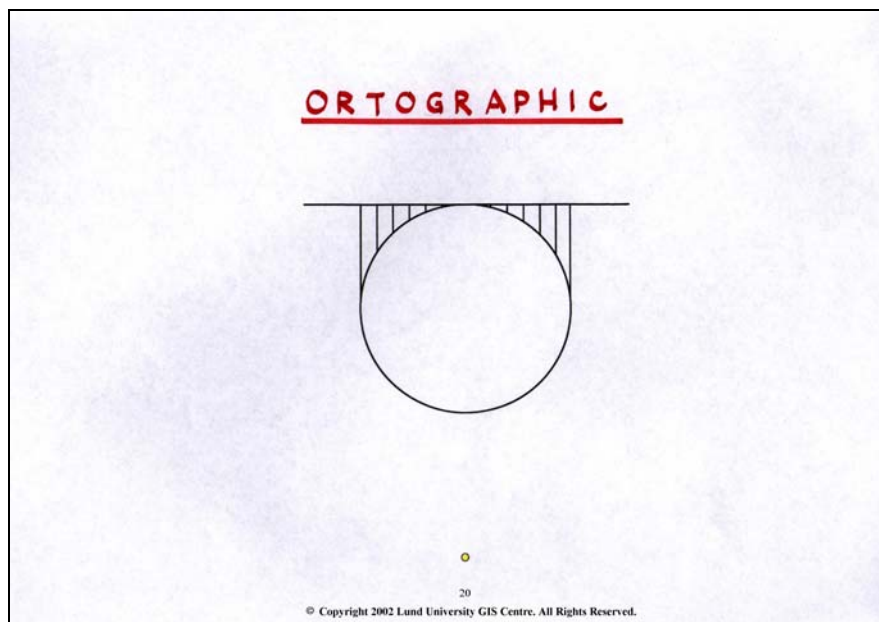
The surface of a sphere can never become completely reproduced on a flat surface, but must be more or less modified to be presented in two dimensions. When making a map, you must choose what parameters you want to have correctly described on the expense of others. You can either have correct length, areas or angles and the choice of which ones to use depend on the purpose of the map. It's impossible to have all three of them correct on one map but you can, in some extent, combine two of them more or less well. You can however never combine areas and angles so that they are both are correct at the same map.



Slide 19

When we are about to project the earth's surface onto the flat surface (the map), we need to think of this as if we were using some sort of light, like a projector. To do this, we can choose between three different sorts of projection types:

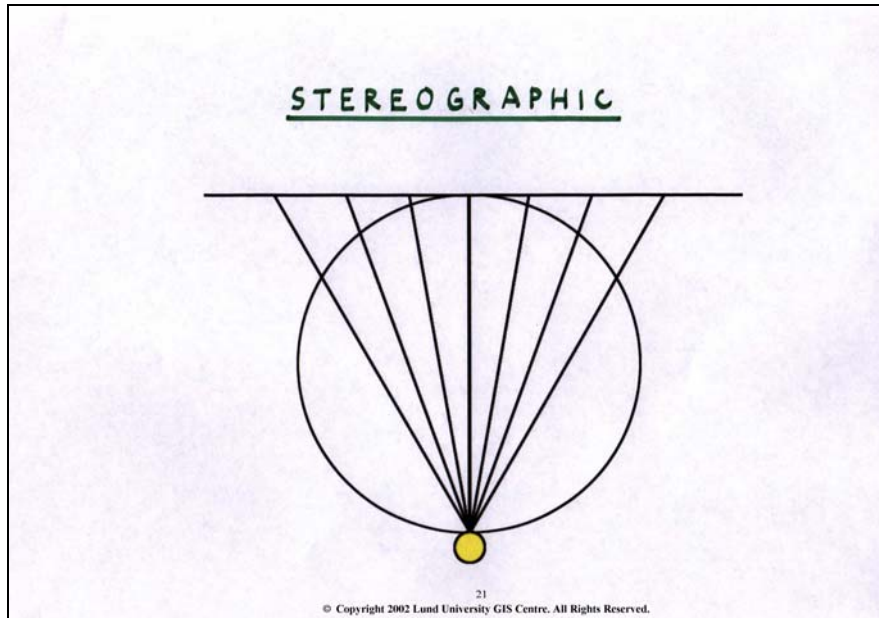
1. Orthographic
2. Stereographic
3. Gnostic



Slide 20

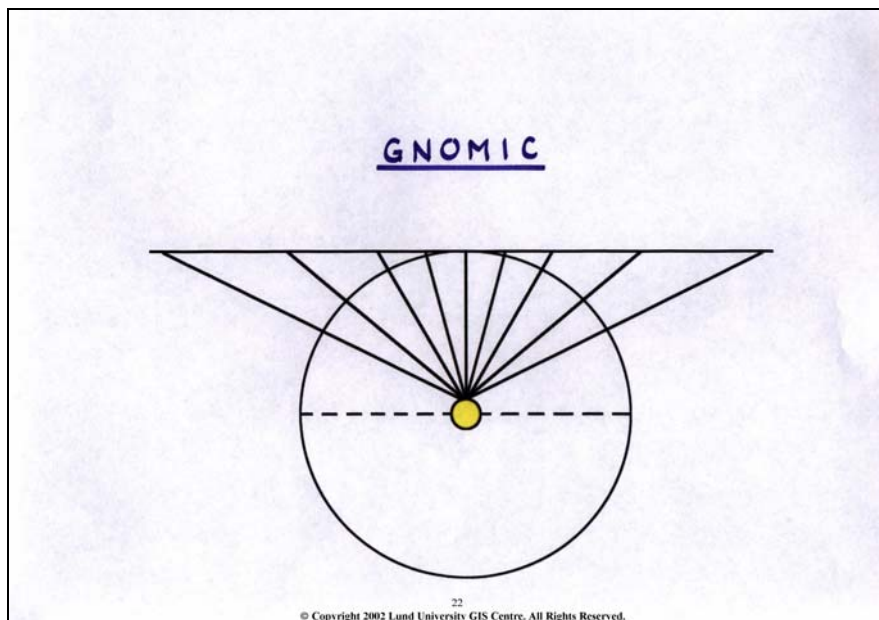
To project something with the orthographic projection, you must think of the imagined light source as being indefinitely far away from the earth. When you then illuminate the earth, the beams of light will strike the

earth as parallels. The earth will then be projected on the flat surface according to where each of the light beams touches the earth. Each of these beams can just touch the earth at one single point.



Slide 21

The stereographic projections put the imagined light source on the Earth surface, on the opposite side of the area that is going to be projected. This means that the light beams will no longer be parallel since they are diverging from the source of light. This causes a quite different image of our surface than the orthographic projection.



Slide 22

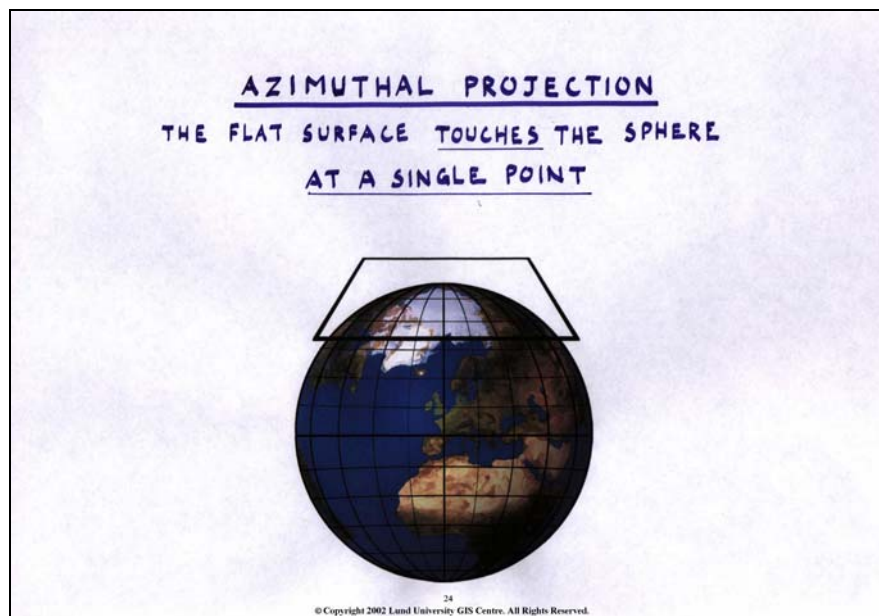
The gnomonic projection puts the source of light at the centre of the Earth, which causes the light beams to strike the Earth's surface almost at right angle. The errors will be rather large along the borders of the map.



Slide 23

Projections are not only based on the different ways of the transference of the earth's surface to the map, but also of the "shape" of the map that we are to transfer them to. There are three different "map shapes":

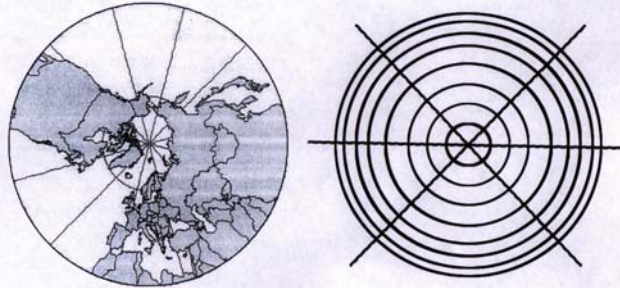
- Azimuthal
- Conic
- Cylindrical



Slide 24

When using the azimuthal projection, the flat map surface touches the sphere only at one single point. In this example it is the North Pole but it could be anywhere on the earth's surface.

- MOSTLY USED FOR THE POLAR AREAS
- THE MAP BECOMES DEFORMED NEAR THE EDGES
- NEITHER CORRECT ANGLES NOR AREAS
- EXAMPLE: LAMBERT'S PROJECTION



25
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 25

These maps are quite often used in the polar areas. They are most accurate in the centre of the image and become more and more erroneous near the borders. They have neither correct angles nor areas. One example of this kind of map is Lambert's projection.

CONIC PROJECTION
THE CONE TOUCHES THE SPHERE ALONG A
PARALLEL

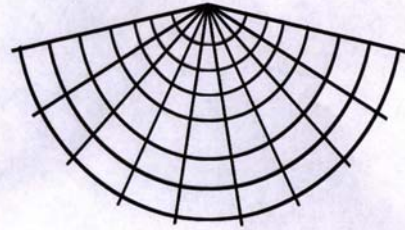
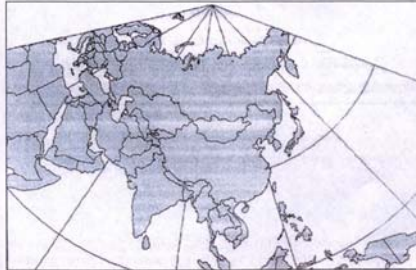


26
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 26

The conic projection consist of a map “wrapped” as a cone around the Earth, and touches the surface in more than one point, along a line as a parallel. The areas in the map that are near this line become more accurate than the areas further away from it. The cone does not have to be placed over one of the poles, but can be put anywhere. It is also possible to let the cone “sink” into the sphere, resulting in two lines touching the surface.

- CORRECT WHERE THE CONE TOUCHES THE SPHERE
- DEFORMED AT HIGHER LATITUDES
- CORRECT ANGLES
- EXAMPLE: ALBERS'S PROJECTION



© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 27

This means that this projections causes deformation in areas both north and south of the line where the cone touches the Earth. This sort of map gives you correct angles and is therefore suitable for navigational charts, where they are most commonly used. In GIS, we do not handle these sorts of maps very often. An example of a conic projection is Albert's projection.

CYLINDRICAL PROJECTION

THE CYLINDER TOUCHES THE SPHERE ALONG
A PARALLEL

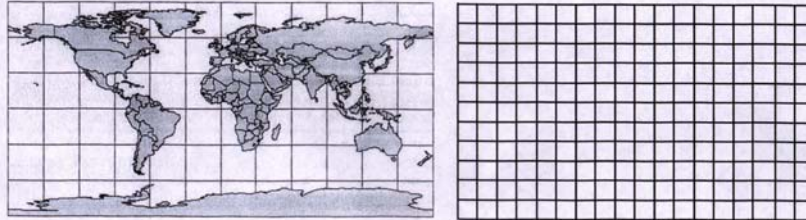


© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 28

The most common projection is the cylindrical one. The map is wrapped as a cylinder and placed over the Earth. It touches the sphere along a specific line/parallel. In the example above the map touches the equator.

- COMMON ON NATIONAL/INTERNATIONAL MAPS
- CORRECT AREAS
- EXAMPLE: MERCATOR'S PROJECTION



© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 29

The cylindrical projection is very commonly used for national and international maps. The main advantage is that it gives correct areas along the line where it touches the sphere, which makes it useful for area measurements. One of the most common cylindrical projections is the Mercator projection, which is often used for world charts. The areas are most correct along the equator but become greatly deformed further north or south.

UNIVERSAL TRANSVERSE MERCATOR

- ▶ A TRANSVERSE PROJECTION IS CHARACTERIZED BY A HORIZONTAL CONE OR CYLINDER
- ▶ UTM (UNIVERSAL TRANSVERSE MERCATOR) IS ONE OF THE MOST WELL-KNOWN TRANSVERSE PROJECTIONS

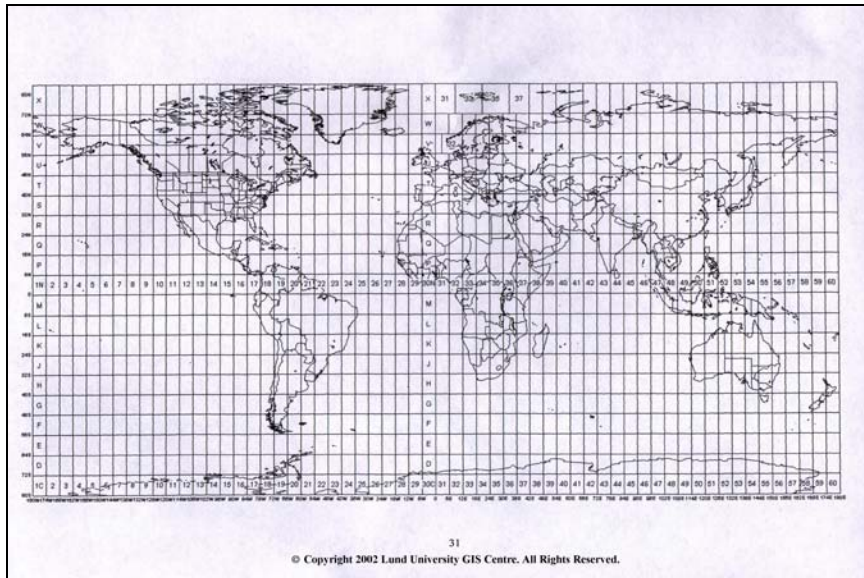


© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 30

The most famous cylindrical projection is the Universal Transverse Mercator, also called the UTM projection. Instead of a standing cylinder, we here have a lying cylinder that gives a good fit between the sphere and the map in a north-south direction. In the example above, we would have a good map in central Africa, central Europe and the Scandinavian countries while the map would be really bad in Japan or North/South America.

However, if we turn this cylinder so it touches the sphere along another meridian, we could then have a much better map for South America or Japan.



Slide 31

Because of the possibility to turn the transverse cylinder, the Earth has been divided into 60 different zones. These correspond to 60 different directions of this cylinder, 6 degrees apart. Most maps covering larger areas now days are using these projections.

TRANSFORMATIONS BETWEEN
DIFFERENT PROJECTIONS:

MATHEMATICAL: MATHEMATICAL DESCRIPTION OF DIFFERENT PROJECTIONS

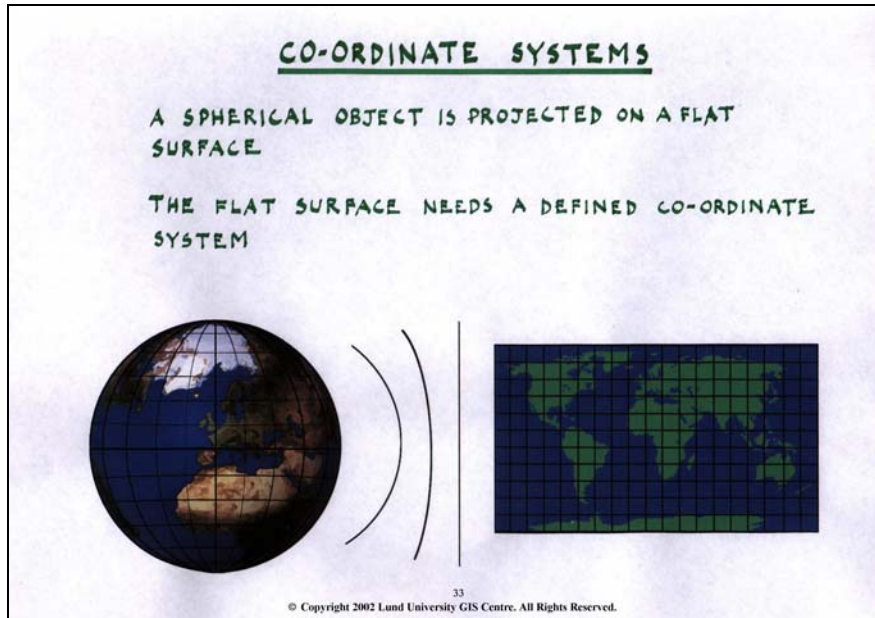
EMPIRICAL: BASED ON EMPIRICAL MEASUREMENTS OF THE DIFFERENCE BETWEEN TWO CO-ORDINATE SYSTEMS

32
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 32

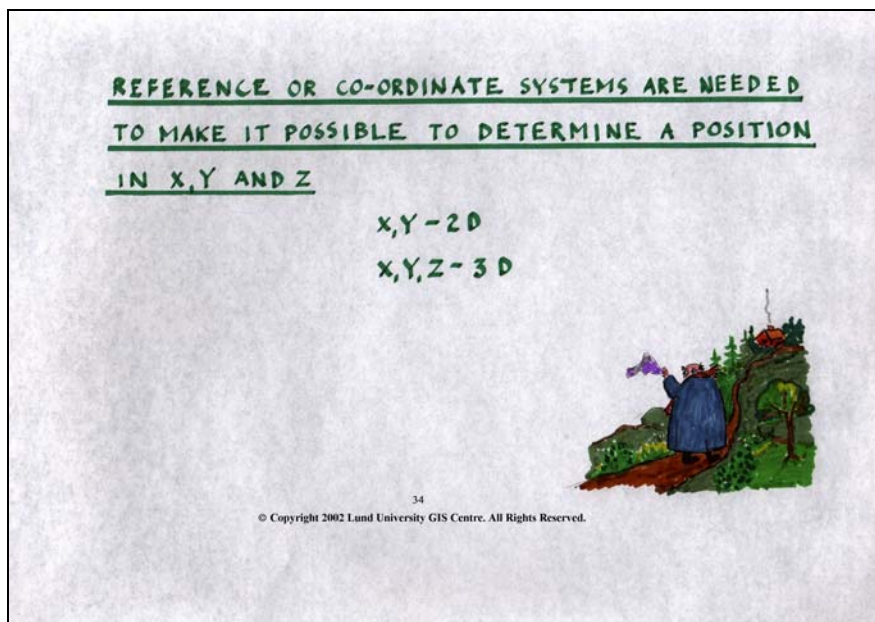
While working with GIS, it could happen (and quite often does) that we get maps that are made in different projections. In order to use them together, they have to be geometrically equal, and we have to be able to transform between different projections. This is often a very complicated mathematical operation. However, most projections are mathematically defined, making it possible to write programs to transform one into another. In most GIS software now days there is some sort of transformation tool that helps the user change map projection. Still, making maps perfectly compatible to each other remains a major difficulty in GIS today. It sometimes even happens that the user does not know what kind of map projections the different

maps he or she works with originate from. We then have to identify the same spatial objects in the maps to be fitted, and then let the computer empirically transform one map to another. Depending on the maps and the software, this will work more or less well. It is also a very time consuming operation, which might induce errors.



Slide 33

When transforming the round surface into a flat map we still need some kind of co-ordinate system (preferably right angled) to locate an object on the map.



Slide 34

We can describe a point's location either in a two-dimensional coordinate system, using x- and y-coordinates or in a three-dimensional system taking the elevation into account. Normally, when working with GIS, we are using the two-dimensional ones.

A REFERENCE SYSTEM IS A DESCRIPTION OF HOW THE TRUE LAND SURFACE IS RELATED TO THE ELLIPSOID

THIS ADJUSTMENT IS LOCAL AND INFLUENCES CO-ORDINATES IN ALL DIMENSIONS (X,Y AND Z)

THE REFERENCE SYSTEM IS DEFINED BY THE CO-ORDINATE/ MAP POSITIONS (X,Y AND Z) OF A NUMBER OF POINTS ON THE TRUE LAND SURFACE

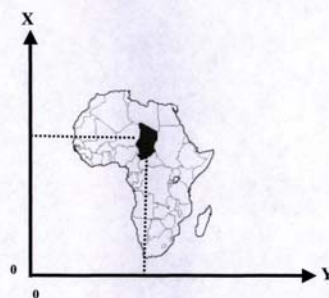
35

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 35

As mentioned above, if we are approximating the Earth using an ellipsoid it does not fit exactly into the Earth's surface. This means that we need some sort of description of how the true land surface is related to the used ellipsoid. We have to move the ellipsoid in all three-dimensions to make it fit our surface as good as possible. To do this we use a reference system. A reference system is defined by a number of points whose positions are defined both on the ellipsoid and on the land surface. The relation between the coordinates tells us about the rotation in three dimensions. These reference systems differ from one country to another. We have to be aware of the different ellipsoids, projections, co-ordinate systems and reference systems to be able to handle and solve problems related to the use of different map data in a GIS.

CO-ORDINATE SYSTEM



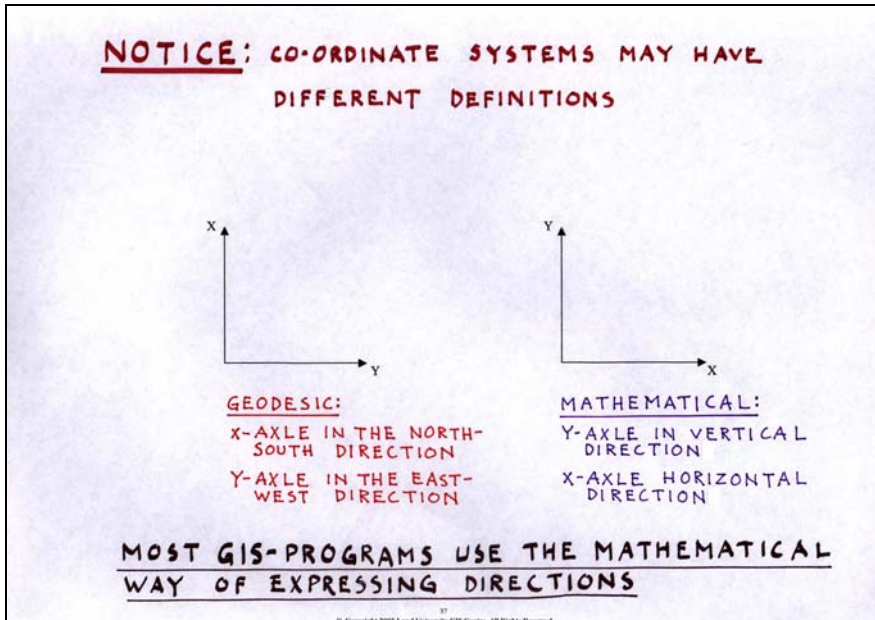
TO AVOID NEGATIVE CO-ORDINATES, ORIGIN IS MOVED (FALSE EASTING)

36

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

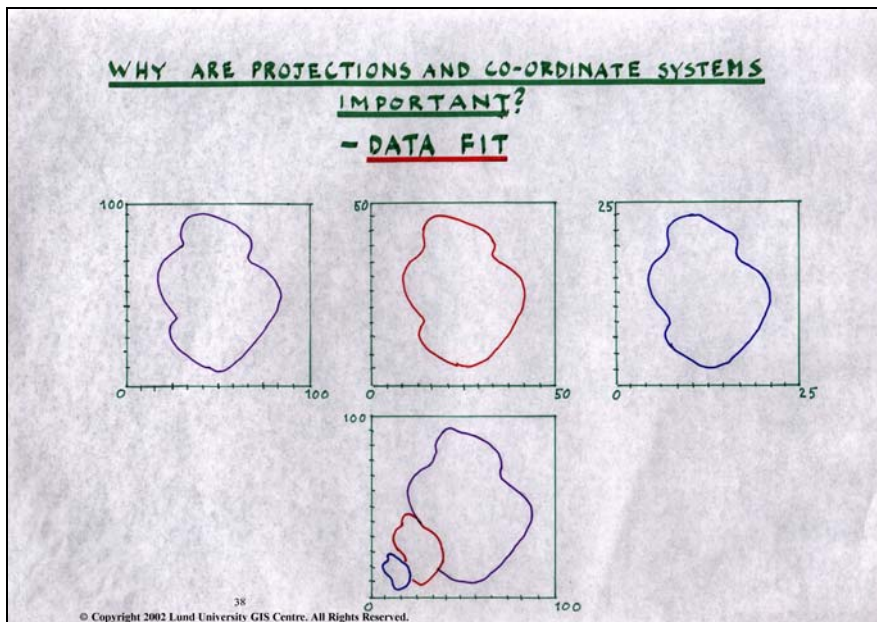
Slide 36

The coordinate systems are often defined so that they avoid negative co-ordinate values. The standard origins are the Greenwich and the Equator, but often the origin of a specific reference system is put outside a country or the area of interest. If the Greenwich origin, for example, is moved further east, we call this false easting and if the Equator is moved further north we call it false northing. By changing the origin we avoid the negative numbers in our co-ordinate system.



Slide 37

Something that is often confusing is that the x and y axes can be reversed, depending if they originate from a geodesic or mathematical point of view. Most GIS use the mathematical way of expressing the co-ordinates while some geographical data can be expressed in the opposite way (x axis as north-south direction and y axis as east-west direction).



Slide 38

If you take all the parameters described above, and different ways of transformations into account you realise that you can produce maps with quite different appearance out of the same surface area. Even though we might need help from surveyors in the geodetic field to solve some of the problems that might occur, we must at least be aware of that the differences exist and what problems they might cause. The geometric condition of

a map could be different using different projections, size might be different t using different co-ordinate systems, the shape of an object might be different if we use different ellipsoids etc.



LUND
UNIVERSITY



VECTOR STRUCTURE

STORAGE OF GEOGRAPHICAL DATA

GEOGRAPHICAL DATA INCLUDE INFORMATION ABOUT

- * COORDINATE SYSTEM (WHERE IS NORTH, SOUTH, WEST, EAST, ETC.)
- * ELLIPSOID
- * MAP DATUM
- * SCALE (FOR PRESENTATION)
- * USED SYMBOLS/LEGEND

- * GENERALIZATION
- * ACCURACY
- * WHO COLLECTED THE DATA
- * PRODUCTION YEAR
- * ETC.



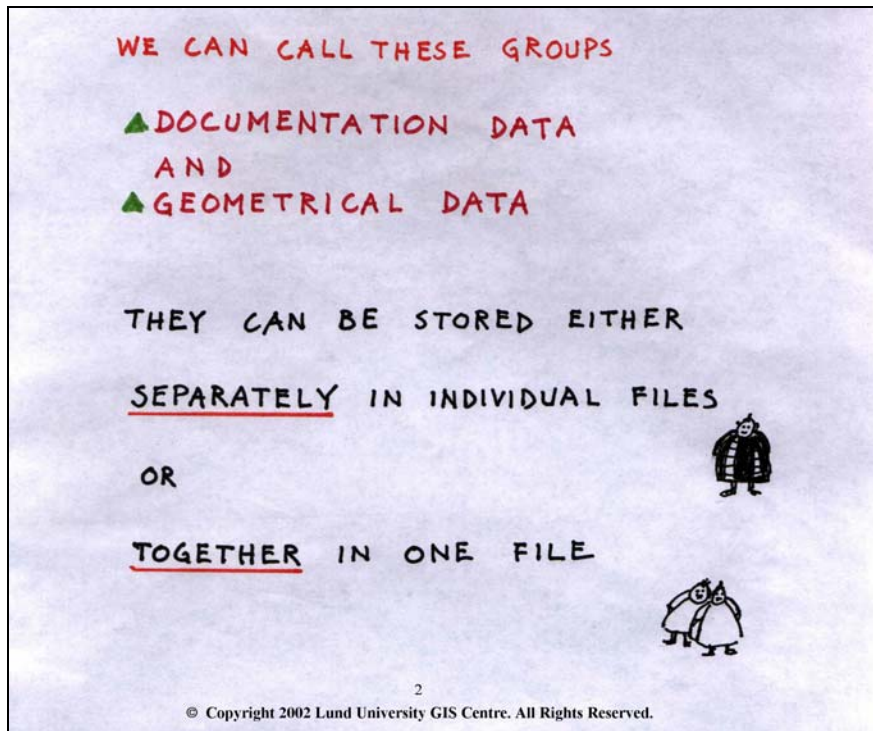
AND
* GEOMETRICAL OBJECTS (SUCH AS POINTS, LINES, POLYGONS)

1
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 1

This chapter is about different data structures in GIS and how to store information (mainly the geographical data). In vector GIS, geographical data are points, lines and polygons. In general, the information for these features can be divided in three parts:

- Information about the geographical objects – X, Y and Z co-ordinates
- Additional information – Accuracy, Scale, Producer, Production year, etc
- Attribute data




Slide 2


The geographical information can be further divided in two main groups:

- 1. Documentary data that describes the data*
- 2. Geometrical data that describes the location (co-ordinates).*

This information can be stored in two different ways; in the same file or in separate files. If stored separately, one file will contain the geometric data and one file the documentary data. Both files must normally be imported to the GIS to make a complete geographical database. Data can also be stored together in a single file. The documentation data is then called a “header”, telling something about the rest of the data in the file.

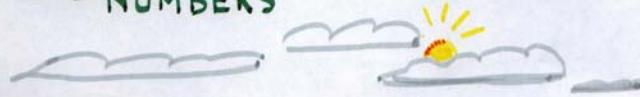
THE DOCUMENTATION DATA
MAY CONSIST OF

 CHARACTERS

 NUMBERS

WHEREAS THE GEOMETRICAL
DATA CONSIST ONLY OF

- NUMBERS



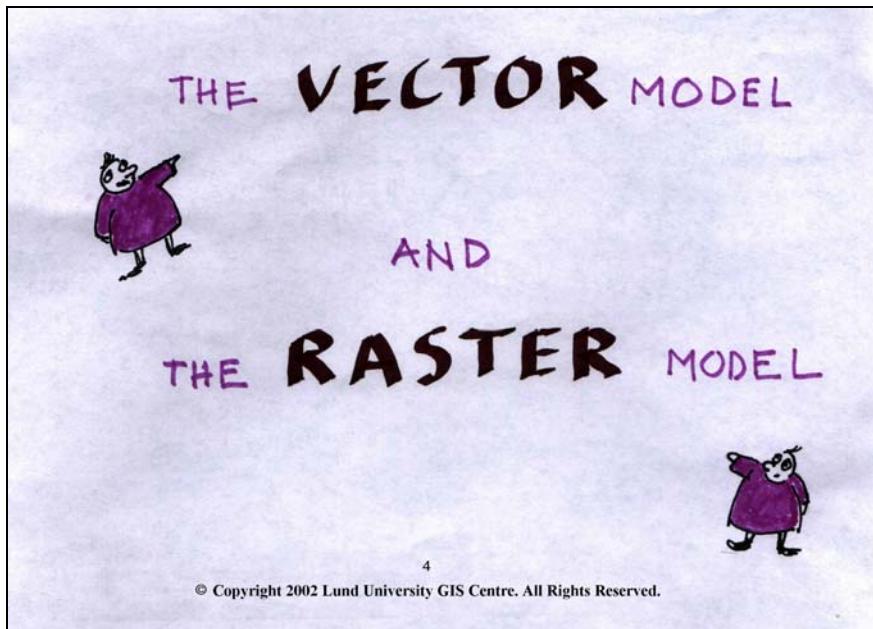
WE WILL NOW FOCUS ON
THE DIFFERENT WAYS
TO STORE GEOMETRICAL
DATA (I.E. SPACIAL OBJECTS)

3

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

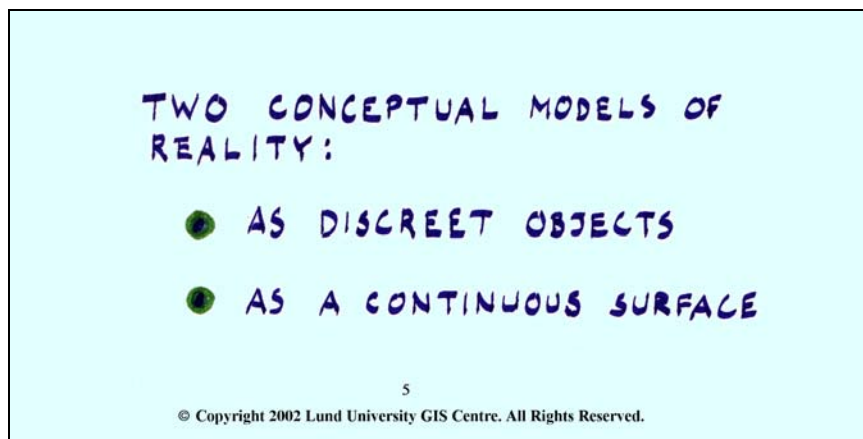
Slide 3

The documentation data can be alphanumerical (characters, i.e. text and numbers) but could also be a video clip, an image etc. The geometric data (the co-ordinates) is always numerical. The lecture will focus on how the geometric objects (points, lines and polygons) are stored in the computer.



Slide 4

There are two conceptually different data models available for storing GIS-databases; the vector model and the raster model. Most advanced GIS software can handle both models and even convert data between the two.

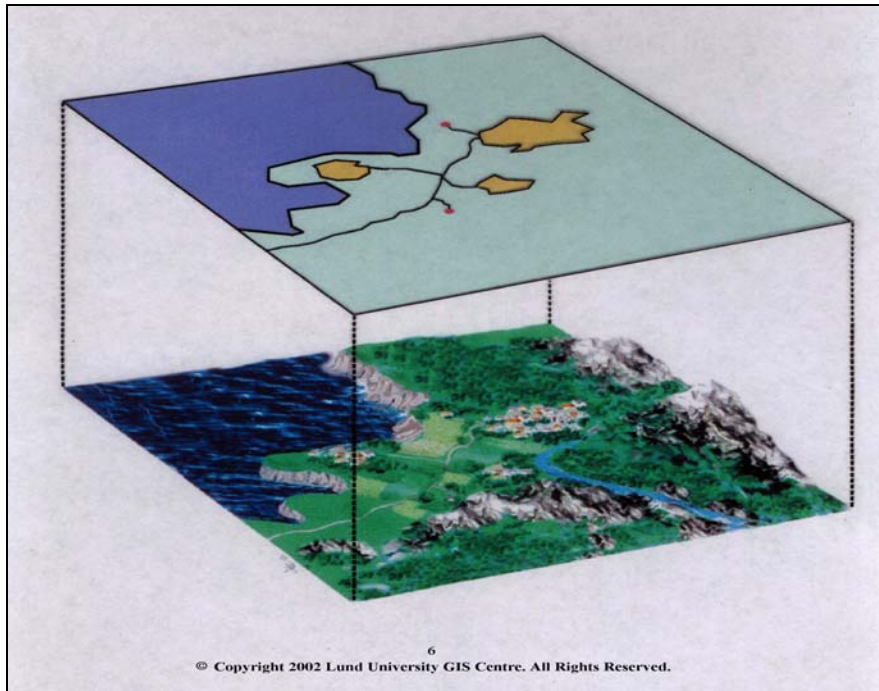


Slide 5

The real world can be described using two conceptually different models:

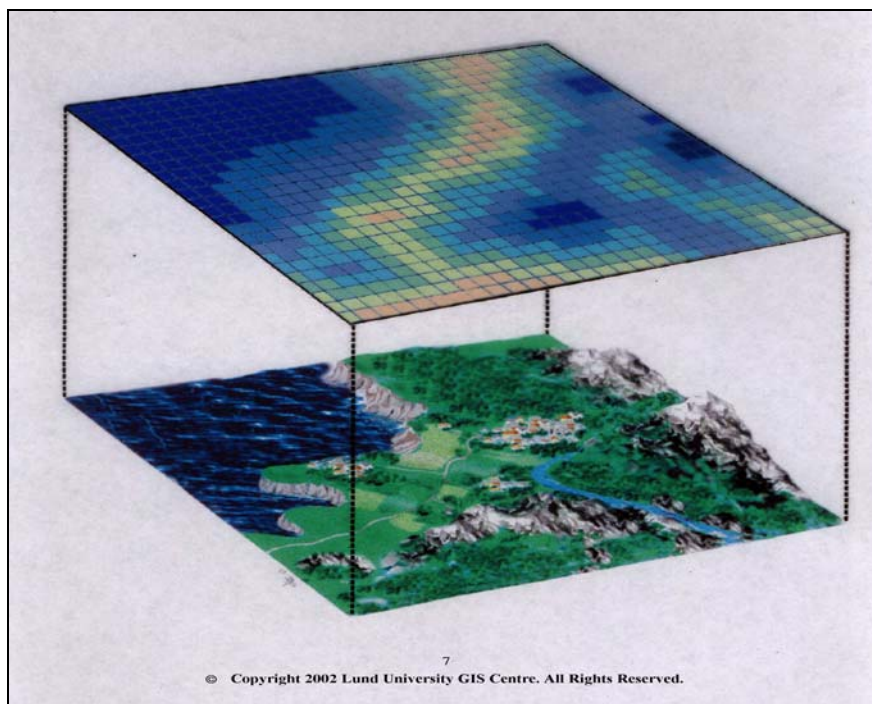
1. As discrete objects, possible to represent as points, lines or polygons.
2. As a continuous surface with no discrete or distinct borders, like temperature and precipitation.

To map houses and roads in an area, discrete objects are more suitable to use since these have a defined spatial extent, but to make a topographic map, a continuous surface should be used since topography has a continuous spatial variation.



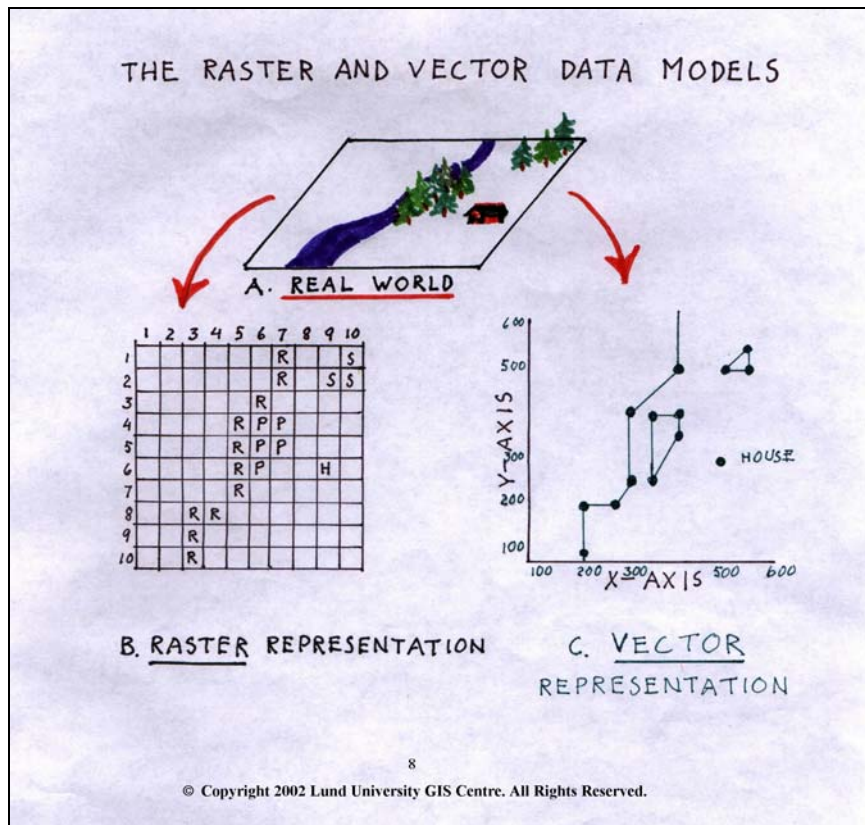
Slide 6

This is an example of a vector model representation of the real world, picturing a lake, some roads, villages and some smaller properties, all separated from each other by defined borderlines.



Slide 7

This is a raster model representation of the real world picturing gradual changes in topography (a continuous surface) with no distinct boarder lines.



Slide 8

The real world (A) can be represented as Raster model (B) or Vector model (C). When representing as raster, a raster/matrix/grid with a fixed cell size is placed over the area, and each raster cell is coded with a code representing the feature in that particular area. In this case, cells covering areas of the river are coded as “R”, cells covering pine forest are coded “P”, cells covering Spruce forest are coded “S” and the cell covering the house is coded “H”. The vector model (C) uses a totally different way for representing the different objects in the real world. The river is represented as a series of lines, with breakpoints every time the line changes direction. The x and y co-ordinate for each breakpoint is stored and the river is represented as a line in the vector model database. An attribute defining the line as river “R” is connected to each line segment. The forests are represented in a similar way by storing the x and y co-ordinate for each breakpoint of the forest’s border line. Defining this line as boarder line to an area permits construction of a surface or polygon and an attribute describing the type of forest is connected to it. The house is stored as a point, described with a single pair of co-ordinates.


VECTOR GIS

GEOMETRICAL OBJECTS
ARE REPRESENTED AS

- ◆ POINTS
- ◆ LINES
- ◆ POLYGONS

ATTRIBUTES CAN BE LINKED
TO THESE

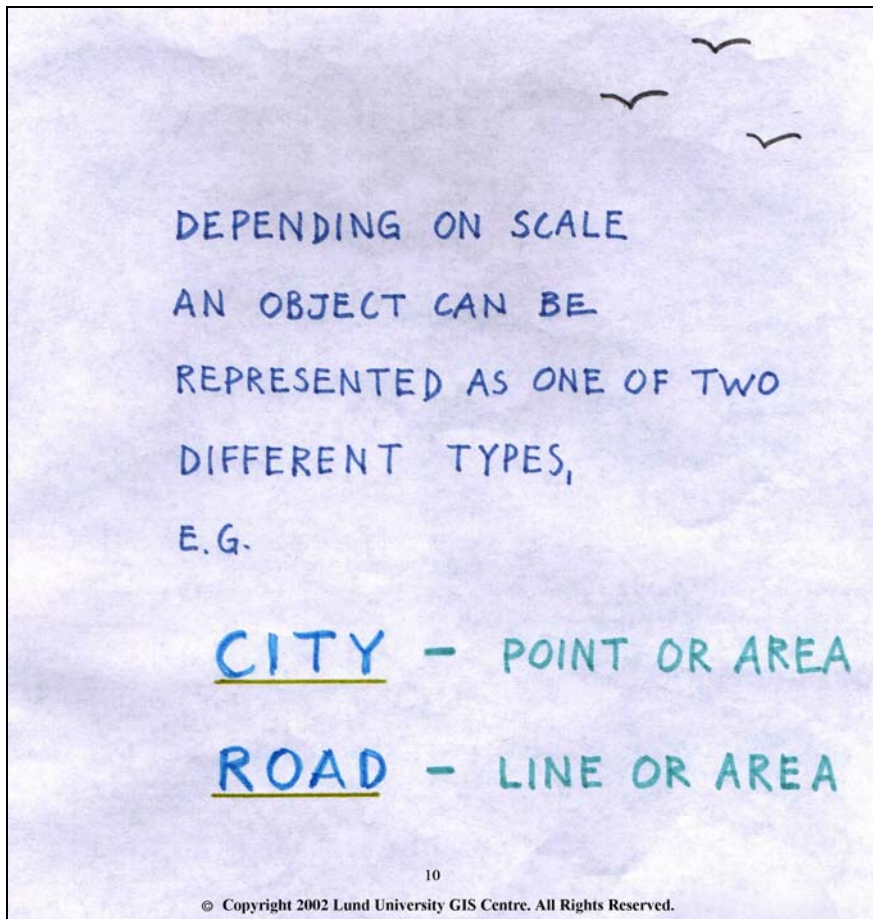
LINE { ARC
CHAIN
STRING
LINK
POLYLINE



9
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 9

Each object type can have an indefinite number of attributes connected to it, and the attributes are used for describing different aspects of the object. Note that in the literature, and in different software manuals, the objects may have different names, all meaning the same type of object, e.g. a line can also be called arcs and/or polylines.



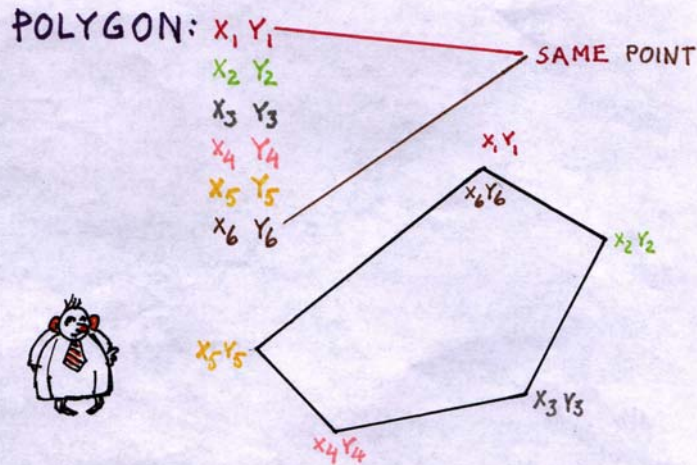
Slide 10

Depending on the map scale, an object in the real world may be represented by different geometric objects. A city on a world map will probably be represented as a point, while on a more detailed map, the city will be represented as a polygon. The same is valid for lines and points. The type of geometric object selected for representing different real world objects is a matter of generalisation and purpose of presentation, which is important to keep in mind when working with GIS-databases.

THE GEOMETRICAL DATA ARE STORED AS COORDINATES,
E.G.

POINT: X Y

LINE: x_1, y_1 - START NODE
 x_2, y_2 - VERTEX
 x_3, y_3 - STOP NODE

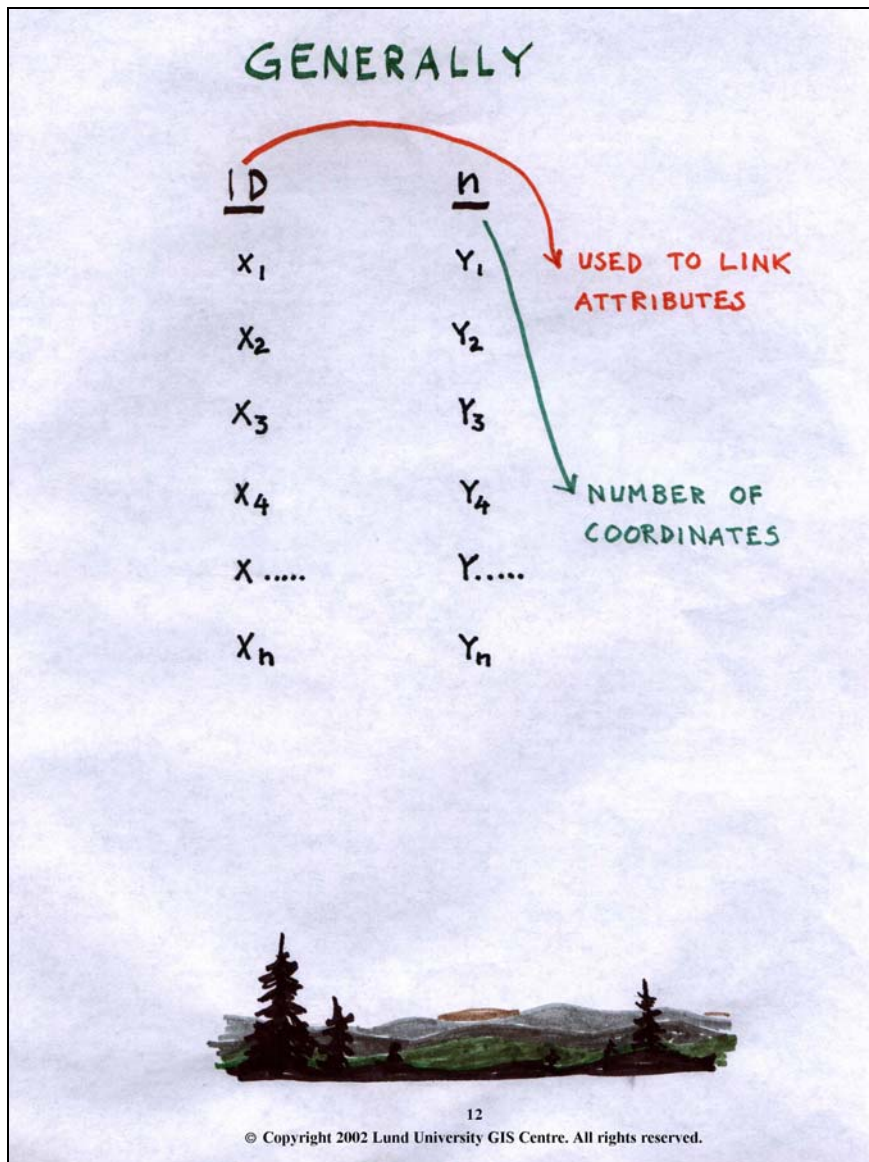


11

© Copyright 2002 Lund University GIS Centre. All rights reserved.

Slide 11

The different geometrical objects in the vector model are stored as one or several pairs of co-ordinates. The basic geometrical object, the point, is easiest to represent since the point has no spatial extent and can be represented by a single pair of co-ordinates. A linear object is more complicated since it has a one-dimensional extension and consequently must be represented with a series of co-ordinate pairs describing this extension. In the simplest case of a straight line between two points, the vector model stores the start point and the end point of the line. More complex lines need several pairs of co-ordinates, one at each breakpoint, where the line changes direction. Smooth real world objects, like for example rivers that change direction more or less constantly, will have to have an indefinitely large number of breakpoints (pairs of co-ordinates) in order to be represented correctly. In GIS terminology, the start point and end point of a line are called “node” and the breakpoints in between are called “vertex” (vertices in plural). Storing of polygon features is similar to line storing. The polygon is defined by its border line, which in turn is described by start node, stop node and a number of vertices. The only difference is that the co-ordinates for the start and stop nodes must be the same, otherwise the polygon will not be closed. The GIS software can thus distinguish between the three geometric objects simply by looking on the list of co-ordinate pairs used to describe it. If it is a point, there is only one pair of co-ordinates, if the object is composed of several pairs of co-ordinates it must be a line and if the co-ordinates for the start node and stop node are the same, the object is a polygon.



Slide 12

The general format for storage of vector data in a GIS database contains three components:

1. The object Id-number, which has to be unique (used to identify the object and to link attributes to it).
2. The n-term that states how many pairs of co-ordinates are used to build the object
3. The co-ordinate pairs as defined by the n-term.

DIFFERENT VECTOR MODELS

- SPAGHETTI MODELS

- TOPOLOGICAL MODELS

TOPOLOGY IS INFORMATION
ABOUT RELATIONS BETWEEN
OBJECTS



13


© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 13


In vector GIS, there are several different ways of storing geometrical information ranging from the simpler, so called **Spaghetti** model, to the more complex Topological model. Both these models are common in GIS and the difference between them is in the level of structure and organisation of the data. In the simple spaghetti model, objects are stored more or less as described in the general vector data format (described on the last slide), while in the topological model, additional information about an objects relations to neighbouring objects is stored together with the co-ordinates. Both models can be used for all types of operations and analysis, but for complex operations the structured topological model will perform more efficiently than spaghetti models. **Topology** is a very common GIS term. It simply means that when topology is present, each object in the database has “knowledge” about its relation to adjacent objects.

TOPOLOGY


- CONTIGUITY



- CONNECTIVITY



TOPOLOGY IS NOT
 NECESSARILY SENSITIVE
 TO CHANGES IN LENGTH,
 SIZE, SHAPE, ETC.

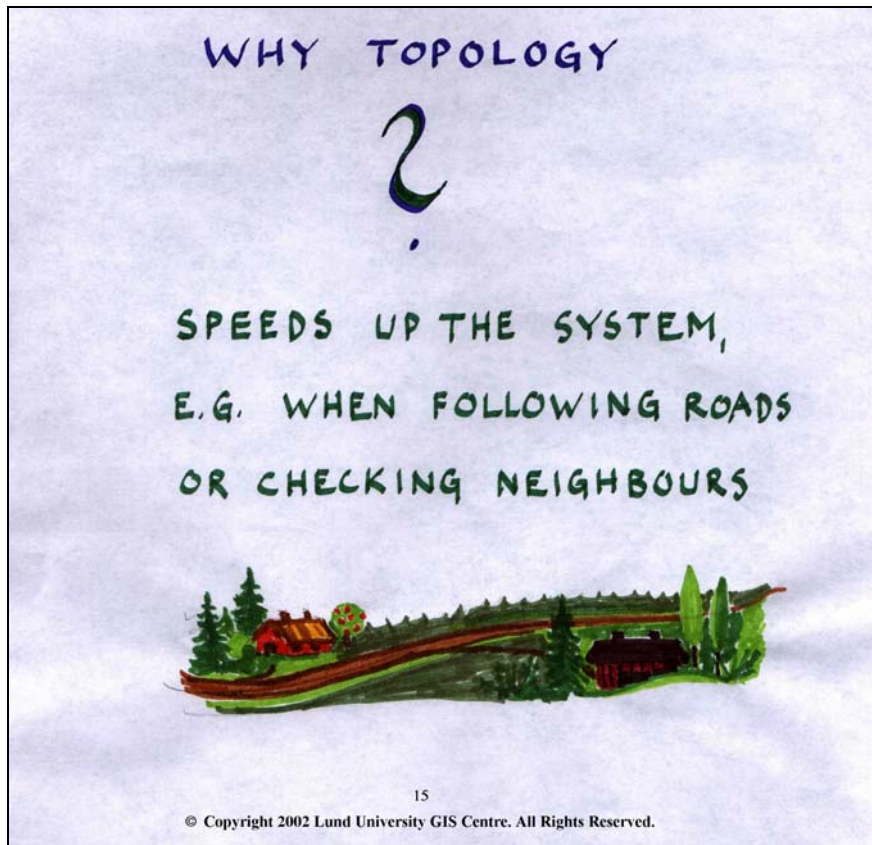


14

© Copyright 2002 Lund University GIS Centre. All rights reserved.

Slide 14

In complete topological models, polygons have **contiguity** (information about their neighbours) and line **connectivity** (information about which nodes are common to other lines). This information decreases processing time, e.g. in network analysis when searching the shortest route between two cities, since the computer doesn't have to check which lines are connected since this is already stored in a table in the database. It is important to note that topology is not necessarily sensitive to changes in length, size and shape of the objects. Changing the shape of a border between to polygon does not change the fact that they are still neighbours and changing the length of a line does not necessarily change the connections to other lines.



Slide 15

The main reasons for advocating the use of complete topology are that processing times are decreased and data storing is more efficient, conserving space on the hard disk.

**THREE EXAMPLES OF VECTOR
DATA STRUCTURES**

- * **SIMPLE POLYGON
STRUCTURE**
- * **SIMPLE POLYGONS
WITH LIST OF
COORDINATES**

SPAGHETTI
MODELS

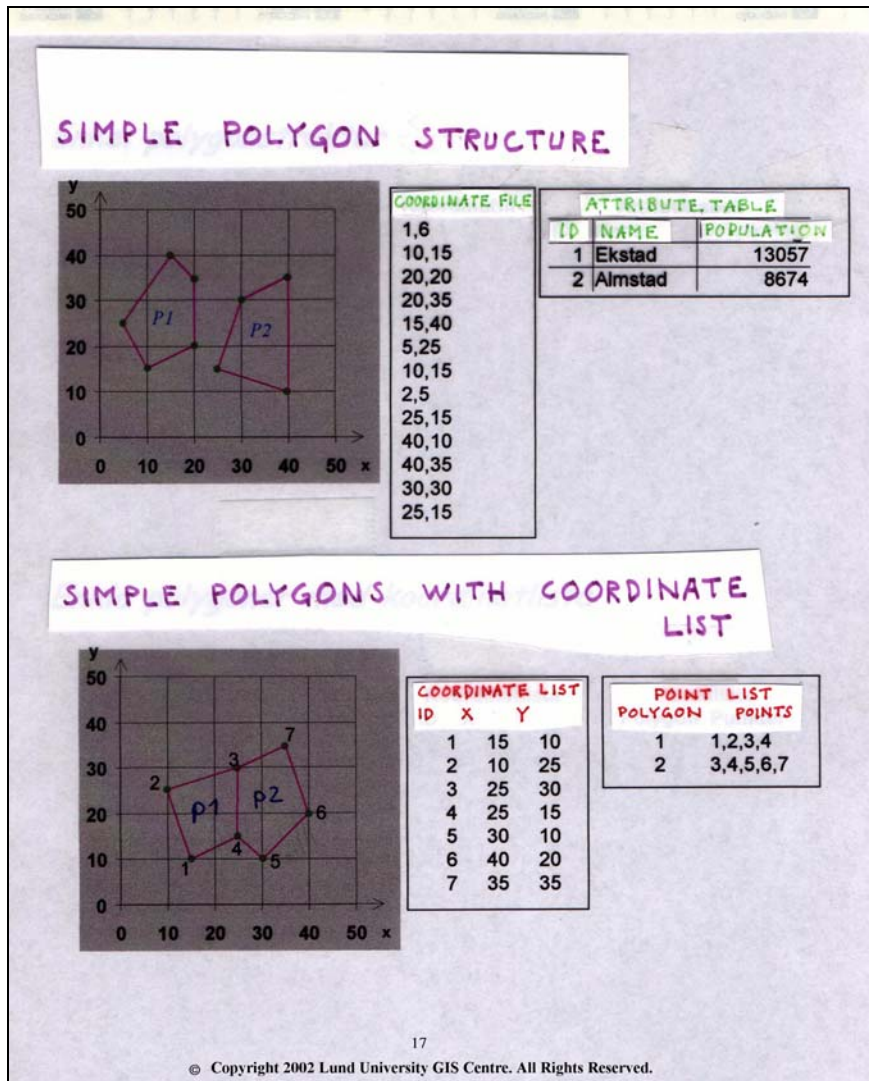
- * **COMPLETELY
TOPOLOGICAL DATA
STRUCTURE**

16
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 16

The following slides will describe three examples of vector models, all of them in frequent use by different GIS software:

- | | |
|--|-------------------|
| 1. Simple polygon structure | Spaghetti model |
| 2. Simple polygons with co-ordinate list | Spaghetti model |
| 3. Completely topological data structure | Topological model |



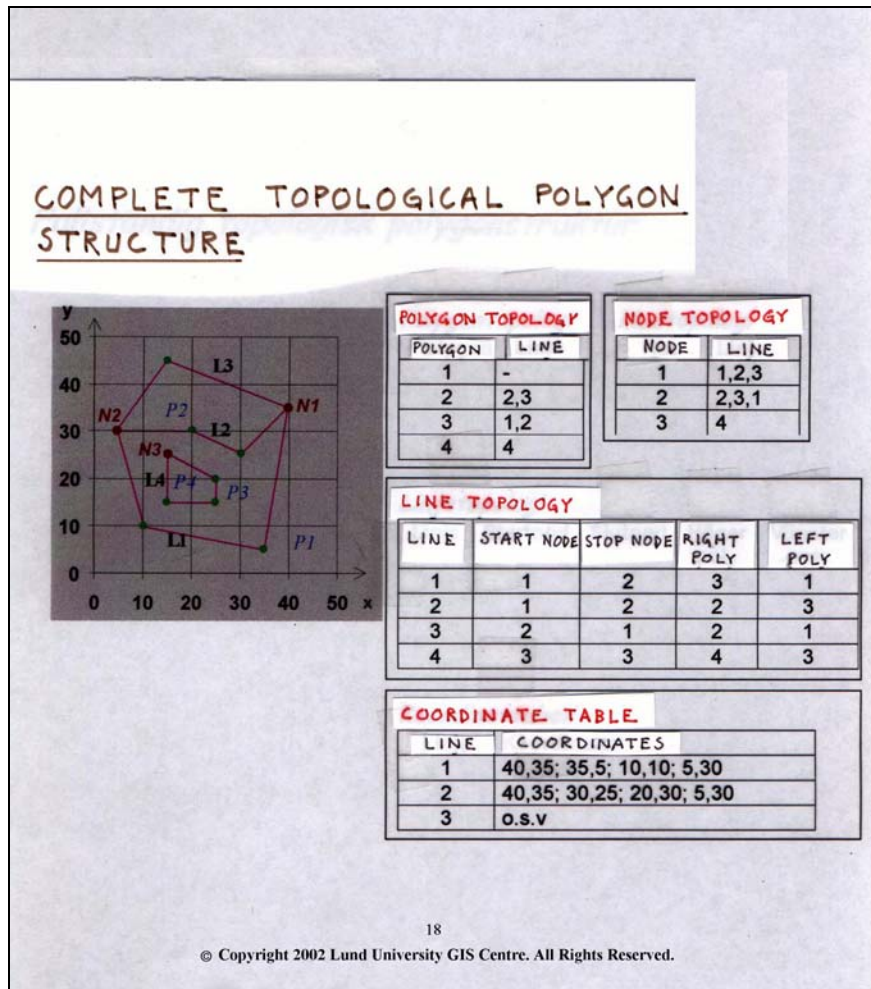
Slide 17

The **simple polygon structure** data storing format is more or less an exact copy of the general vector data storing model, i.e. Id-number, number of pairs of co-ordinates (n) and finally a list of the x and y co-ordinates. In the example, polygon Id = 1 has 6 pairs of co-ordinates ($x=10, y=15; x=20, y=20; x=20, y=35...$). The Id-number is used to link attributes to the geometrical objects, yielding an operational GIS database structure. Note that the number of breakpoints on the polygon boarder is $n-1$ (one less than the number of pairs of co-ordinates) due to the reason that both the start and stop node co-ordinates has to be stored. There are some disadvantages with this type of polygon structure:

1. No information about neighbouring polygons is stored.
2. All pairs of co-ordinates that are common to neighbouring polygons have to be stored in the co-ordinate list of all polygons.
3. When digitising maps, this means that common boarders have to be digitised once for every polygon, increasing the risk of miss-match along the common boarders.

In the **simple polygon structure with coordinate list structure**, each vertex has a unique Id-number a co-ordinate list, meaning that a polygon can then be defined by listing the vertex Id-numbers for the vertices on the boarder line of that particular polygon. In the example, polygon 1 is described by vertex 1,2,3,4 and polygon 2 by vertex 3,4,5,6,7. Storing the polygon structure in this way gives limited topological information since, by comparing the list of vertices, it is possible to identify polygons with vertices in common, which means that they must have a common boarder line, thus being neighbours. This way of storing the data

structure is a bit more memory efficient than the simple polygon structure. As in the simple polygon structure, the polygon Id-number is used to link attributes to the geometry.



Slide 18

*The complete topological polygon structure describes the interrelationship between all geometrical objects in the database. At first glance, this structure might look much more complicated and inefficient, but in fact the opposite is true. The complete topological structure is composed of four tables; the **polygon topology table**, the **node topology table**, the **line topology table** and the **co-ordinate table**. The co-ordinate table lists all the co-ordinate pairs for the lines in the database. It begins with the co-ordinate pair for the start node followed by a number of vertex co-ordinate pairs and ends with the stop node co-ordinates. In the example, line number 1 starts with a node at co-ordinates 40,35 followed by vertices at 35,5 and at 10,10 and ends at the coordinates 5,30. The line topology table lists the start and stop nodes for each line, which describes line connectivity and direction of each line. The table also contains information about the polygons at the right-hand and left-hand side of the line, based on the direction of the line (from start node to stop node). This is important since it gives information about what polygons that are neighbours. The node topology table also contains useful connectivity information since lines connect at nodes. In the example, lines 1, 2 and 3 are connected in node number 1. The polygon topology table store information about lines building the polygons. Instead of comparing all pairs of co-ordinates when searching for neighbouring polygons, the computer only has to search the polygon topology table to identify them. Note: Polygon number 1 is called the background polygon surrounding all other polygons and is used internally by the GIS software, with no application for a user. It is very important to be able to distinguish the different data structures from each other and to know the difference between spaghetti structure and topological structure since it affects how the software handles your data. In the example, a complete polygon topology has been described. If a database doesn't contain polygons, the polygon topology table is omitted from the structure.*

NORMALLY, DIFFERENT TYPES
OF SPATIAL OBJECTS

(ROADS, LAKES, BORDERS, ETC.)

ARE STORED IN DIFFERENT



LAYERS

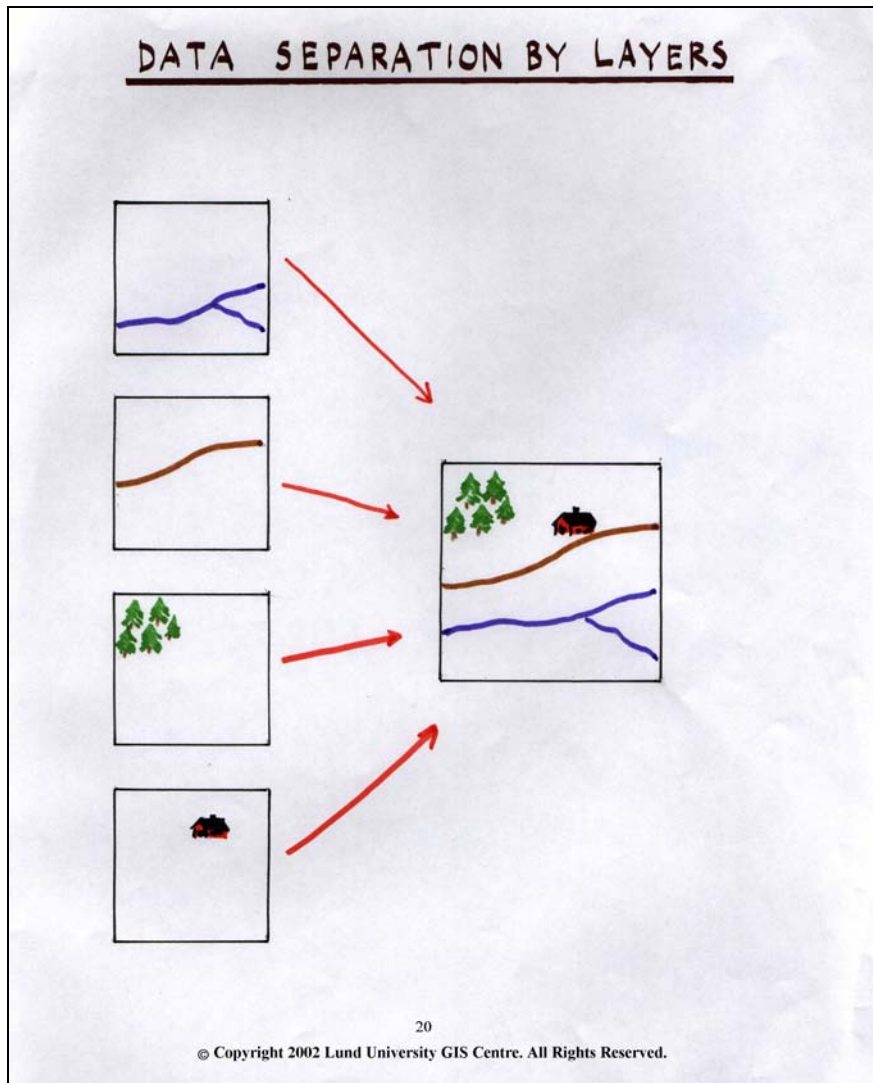
A LAYER CONSISTS OF
ONE OR MORE DATA FILES

19

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 19

Different types of geometrical objects (points, lines and polygons) are normally stored in different map layers (data files) in order to handle them more efficiently. Different thematic information of the same character, e.g. roads, railroads and rivers (assuming they are all represented as lines) are frequently also stored in different databases, but may also be stored together in the same database and separated by attaching attributes describing their character.




Slide 20

Different thematic map layers, in this case consisting of a river layer, a road layer, a forest layer and a house layer are combined with the GIS software to present the desired information on a map. Provided that the cartographic settings are the same for all layers, it is simple and straightforward to combine any type of data on the same map.

THE VECTOR GIS IS

- - EXACT (GOOD MEASUREMENTS, E.G. OF LENGTH AND AREA)
- - SOMETIMES FAST (E.G. WHEN TRACING LINES)
- - GOOD FOR VISUALIZATION
- - COMPACT IN TERMS OF DATA STRUCTURE
- - OFTEN TOPOLOGICAL (SPEEDS UP DATA RETRIEVAL)



21

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

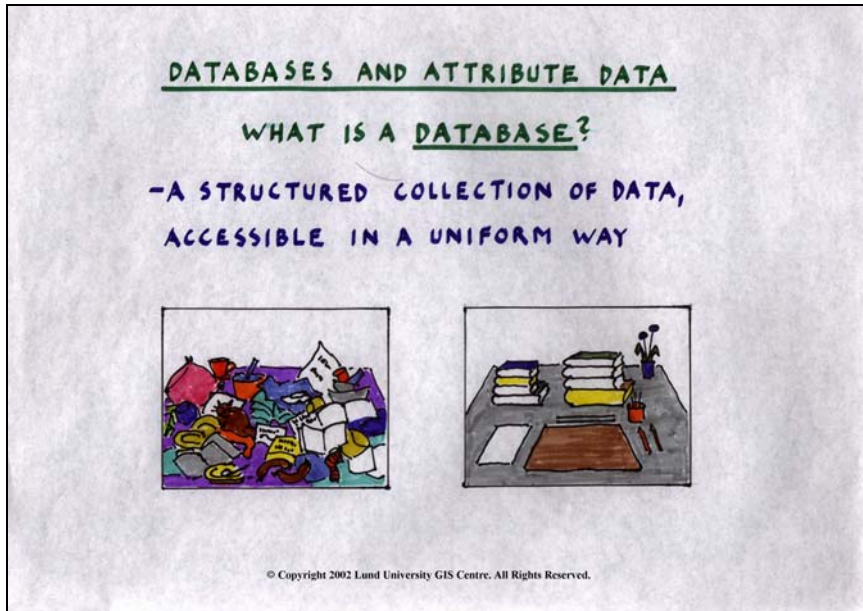
Slide 21

To summarise, vector GIS is:

- Exact – The geometry of the vector model inherits the accuracy of the original data, as collected by field surveyors, GPS, photogrammetry, etc, since the structure of the model is based on storing the actual co-ordinates describing the location of different objects. This means that measurements done in vector database are as exact as the original data.
- Sometimes fast – Many operations are easy to perform on vector model data, e.g. network analysis (tracing lines and measuring distances along networks).
- Good for visualisation – For the same reason as it is exact, the vector model yields neat looking maps for all types of objects that are suitably represented by the model.
- Compact in terms of data structure – The only information that has to be stored are the nodes and vertices for lines and borderlines. Vector data structures demand much less computer storage space than raster data structure.
- Often topological – Most GIS software handles complete topological data structures, which speeds up the data retrieval and gives information about contiguity and connectivity.

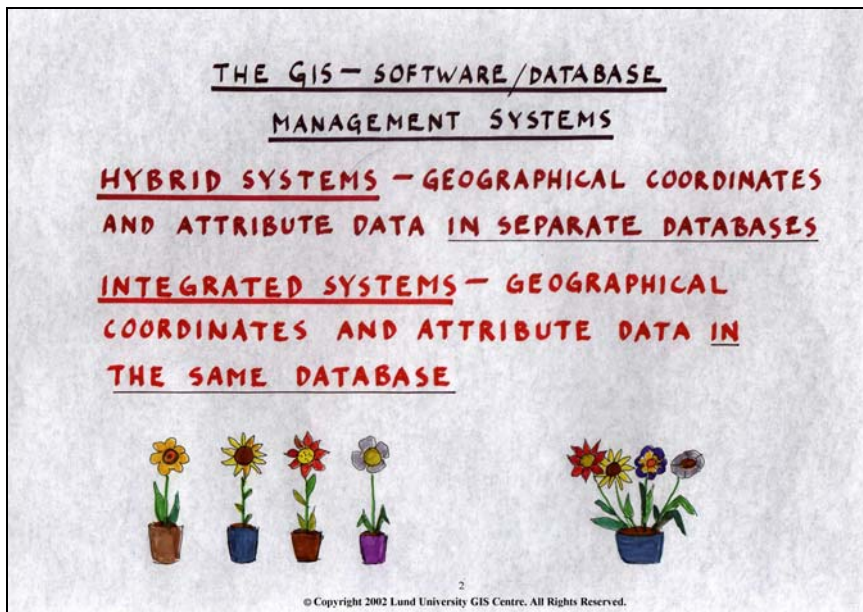


DATABASES



Slide 1

The following lecture will describe database construction and database management systems (DBMS). While reading this lecture, it's important that you keep in mind that the lecture mostly concerns databases for the attribute data and not the geometrical data. A database can be defined as a structured collection of data that is accessible in a uniform way. A database is a really important part in a GIS since the data is organised using different types of DBMS.

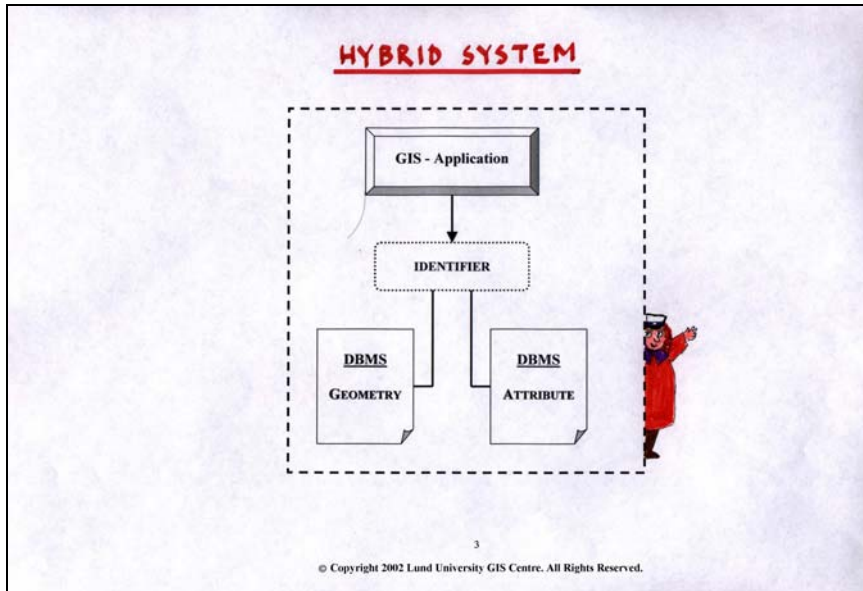


Slide 2

There are two different types of database management systems:

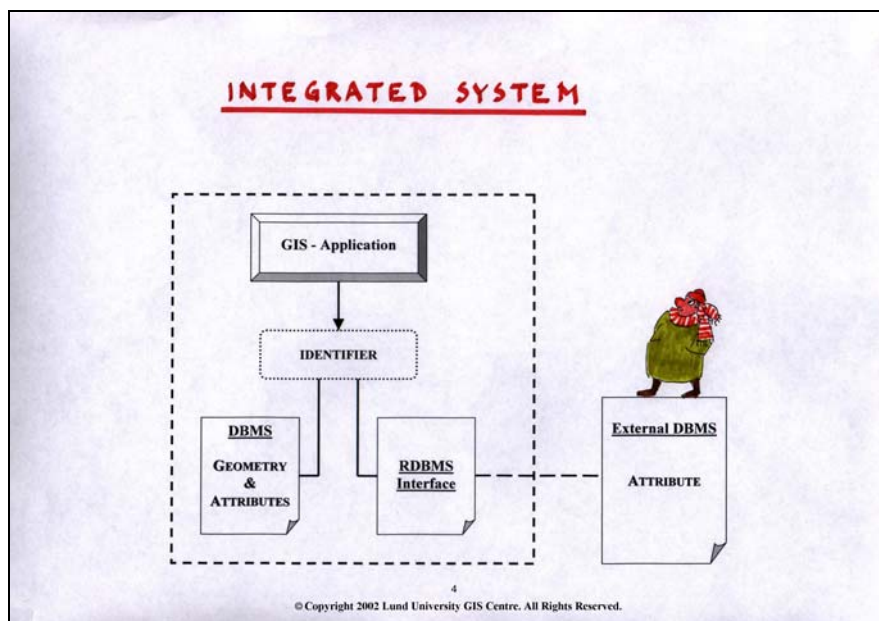
1. Hybrid System
2. Integrated System

In the Hybrid System the geographical coordinates (the geometrical data) and the attribute data are stored in separate databases, while in the Integrated System the geometrical data and the attribute data are stored in the same database.



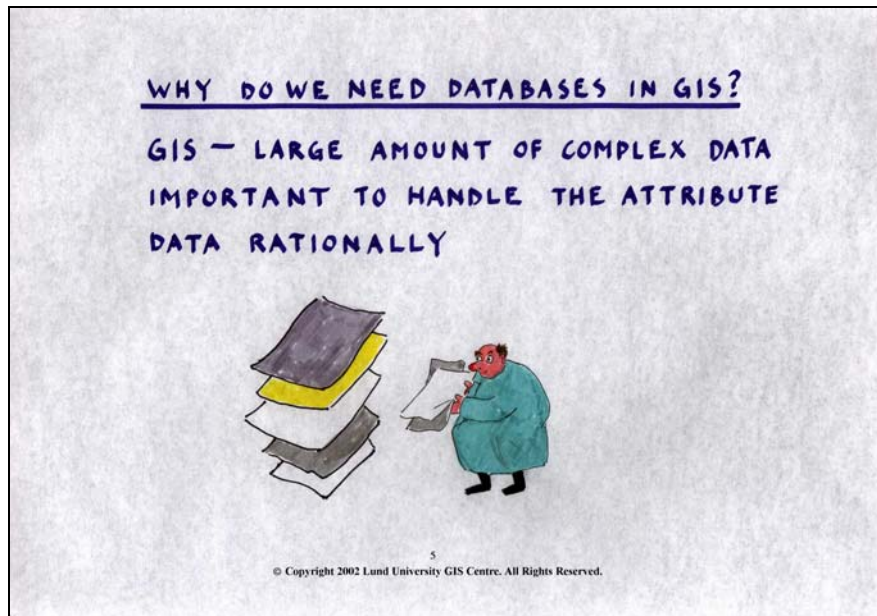
Slide 3

The Hybrid DBMS contains attribute data and geometrical stored in separate databases. Entries in the two databases are linked with a common unique ID-number. A part from the internal linking, it is always possible to link the ID-number with external databases, even via networks such as Internet or a local Intranet.



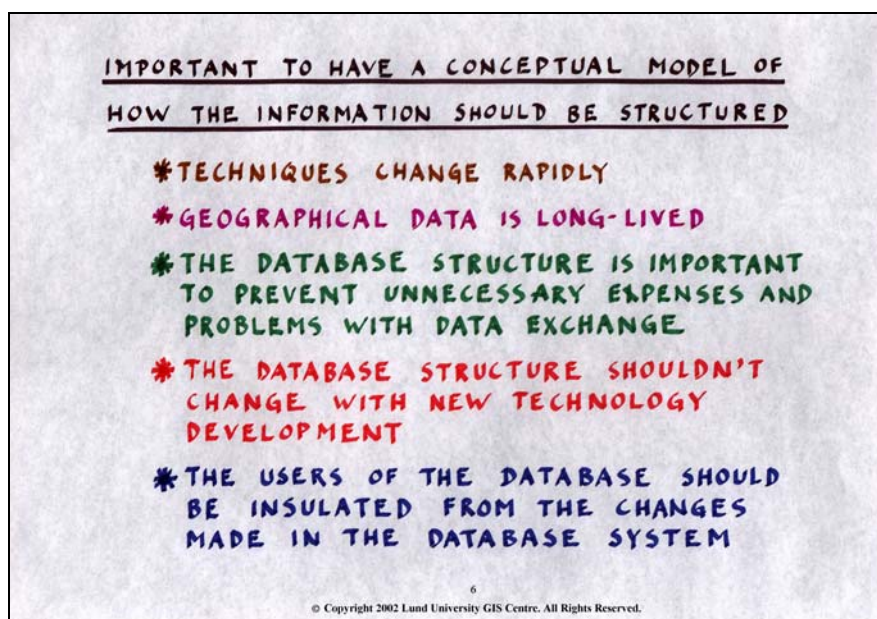
Slide 4

In the Integrated DBMS, geometrical data and attribute data are stored together in the same database. Using the unique ID-number, it is possible to link to external data sources via a RDBMS interface. The most commonly used DBMS in GIS-applications is the hybrid system.



Slide 5

GIS databases often contain large amounts of data, both co-ordinates and attributes, and the need for structuring well the data is obvious. If data is not handled in a rational manner, system responding times will be unreasonably long. The DBMS also assists in selecting specific types of data or specific parts of a database.

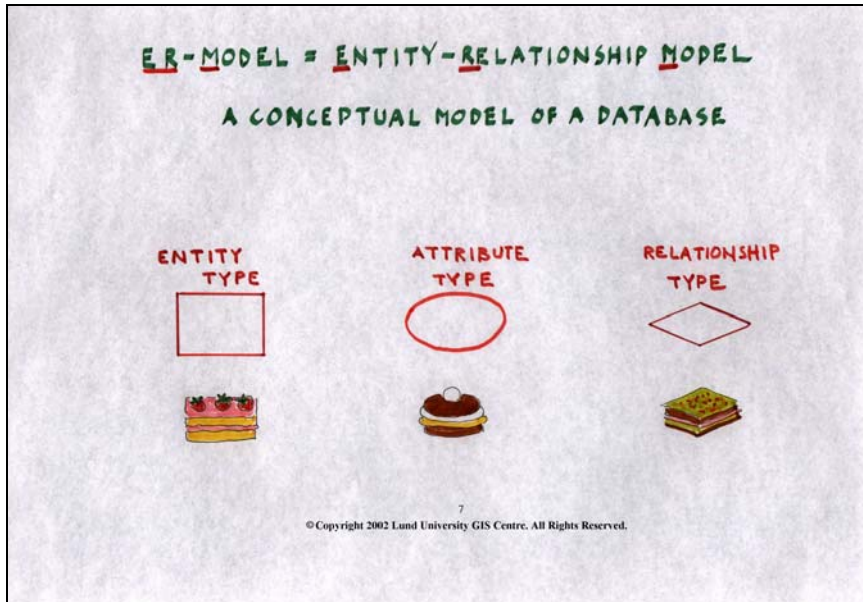


Slide 6

Before constructing a database, it is important to have some kind of conceptual model of how the data should be organised and how the information should be structured. The task of constructing this conceptual model is often referred to as Database Modelling. The database modelling should consider all stakeholders and cover both technical and non-technical aspects:

- The technique must be stable (in terms of hardware and software) so that the structure of the database does not have to change according to every rapid change in the development of these. A geographical database is often very long-lived and should therefore be planned to sustain for a longer period of time.
- The database structure should be kept as simple as possible to facilitate changes, input and output of data. It is unnecessary to try to make fancy solutions to solve simple problems!

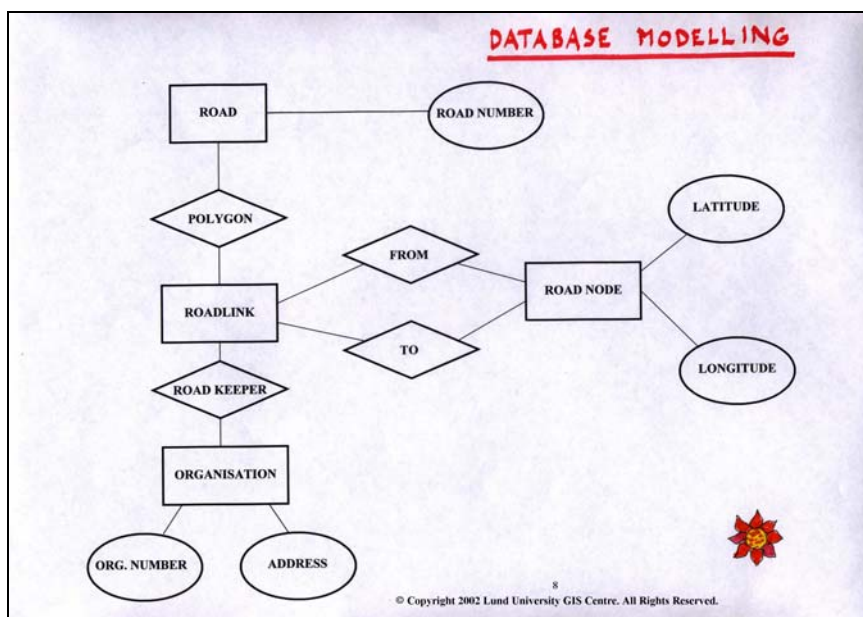
- The users should be insulated from changes made in the database system so that changes in technologies will not affect the handling of the data (editing, in- and output etc.) or the connection between database and user interface.



Slide 7

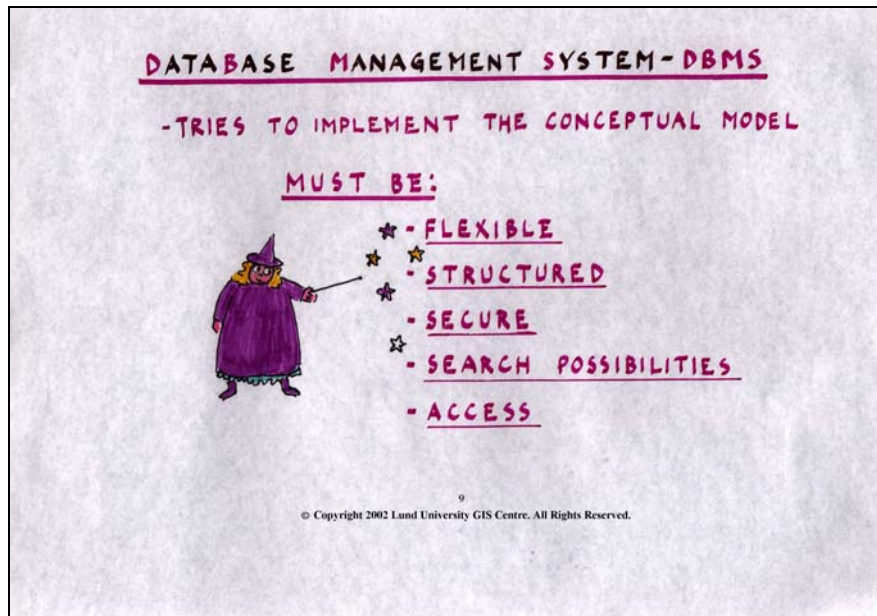
Database modelling can be based on different concepts. One frequently used is the Entity-Relationship model (ER-model). This model consists of three different objects that are symbolising the different parts in the database.

1. Entity type – defines what kind of object we are dealing with.
2. Attribute type – describes different entity types.
3. Relationship type – defines how the entities and the attributes are related in the database.



Slide 8

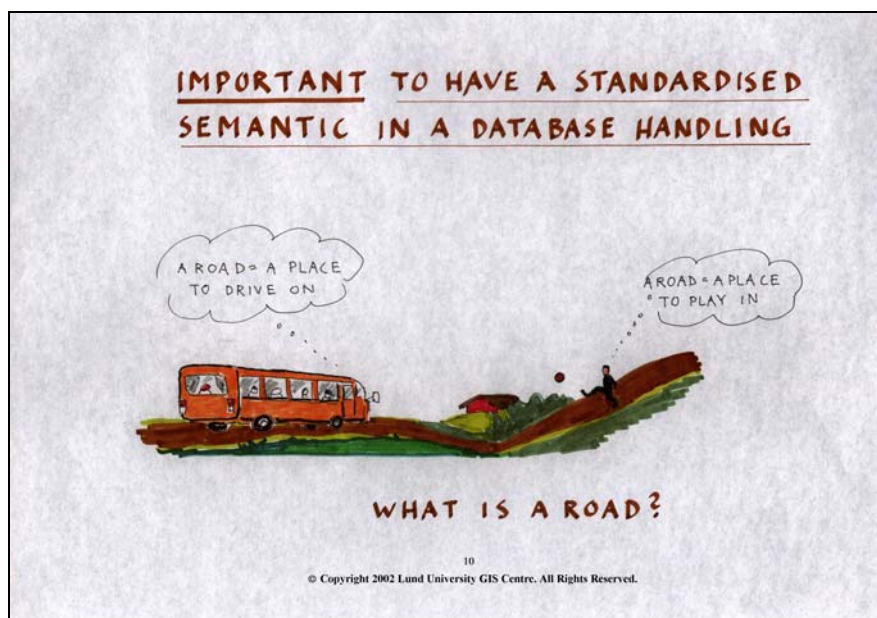
Even though ER-modelling might seem like a quite basic, and perhaps simple, part of the database modelling, it is a very important field in the database construction process. Slide 8 is a small example of an ER-model where you have a few entities linked together with some specific attributes, in this case, a database for management of a road network. Depending on the complexity of the database, the ER-model will be more or less complicated. With this type of database modelling, strengths and weaknesses with different database constructions can be illustrated graphically and analysed. In this example the entity *Road* is linked to the attribute *Road number* and to the entity *Roadlink*.



Slide 9

After finishing the conceptual modelling, the database structure shall be implemented into a DBMS (software). While doing so, there are some things to take into account:

- The construction must be flexible enough to take care of different actions within the DBMS.
- The data must be structured to facilitate data retrieval.
- The construction of the DBMS must minimize the risk for errors within the system. It must be impossible to enter the wrong type of data into the wrong part of the system, e.g. text in a numerical field.
- The database must be accessible and have some sort of search possibility through the software system. This might include the construction of interfaces to users that have no skills in database management and therefore otherwise might have problems retrieving data.



Slide 10

It is important to have standardised semantic to ensure that the users can communicate and make sure that everybody is talking about the same thing. The meaning of a road for a bus driver might be a place where he can drive his bus, while a child might think of it as a place to play football and other persons might have other ways of thinking of a road. It is very important to have a standardised terminology for every entity in your database in order to enable communication between users.

THERE ARE DIFFERENT SORTS OF DATABASES:



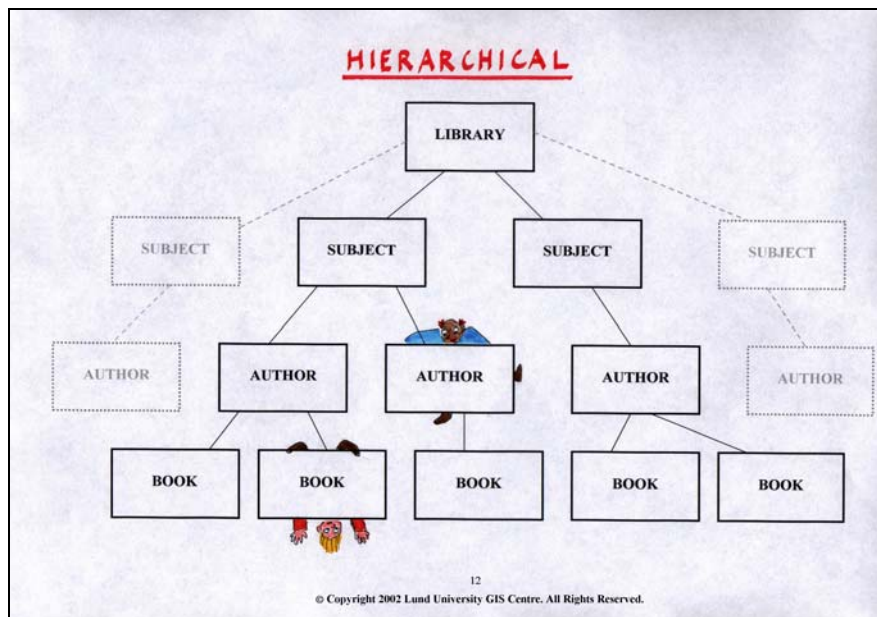
11

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 11

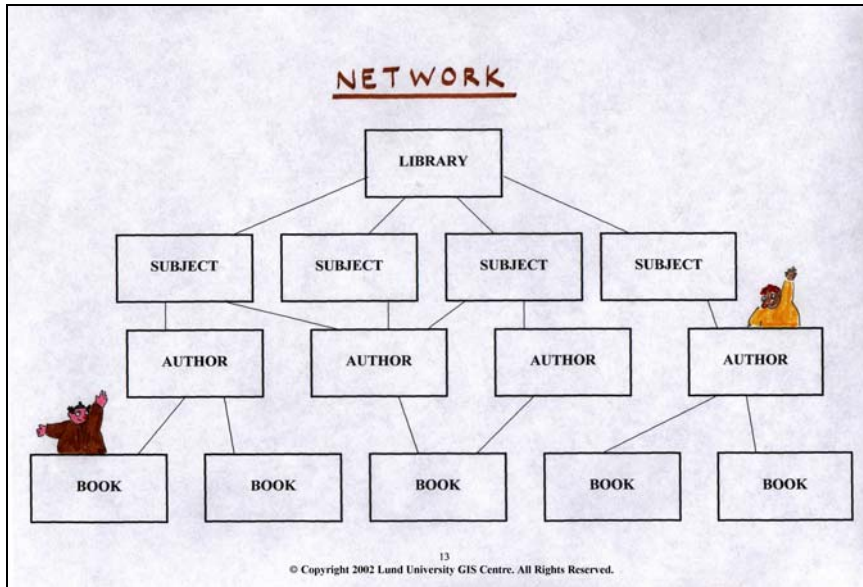
There are three different types of commonly used databases structures:

1. Hierarchical databases
2. Network databases
3. Relational databases



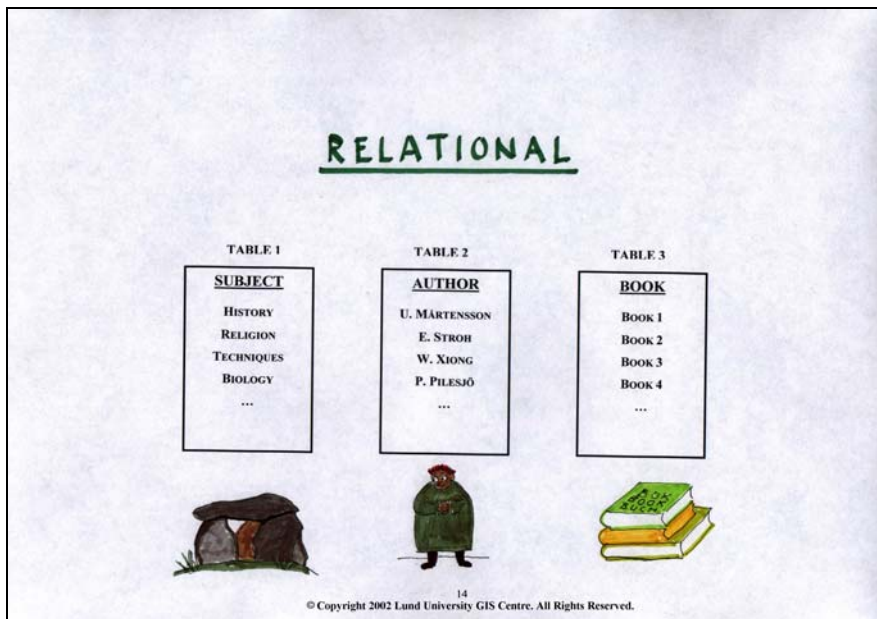
Slide 12

The Hierarchical database structure creates a tree with quite simple links between the different levels. The structure makes it possible to link just one object from the bottom to the top. The slide shows an example of a hierarchical database for a library with different subjects, authors and books. It is only possible to link one book to one author and one author to one subject, which makes the database inefficient since a book can be written by many authors or an author might have written books in many different subjects. This sort of database is not very common anymore.



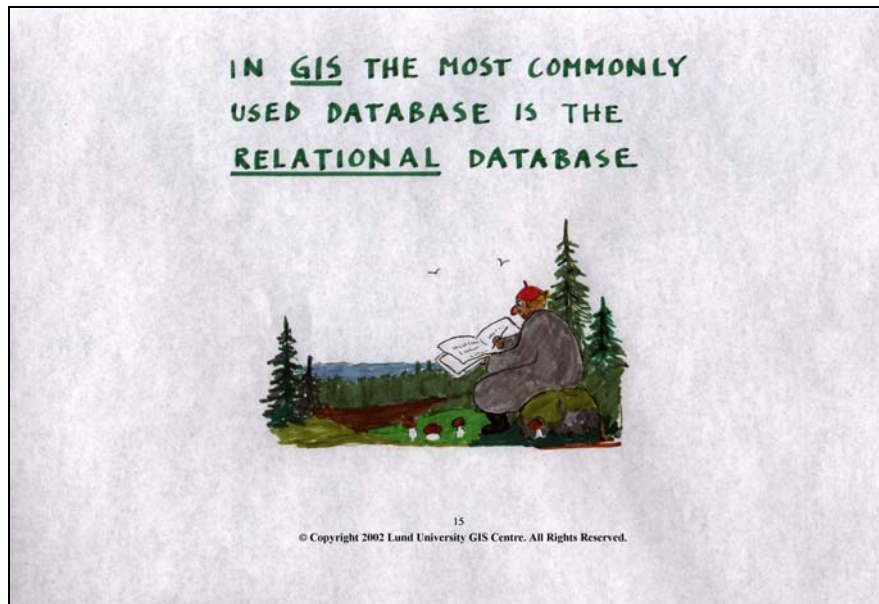
Slide 13

The Network database structure is more complex than the Hierarchical database structure. The different entities can be linked to one another in a more flexible way. A book in the library database can now be linked to several authors and the authors can be linked to several subjects and so on.



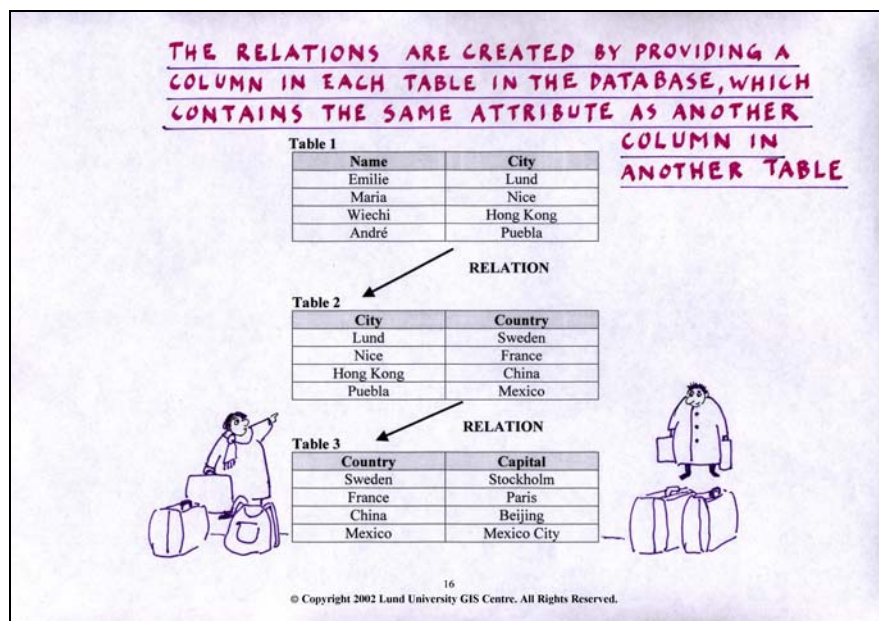
Slide 14

In the Relational database structure, different entities are stored in tables in which the entities are described and linked (related) to others in other tables. The slide shows the library database in which the entity “subject” is stored as a table with different subjects listed and the subjects are linked to the entity table “author” in which the different authors are listed and so on. This might seem a little bit complex but it will be explained more thoroughly further on. The relational database structure is much more flexible than the two previous ones.



Slide 15

In GIS software, the most commonly used structure is the relational database structure.



Slide 16

The Relational database structure is organised in tables, which in turn are organised in columns containing different information. The fundamentals of the relational database is that a column in a table contain the same information as another column in another table – this makes it possible to link the two tables together and connect the information stored in them. This is called a *relation*, hence the name of the database structure. The slide shows three tables with two columns. The first table stores personal names and the names of hometowns. The second table also consist of a column with the names of the hometowns but here, there is also a column with the name of the country in which the different cities are located. This “duplication” of the columns with the name of the hometowns makes it possible to link the different tables together. The third table consist of a column with the name of different countries, but also of a column with the names of the capitals in each country. This makes it possible to link table 2 with table 3 since they both have a column with the names of different countries and it also makes it possible to link table 1 and 3 since they have a relation to each other through table 2. This use of “duplicated information” to create relations between the different entities in the relational database is the same thing as the use of ID-numbers to connect geometrical data and attribute data.

RDBMS

RELATIONAL DATABASE MANAGEMENT SYSTEMS

DATA IS ORGANISED IN DEFINED COLUMNS

OBJECT
POST
RECORD
TUPEL

→

Number	Surz-name	Other name	Acode	Tel
1	Mårtensson	Ulrik	046	222 4026
2	Pilesjö	Petter	046	222 9654
3	Stroh	Emilie	046	222 3946
4	Xiong	Wiechi	046	222 3939

↑

FIELD
ITEM
TABLE
COLUMN

17


© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

1.1. Slide 17

A relational database is handled by a Relational Database Management System, abbreviated RDBMS in analogy with the DBMS mentioned before. Data in the database is organised in columns and rows, the number of columns and rows in the tables must be defined. The columns could also be called “fields”, “items” or “tables”, which might be a bit confusing sometimes, and the rows might also be called “objects”, “post”, “records” or “tuples”.

THE DATABASE STRUCTURE IS DEFINED WHEN THE DATABASE IS CREATED

FIELDS - ALWAYS A DEFINED TYPE:



TEXT
NUMERIC
BOOLEAN
ETC.

AND SIZE:

TEXT: NUMBER
OF LETTERS
NUMERIC: BYTE,
INTEGER,
REAL, ETC.

18

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 18

When creating a new database, it is necessary to define the structure of the different fields (columns), e.g. “type”. What type of information should the users be allowed to enter into the column? If it is defined as text, it should only be allowed to enter ASCII-text and the information should only be used as a text string while defining it as “numerical” makes it possible write in numbers and use this column in calculations. The strict definition of each column reduces possible errors since different data types are separated and mixing is prevented, which is an important difference between a RDBMS and spreadsheet software, where data mixing is normally permitted. It is also necessary to define the size of your column, in the sense of how many characters that are allowed. For example, a text column may contain a maximum of 50 characters and a

numerical field maximum 30 digits with 10 decimal places. You could also define the type of storage for different values, e.g. binary, integer or real.

THE DATA IS STRUCTURED IN TABLES AND THESE
ARE LINKED TOGETHER

- IT'S THIS SORT OF "LINKING" THAT CREATES
THE RELATIONS BETWEEN THE DIFFERENT DATA



19

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 19

Data is structured in tables that are linked together; the link creates the relations between different attribute data. And even though two tables might not be linked directly to each other, they might have a relation to each other through a chain of other tables which are linked together, which makes it possible to combine the information in the two tables. In wider sense, most features (and humans) are linked in a more or less complex way.

A RELATIONAL DATABASE IS BUILT
UPON CERTAIN RULES AND
RESTRICTIONS:

- ✿ NORMAL FORMS
E.G. THE DATA SHOULD BE SIMPLE
VALUES — NOT QUANTITIES
- ✿ FUNCTIONAL DEPENDENCE
IF THE VALUES IN A COLUMN ARE
DEFINED BY THE VALUES IN
ANOTHER COLUMN



20

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 20

When constructing tables in a relational database, certain restrictions must be considered: It's important that only one value is stored in each cell in the table. It is not possible to store age and number of people in the same cell or age and name. Every cell should contain unique values related to that specific entity. Functional dependencies should be avoided, which means that a column that is totally dependent on another column should be avoided. It is probably then better to get rid of the data since it's just a form of transformation using an algorithm on another column.

AVOID REDUNDANCY!
E.G.

Name	City	Country	Capital
Ulrik	Lund	Sweden	Stockholm
Christopher	New York	USA	Washington D.C.
Petter	Lund	Sweden	Stockholm

↓

Name	City
Ulrik	Lund
Christopher	New York
Petter	Lund

City	Country	Capital
Lund	Sweden	Stockholm
New York	USA	Washington D.C.

21
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 21

Redundancy should be avoided as much as possible when creating a database. Redundancy means that the same information is stored many times, which will make the database larger and slower to handle than it needs to be. Normally redundancy is avoided by splitting larger tables into several smaller tables and use common columns to link the tables together. In the example data storage will be much more efficient if the table is divided into two smaller tables that are linked together.

INDEX


- **PREVENTS DUPLICATION**
- **INCREASES THE SEARCH SPEED**
(TELEPHONE DIRECTORY)
- **SORTS (APPARENT)**
- **MAINTAINS THE REFERENTIAL**
INTEGRITY


22
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.


Slide 22

The use of indexes, similar to the ones used in a telephone book, will increase the search speed in a database. Data can e.g. be indexed by storing it in alphabetical order. The creation of an index decreases retrieval time but the index could also be used for linking different tables. Indexes generally make database handling more efficient since they prevent duplications, increase search speed and maintain the referential integrity.

THERE ARE DIFFERENT WAYS OF LINKING TABLES:

1. ONE TO ONE 

2. ONE TO MANY 

3. MANY TO MANY 

23
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 23

There are different ways of linking the data. Logically a relation can be of either one of the three basic types:

1. “One to one”-relation
2. “One to many”-relation
3. “Many to many”-relation

ONE TO ONE

THE VALUES THAT ARE USED TO RELATE THE DIFFERENT TABLES TO EACH OTHER ARE UNIQUE, E.G. THEY CAN ONLY OCCUR ONCE IN EACH TABLE

Order	Name	Address	City	Date	Object	Amount	Price
54	Ulrik	PO box 3	Lund	14/11/02	A34b	4	500: -

Order	Name	Address	City	Date
54	Ulrik	PO box 3	Lund	14/11/02

Order	Object	Amount	Price
54	A34b	4	500: -

Slide 24

“One to one” relations link tables by a unique value that can only occur once in each table. The example has a common “order” column in each table. The order number is unique and can only occur once in each table. This relation can link one row in one table with one row in another table.

ONE TO MANY

ONE POST IN ONE OF THE TABLES CAN CORRESPOND TO MANY IN ANOTHER
THIS IS A GOOD WAY TO AVOID REDUNDANCY

MANY NON-UNIQUE VALUES

Country No.	Country
1	Sweden
2	Norway
3	Denmark

Country No.	City
1	Lund
1	Stockholm
1	Gothenburg
2	Oslo
3	Copenhagen
3	Odense

25
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 25

“One to many” relations occurs when one row in one table can be linked to many rows in another table. This is a typical way of avoiding redundancy in databases. The example above shows two tables. The first table has a column for different countries and a column for the country ID’s. The second table has a column for the country ID’s and one containing names of different cities in each country. Since a country can have many different cities, the country ID can apply to many cities (rows) in the city name table. In a more complex database, the two tables would certainly contain much more information that only has to be stored once using this technique. The first table may store information about the different countries such as: population, area, capital etc. and the second table could store information about the different cities, such as population, size, railway connections etc. To divide the table using a “one to many”- relation is much better than to store everything in the same table, since redundancy is avoided.

MANY TO MANY

THERE ARE NO UNIQUE VALUES IN ANY OF THE TABLES

Weather St.	Year	Month	Rain
100	1996	May	100
100	1996	June	150

Weather St.	Year	Month	Temp
100	1996	May	19
100	1996	June	21

TO LINK THESE THERE MUST BE A MULTIPLE INDEX:
WEATHER ST. + YEAR + MONTH

26
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 26

“Many to many” relations have no unique values in any of the table’s columns. The example shows two tables with information from a weather station. The tables contain station number, year and month for both tables and rainfall and temperature data respectively. In these tables, there are no unique identifiers and in order to link unique values, e.g. for the same month, several columns must be used (station number, year and month).

WHAT HAPPENS IF SOME VALUES ARE MISSING?

TABLE A

No.		
1		
2		
3		
4		
5		

TABLE B

No.		
3		
4		
5		
6		
7		

EITHER "INNER JOIN" OR "OUTER JOIN"

27
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.


Slide 27

Often, missing data creates problems when working with attribute data. If two tables containing missing data are to be merged or joined, this can be done using two different approaches, **inner-join** or **outer-join**.

INNER JOIN

-RECORDS THAT EXIST IN BOTH THE TABLES ARE KEPT

No.	No.
3	3
4	4
5	5



28
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.


Slide 28

When applying an inner-join, only rows that can be linked using a one to one relation for the selected join attribute will be kept. All other records will be omitted from the resulting database. This is one way of omitting data that is not relevant for a specific task when merging databases.

OUTER JOIN

- THE RECORDS THAT EXIST IN ONE OF THE TABLES ARE KEPT

A	B
No.	No.
1	
2	
3	3
4	4
5	5



29
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 29


The “outer-join” operation keeps all records from one of the tables and adds those records from the other table that correspond. In the example, all the records from table A are kept and merged with the corresponding records of table B. The records in table A that do not have any corresponding records in table B will have missing data in the part of the resulting database that comes from table B. This is a way of partially updating your database some fields with some missing values. Depending on how the database is going to be used either the inner join or the outer join could be used. If all records in both databases have the same records (same record identifiers) this problem will not occur.

HOW TO HANDLE A DATABASE

SQL - STRUCTURED QUERY LANGUAGE

USED FOR DIFFERENT SORTS OF ATTRIBUTE OPERATIONS SUCH AS:

1. SEARCH
2. SEARCH AND EDIT/DELETE
3. SELECTIONS



30
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 30


To retrieve data or perform different sorts of operations on a database, you use the Structured Query Language (SQL). SQL is used to search, replace, edit, update or delete attribute data and to create output from your database. It is most often used for selecting parts of a database according to the users specific criteria's that are used in an analysis or in other operations. The SQL uses a strict syntax, and it is important to understand the exact meaning of the conditional operators that are used.

EXAMPLES OF SQL-SENTENCE:

SELECT DEFINES WHAT THE RESULT SHOULD CONTAIN
(OFTEN A SET OF COLUMN NAMES)

FROM DEFINES WHAT TABLES THE DATA SHOULD
BE COLLECTED FROM

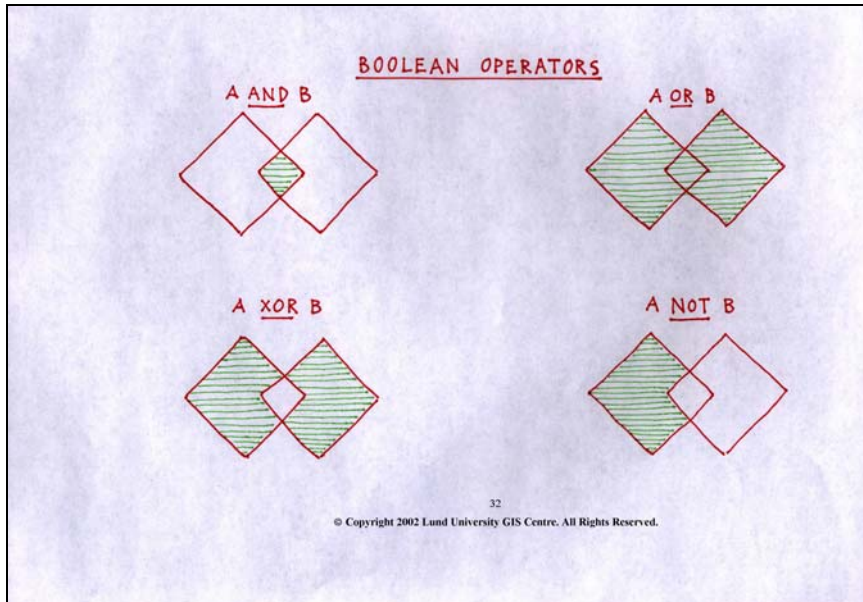
WHERE DEFINES THE ROWS WHICH FULFIL OUR
BOOLEAN CONDITION



31
© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 31

The SQL-sentence consists of three different parts. The SELECT-part defines what attributes should be included in the results, e.g. you are only interested in selecting information from a set of specific columns in a large database, but uninterested in the rest. The FROM-part names the source table or source database from which data should be retrieved. The WHERE-part defines what conditions the records in the database must fulfil to be selected. The conditions could be logical expressions, mathematical operators, etc, and several conditions could be set, using Boolean operators to join them.



Slide 32

A Boolean operator states a condition and the condition is evaluated and returns a result that can be **TRUE** (condition fulfilled) or **FALSE** (condition not fulfilled). The Boolean operators used to join several conditions in the WHERE part of the SQL-sentence are:

AND – looks for conditions that are true in two database or tables. I.e. both condition A **AND** condition B must be fulfilled, e.g. to select parents who have a boy (dataset A) and a girl (dataset B) this operand would be used.

OR – looks for conditions that are true in either one of the datasets or in both, e.g. this operand to select parents who has either a boy **OR** a girl (or both a boy and a girl).

XOR – looks for conditions that are true in one of the databases but not in both, e.g. the operand **XOR** should be used to select parents who has either a boy or a girl, but excludes those with children of both sexes.

NOT – looks for a condition that must be false, e.g. the operand **NOT** is used to select all parents with only boys (and NOT girls) or all parents with girls (and NOT boys).

EXAMPLES:

Table: Students

ID NO.	Name	Subject	Mark
14	Robert S	History A	5
32	Therese G	English B	2
65	Jacob T	History A	4
21	Sara L	Sports	3

Search: Find all the students who study History A

SELECT Name, Subject, Mark FROM Students WHERE Subject = "History A"

RESULT:

Name	Subject	Mark
Robert S	History A	5
Jacob T	History A	4

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 33

This is an example of search or retrieval of data using by SQL. To find all the students that studies history A, SELECT the fields that should be in the result, i.e. "Name", "Subject" and "Mark" FROM the table "Students", WHERE the condition is "Subject = history A". The computer search the "Subject" field in the table "Students" for records that fulfils the where-condition (the condition is True), these rows are selected and retrieved together with the corresponding data in the fields for "Name" and "Mark".

SELECT * FROM Students WHERE Name = Robert Strand AND Mark > 4

RESULT:

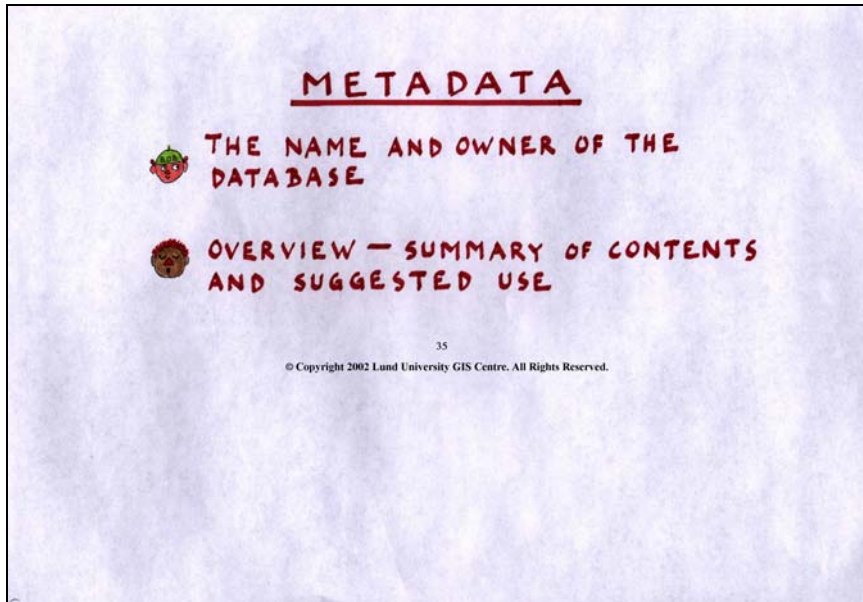
ID NO.	Name	Subject	Mark
14	Robert S	History A	5

34

© Copyright 2002 Lund University GIS Centre. All Rights Reserved.

Slide 34

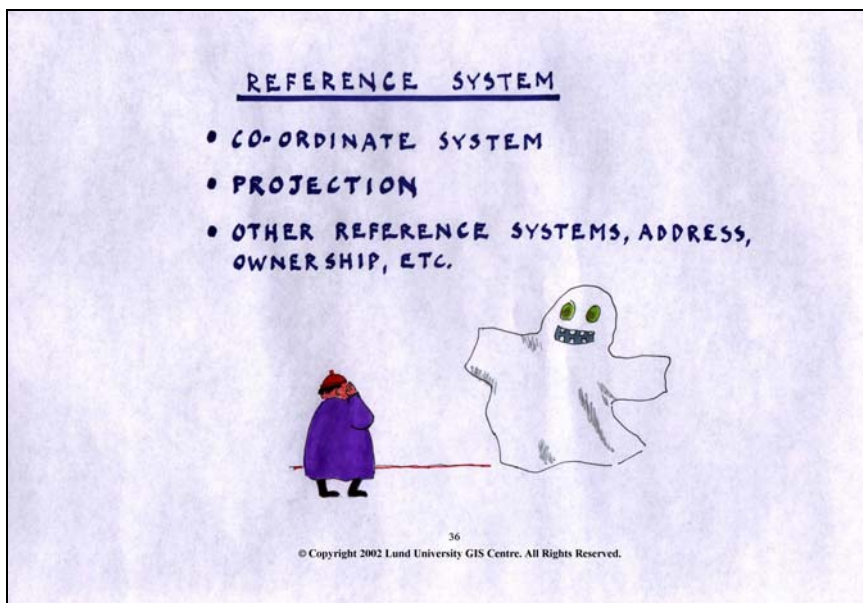
This is a multiple condition SQL example selecting all fields using the wildcard sign * in the SELECT statement, FROM the table "Students", WHERE the condition is "name = Robert Strand" AND "mark > 4". The result is only one record since Robert Strand obviously only has one mark greater than four.



Slide 35

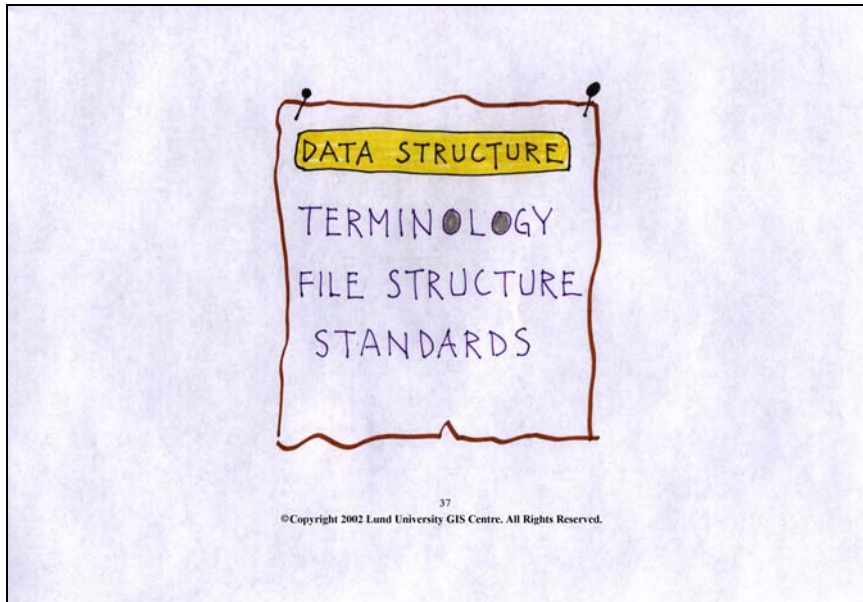
Metadata is information about the data in the database (both for geometrical data and attribute data). Metadata is a very important issue, since database documentation is important to sustain a certain standard of a database. It assures a certain level of quality, reliability and flexibility for the users of the database. Most countries have or are in the process of establishing standards for metadata and the following slides gives an example that has been adapted from the European standard:

1. Name of the database and the owner of the database
2. Overview of geometrical and attribute data and the suggested use of the database.



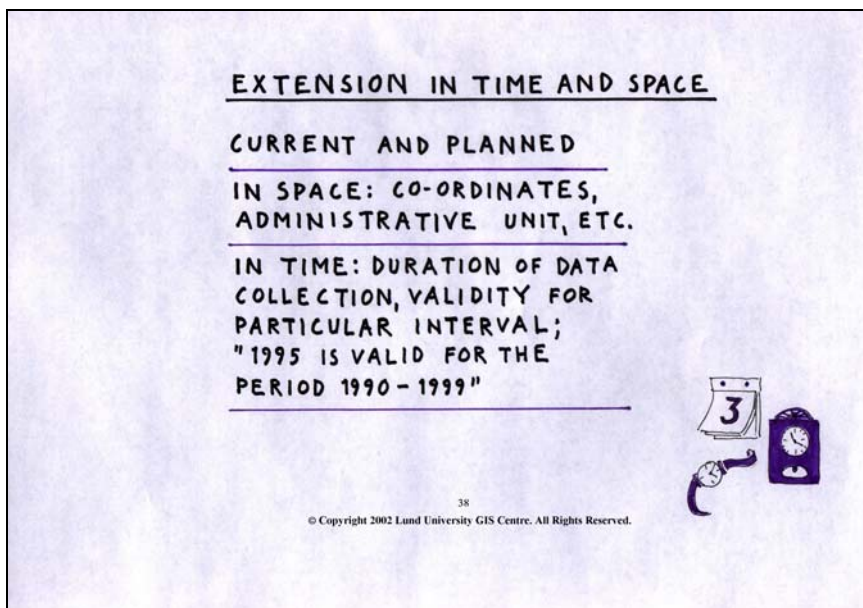
Slide 36

Metadata should also include cartographic information, such as projection, co-ordinate system, map datum etc. When such information is missing it is difficult to know what kind of projection to use without deforming the data or how to transform databases between different co-ordinate systems, etc.



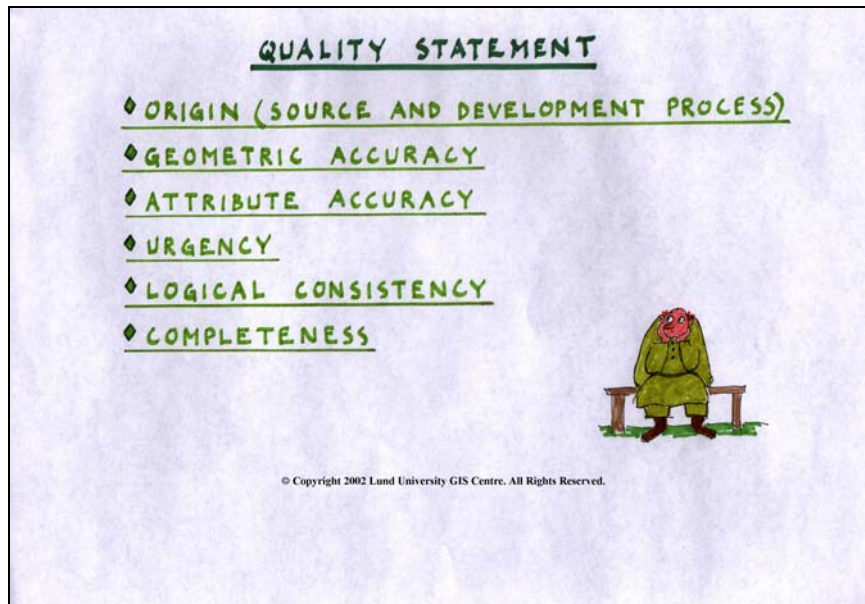
Slide 37

The structure of the database, e.g. how data is stored (both geometrical data and attribute data), what terminology, file structure, the data format, etc. that have been used for the database is also important to document (to facilitate data exchange between different systems) and the standards for storing, terminology, file format, etc that has been used.



Slide 38

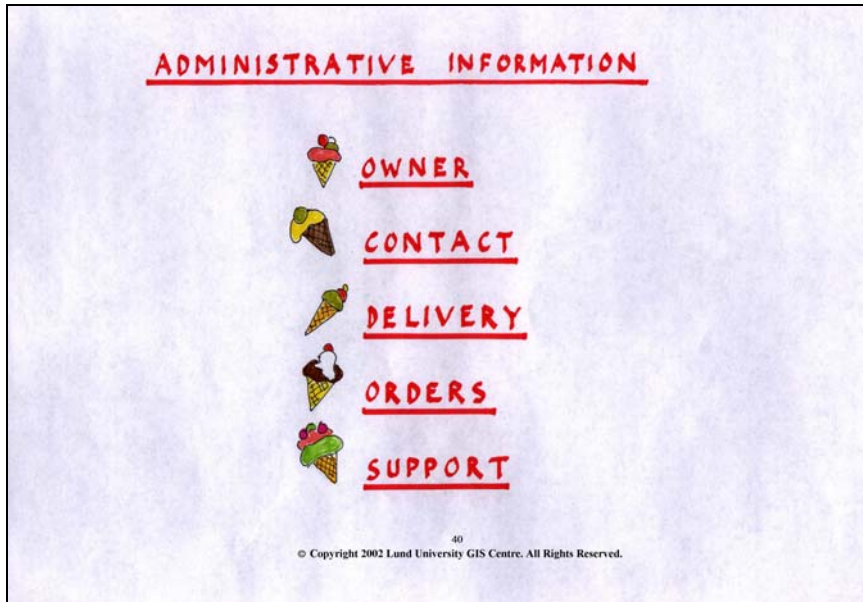
Actual Extension in time and space of the database as well as planned extensions should also be part of metadata. Date(s) of data collection and interval of updating are also important features to describe in the metadata. This information will help a user to determine if the database is appropriate for use. It is also an opportunity to inform the user about future extensions of the database. Some databases, e.g. census data, may be valid for a certain time span, which should be indicated in the metadata.



Slide 39

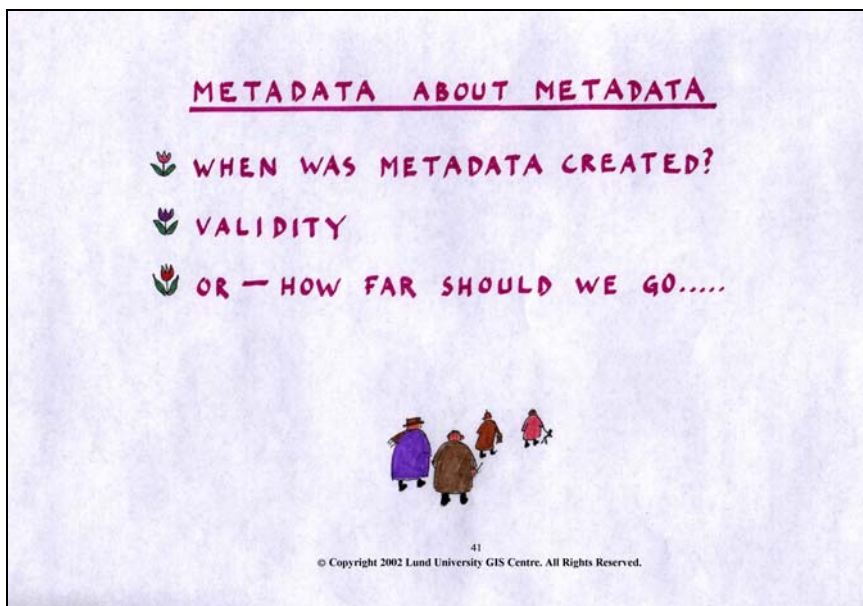
Metadata is often considered as equivalent to quality statements, and naturally, these are a very important component of the metadata.

- Origin of the database including a description of the source of data and the database development process – information about where the data was retrieved, from and how it has been processed before entering the database.
- Geometric accuracy - geometric data collection methodology and with what accuracy the geometry was measured, e.g. it can be assumed that data collected using advanced field survey equipment is more accurate than data collected from aerial photographs. This affects the co-ordinate accuracy. For certain applications, error-terms as Root Mean Square (RMS), or similar are used.
- Attribute accuracy – as for geometry, the data collection methodology will influence on accuracy. Is the attribute an actual measurement, estimation or interpolation of attribute values.
- Urgency – temporal accuracy of the database.
- Logical consistency – database production process, lineage of production as well as operations employed in the process. Topological statements, spaghetti topology or full polygon topology.
- Completeness – is the above true for the whole area covered by the database or are there differences in quality, temporal accuracy, consistency ,etc.



Slide 40

As a service for the presumptive user of the database, the metadata should also include administrative information about the database. Since, at least in most cases, databases are sold, not given away for free, this is one way of advertising to future clients, to whom to address their inquiries and how to order data, delivery formats and media (by e-mail, CD-ROM, diskette), additional services that the producer can provide (e.g. analysis).



Slide 41

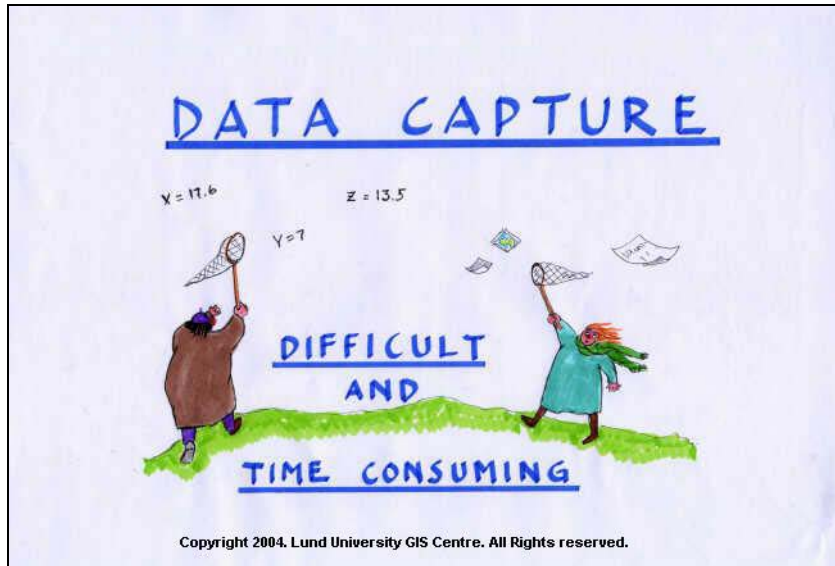
There is also such a thing as metadata about metadata. This is important since a database probably will be updated and the metadata therefore needs to be updated as well. But it is also important to keep in mind that the primary task is not metadata and the complexity of the metadata must be reasonable, in order to prevent the producing agent's staff from producing only metadata instead of the database it self!



LUND
UNIVERSITY

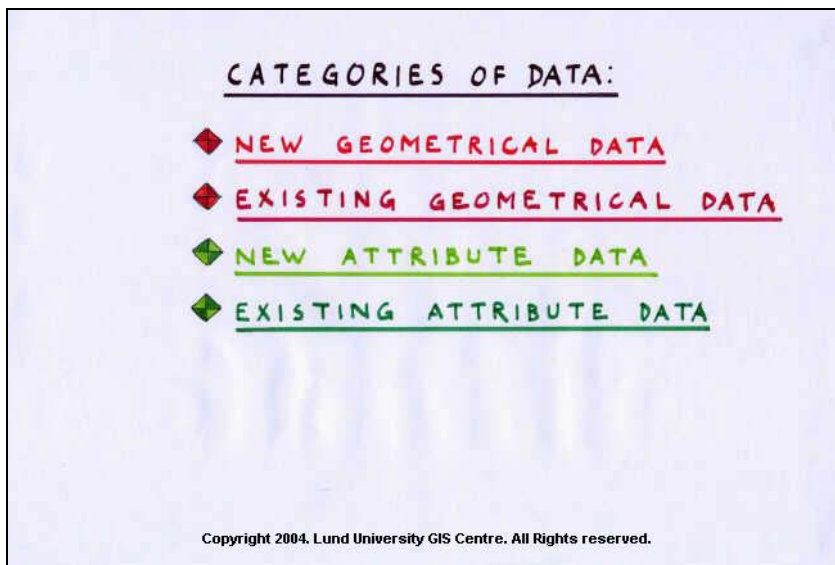


DATA CAPTURE



Slide 1:

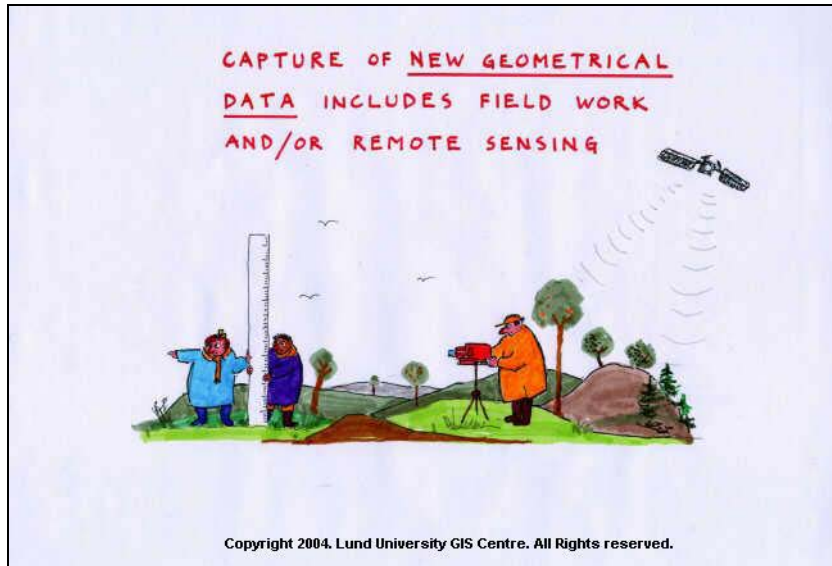
Without geographical data, it is impossible to work with a GIS. The capture and import of data are often the most time consuming, expensive and difficult tasks within the whole chain of work in the GIS environment.



Slide 2:

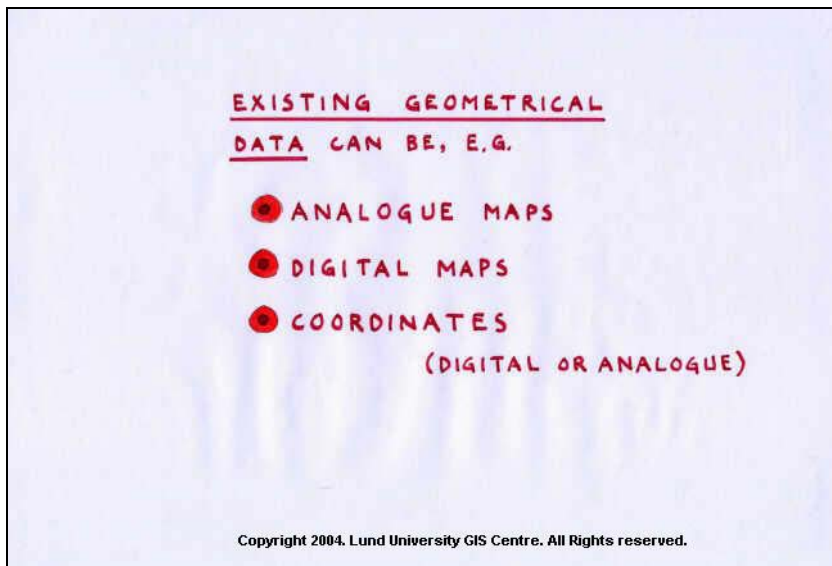
The data capture can be divided in four categories, i.e. capture of:

- new geometrical data;
- existing geometrical data,
- new attribute data, and
- existing attribute data.



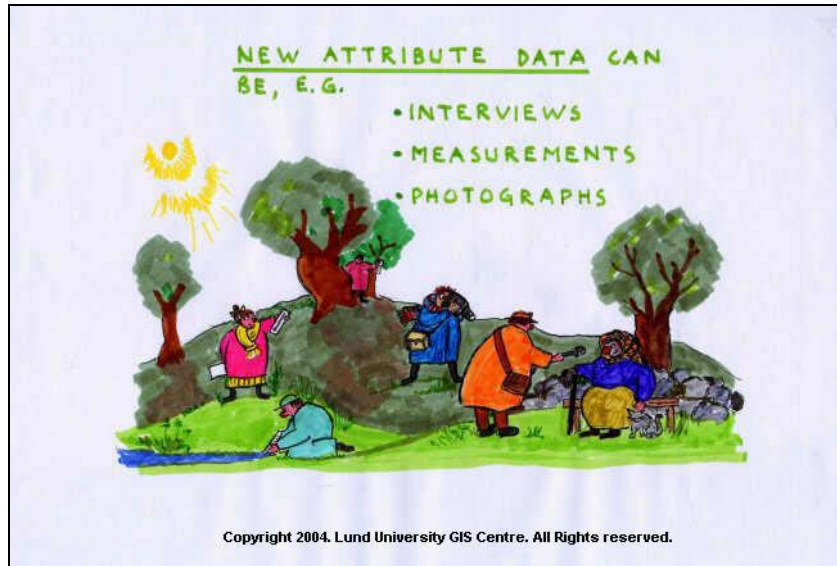
Slide 3:

New geometrical data are equal to objects (i.e. positions of objects) that we want added to our database and that have not yet been mapped in the field. This means that fieldwork is often needed to get the positions of these “new” objects. An alternative to fieldwork is the use of remote sensing, like aerial photographs and satellite imageries.



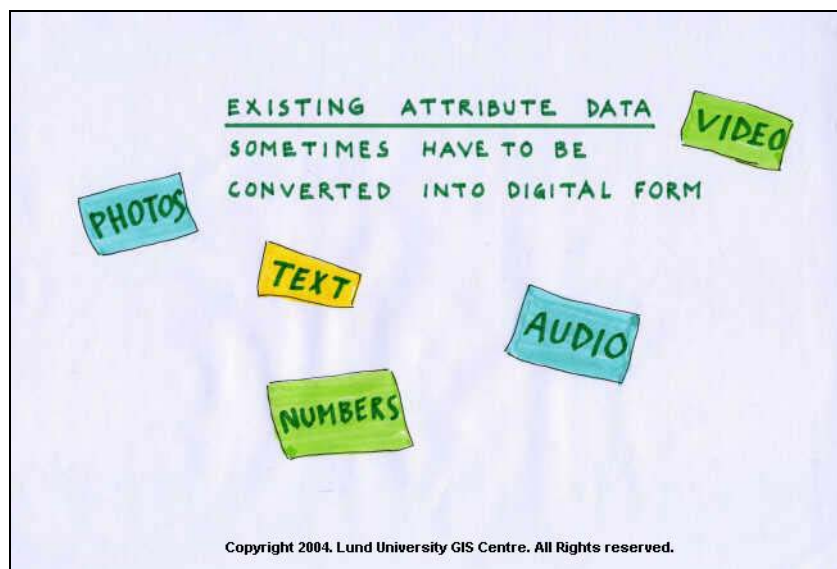
Slide 4:

When we talk about existing geometrical data, we refer to e.g. analogue maps on paper or plastic film, digital maps and coordinate tables on paper or in digital format. These data have in common that they are not yet imported into the geographical information system. Thus, the data has to be imported, and before that, sometimes also converted into digital form.



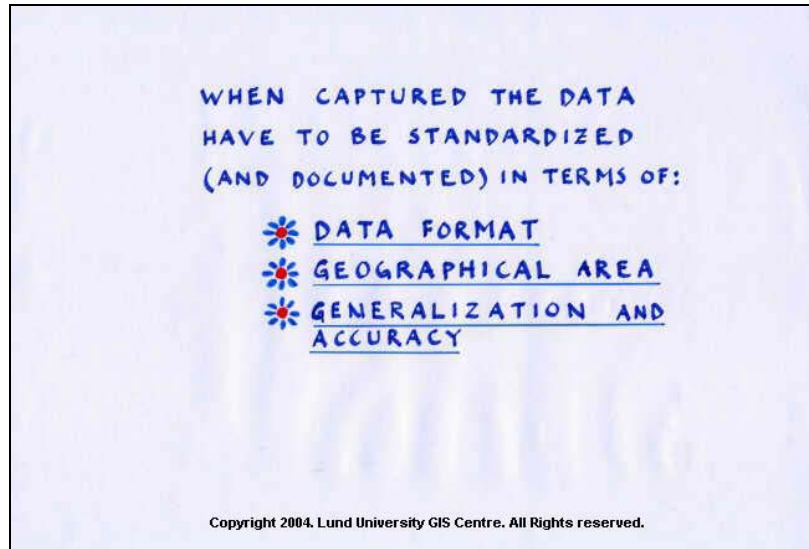
Slide 5:

New attribute data can be captured e.g. by interviews, measurements of variables of interest and photographs of the geometric objects that we want to document. Quite often, this data capture involves data sampling. At this stage, we want to remind the reader that the sample should always be representative, meaning that it should reflect the population in the best possible way.



Slide 6:

Existing attribute data could be e.g. text, numbers (tables), audio, video and photographs in either digital or analogue form. Analogue data have to be converted into digital form (e.g. by typing or scanning, see below) and imported into the GIS. Digital data have to be formatted in order to be compatible with the used GIS. Even if most systems can read the standard formats, it is common that one gets data stored in an unusual way in order to fit specific software. The user then has to figure out how the data are stored, and sometimes manually (e.g. by writing a small program) convert it into the desired form.



Slide 7:

When captured and imported into the GIS, all the data types have to be thoroughly documented. If possible they should be standardised in terms of:

- data format (in order to be imported, the data have to be in a format that the software accepts);
- geographical area (if the data do not cover the same area, they cannot easily be combined in the GIS), and;
- generalisation and accuracy (it is not appropriate to mix data with different levels of generalisation and/or accuracy. Imagine e.g. what will happen if we try to superimpose a very generalized road network on the top of a detailed “map” showing property boundaries).

It is very important that the data capture can be repeated. In order to do this, standards and documentation are needed.



Slide 8:

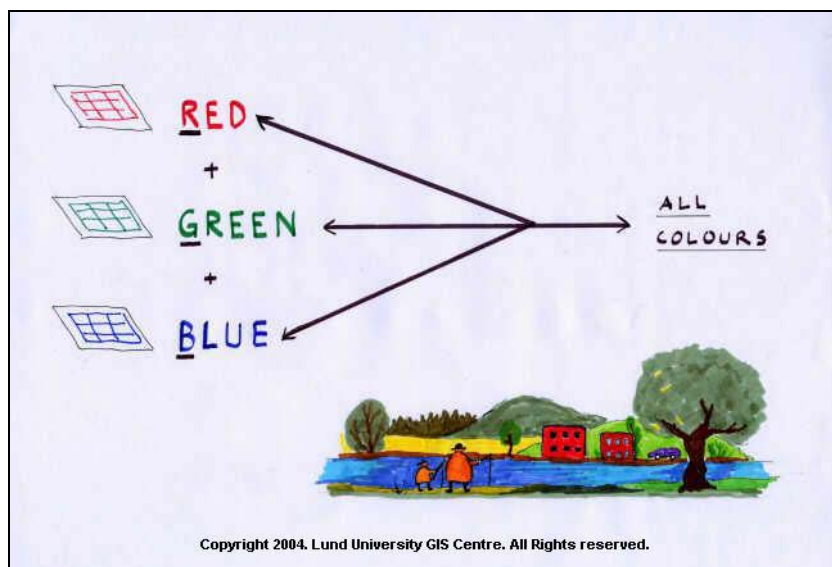
This is an example of a satellite image. Each cell in the image is 20 x 20 metres. Satellite data are almost always captured in digital form and delivered as raster files. From satellites, you can get grey scale as well as colour data. However, satellites are often used to register colour (i.e. reflection of light from an object in a certain wavelength interval) in non-visible wavelengths. Both shorter wavelengths, like ultra violet (UV), and longer wavelengths, like infra red (IR) and thermal radiation, can be registered. IR can be used to detect variations in vegetation status on the ground. This example is an infrared registration, where high reflection of infrared light appears as red

(vegetated agricultural land). A standard altitude of a land resource satellite (like the one in this example) is approximately 700 km.



Slide 9:

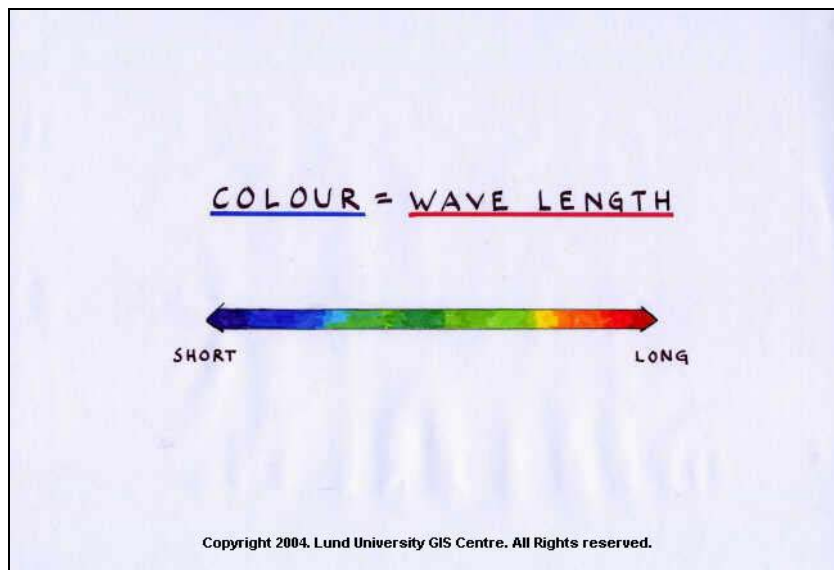
As mentioned above, new geometrical data are also often captured by the use of aerial photographs. Aerial photographs can be delivered in either analogue or digital form. If they are analogue, they have to be digitised before imported. This is usually performed by the use of a scanner that creates one or three raster data layers. If the photograph is black and white only one layer is needed, containing the “level of brightness” (i.e. degree of white, 0 – 100%) in all the raster cells. Normally, a cell value of 0 corresponds to black, a cell value of 127 corresponds to “medium grey” and a cell value of 255 corresponds to white (if 8 data bits are used). After digitalisation, documentation data (often referred to as metadata) have to be created. These data tell the system and the user about the geographical position of the raster layer (e.g. x_{min} and y_{min}), cell size, data quality, projection, data type etc. If data are to be stored in vector form, they have to be of course vectorized (see below). Aerial photographs are normally registered from an altitude of 4000 to 6000 metres.



Slide 10:

If a colour photograph is scanned, three raster files are created, one explaining the red colour in the cell, one explaining the green colour and one explaining the blue. When these layers are combined

the “true” colour from the original image can be presented on the screen or as a printed map. Satellite images and aerial photographs can, like any digitised colour photograph, be divided into different files representing different colours (i.e. wavelength bands), for example red, green and blue (RGB).



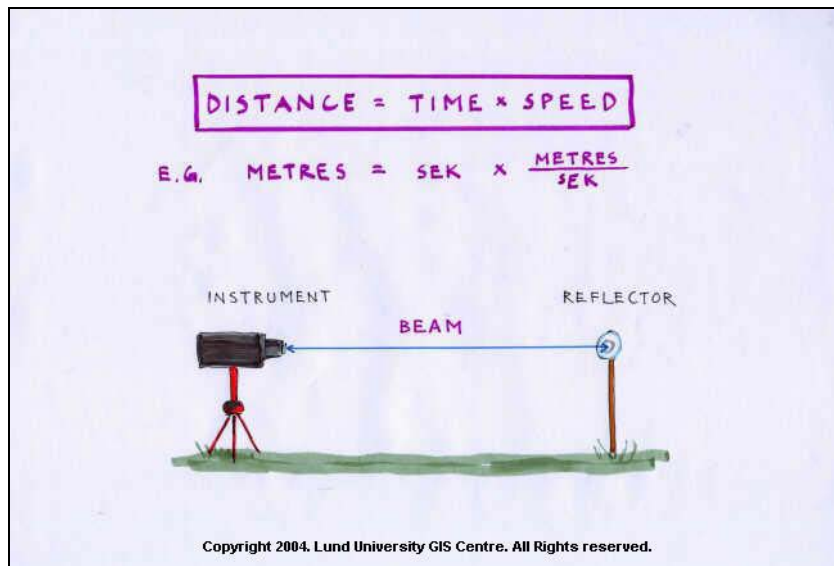
Slide 11:

Even if we often talk about colour, what the satellite sensor and the camera really measure is reflected light. Light with shorter wavelengths appears as blue, while light with longer wavelengths appears as red.



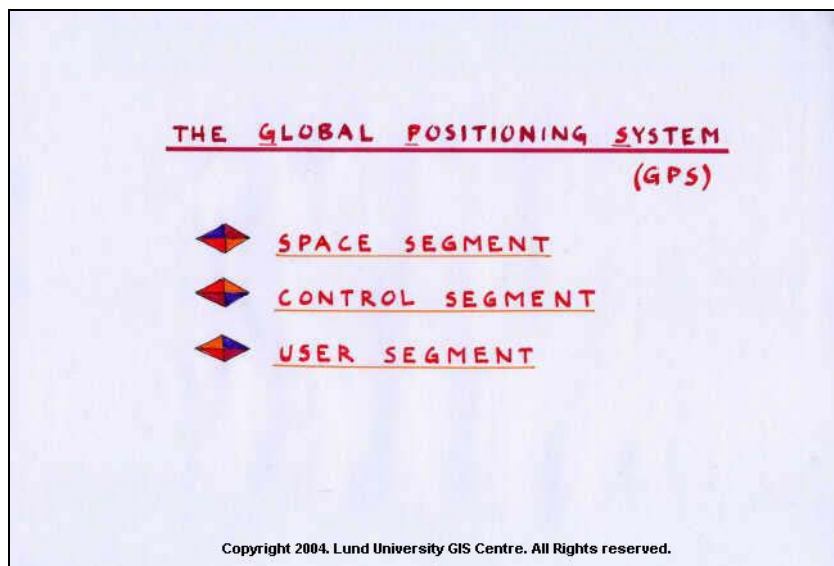
Slide 12:

Fieldwork (measurements on the ground) is basically performed in three ways, by “subjective” updating of an existing map, by the use of a total station (or other ground measurements) or by the use of the global positioning system (GPS). The easiest way to collect data is to observe the object to be mapped in field and then draw it on the top of an existing map. Depending on the quality of the map and the geometrical demands on the data, this can sometimes be acceptable. If the new data is added to an already existing GIS database in digital form, no more processing is needed. If the data is added to a paper map, it has to be digitised and imported into the GIS (see section concerning “capture of existing geometrical data” below).



Slide 13:

A total station is an instrument that actively sends out a beam of e.g. infra red light that is reflected at the object that will be measured, and then receives and records the reflected light. Since the speed of the beam is known and the station measures the time difference between sending and receiving the beam, the distance to the object can be estimated (distance = time difference/2 * speed). The total station also has a built in compass that records the angle (normally in degrees from North) to the object. Apart from this, the total station measures the horizontal angle from the instrument to the object. If the position (in x, y and z) for the total station is known the position of the object can easily be estimated by the use of distance, direction and horizontal angle. The user can either construct his or her own coordinate system (e.g. decide that the station is placed in x=0, y=0, z=0) or use at least two known points to decide the position of the station in an existing coordinate system.

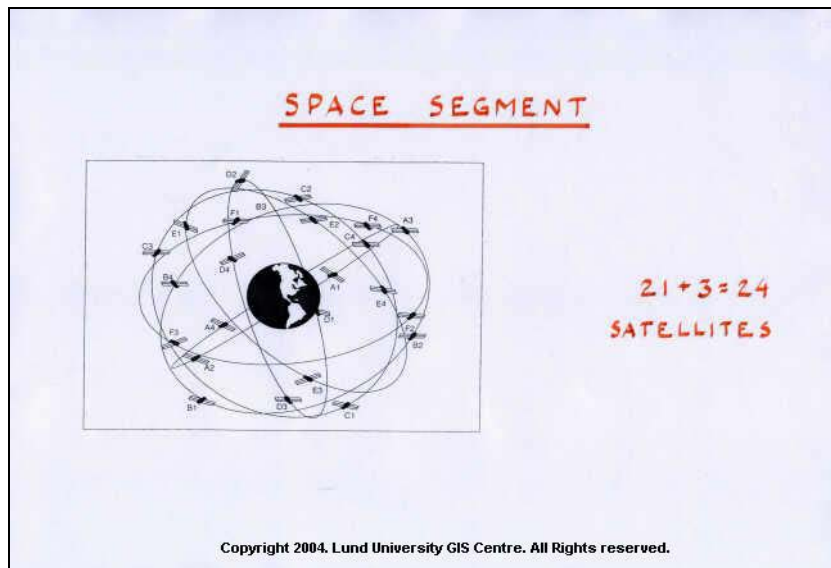


Slide 14:

The global positioning system was ready to use in the mid 80's. It is developed and owned by the United States Department of Defence (USDD) and consists of three parts:

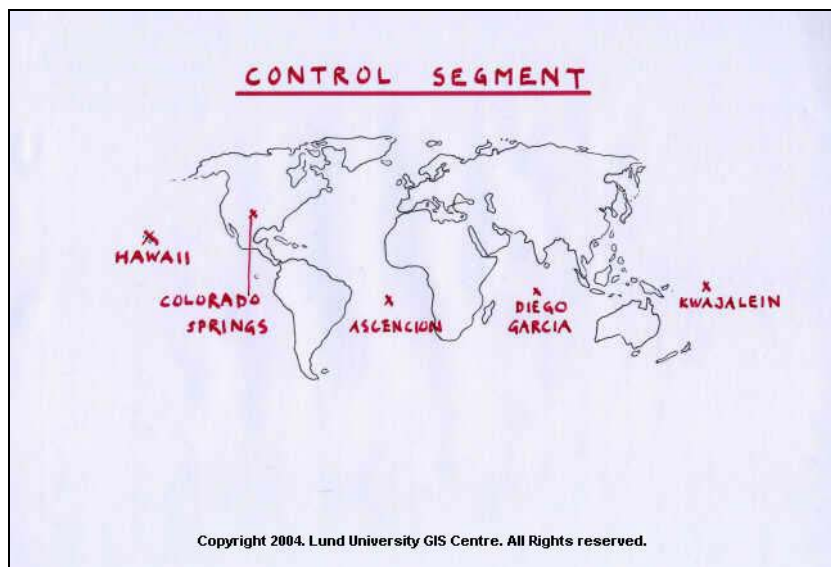
- the space segment (satellites);
- the control segment (ground stations), and

- the user segment (receiver).



Slide 15:

The space segment consists of 24 satellites (called the NAVSTAR satellites), of which 21 are in use and the others are spare, circulating around the Earth at an altitude of approximately 20 200 km. The satellites are arranged in 6 orbital planes with 4 satellites in each plane. All orbital planes have an inclination of 55 degrees relative to the Earth's equator. The satellites complete one orbit around the Earth in approximately 12 hours. Each satellite is equipped with four atomic clocks and measures data about its position at different times. The satellites continuously send signals (pulses) down to Earth including exact time when the pulse was sent and satellite position. The position is given in x, y and z WGS84. Independently to location and time on the Earth's surface, at least 4 satellites are always in view, at least 15 degrees above the horizon.



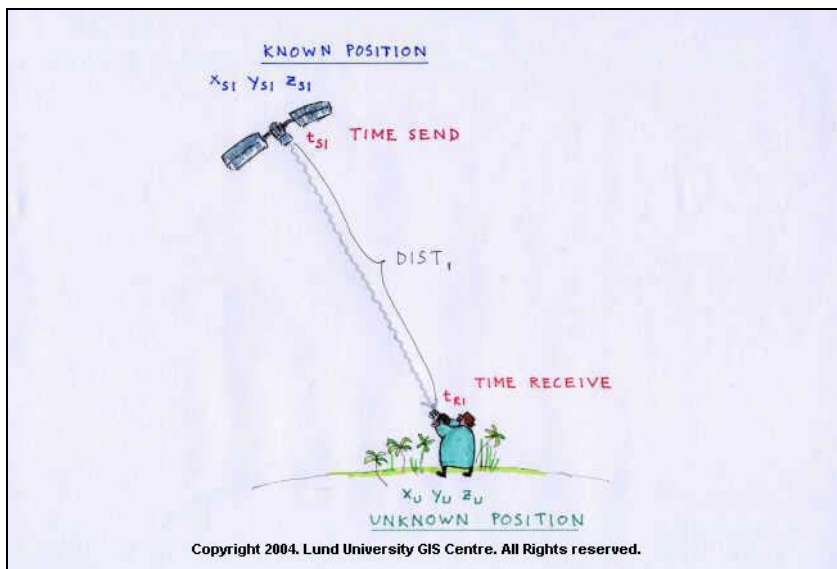
Slide 16:

The satellites are controlled from 5 ground stations distributed around the equator. The main station is located in Colorado Springs, USA. The purpose of the ground stations is to monitor and correct the positions of the satellites (that are pre-designated), the atomic clocks (that have to be correct and synchronized for all satellites) and the data transmission.



Slide 17:

The user segment consists of one or more GPS receivers that have the possibility to receive data from many satellites (depending on number of bands) at the same time. The receiver is equipped with a clock and has the calculation ability to estimate its position. A standard hand-held receiver is the size of a large mobile phone.



Slide 18:

In most cases, an unknown position is estimated by a method called code measurement. This is based on the knowledge of distances to, and positions of, at least four satellites. An example of a code measurement is illustrated in the figure above. x_u , y_u and z_u denote the unknown position of the receiver. x_{s1} , y_{s1} and z_{s1} denote the positions of satellite number 1. t_{s1} denotes the transmission time for the signal from satellite number 1, and t_{r1} denotes the receiving time for the signal from satellite number 1. $Dist_1$ denotes the distance to the satellite.

TIME FOR SIGNAL SAT. → RECEIVER:

$$dt = t_{R1} - t_{S1} + b \quad \searrow \text{CLOCK ERROR (RECEIVER)}$$

SPEED OF SIGNAL:

c (SPEED OF LIGHT, APPROX. 3×10^8 m/s)

DISTANCE SAT. → RECEIVER:

$$\text{DIST}_1 = c \times (t_{R1} - t_{S1} + b)$$

Copyright 2004. Lund University GIS Centre. All Rights reserved.

Slide 19:

The time it takes for the signal to travel from the satellite to the receiver can be estimated by subtracting t_{R1} by t_{S1} . We can call this difference dt . However, since the clock in the receiver is not synchronized to the satellite clocks, dt has to be corrected by an unknown factor, here called b . The travel time for the signal (dt) then equals $t_{R1} - t_{S1} + b$, and can be expressed in e.g. seconds. Since the speed of the signal is known ($c = \text{speed of light} \sim 3 \times 10^8$ m/s) we can estimate the distance to a satellite by multiplying c and dt . In the figure above we are estimating the distance to satellite number 1.

ALTERNATIVE SOLUTION, DISTANCE SAT. → RECEIVER:

$$\text{DIST}_1 = \sqrt{(x_{S1} - x_U)^2 + (y_{S1} - y_U)^2 + (z_{S1} - z_U)^2}$$

(THEOREM OF PYTHAGORAS IN THREE DIMENSIONS)

COMBINATION OF THE TWO WAYS:

$$c \times (t_{R1} - t_{S1} + b) = \sqrt{(x_{S1} - x_U)^2 + (y_{S1} - y_U)^2 + (z_{S1} - z_U)^2}$$

↑ KNOWN
 |
 KNOWN

 ↑ KNOWN
 |
 KNOWN

 ↑ KNOWN
 |
 UNKNOWN

 ↑ KNOWN
 |
 UNKNOWN

 ↑ KNOWN
 |
 UNKNOWN

 ↑ KNOWN
 |
 UNKNOWN

4 UNKNOWN

Copyright 2004. Lund University GIS Centre. All Rights reserved.


Slide 20:

Another way to estimate distances is by applying the theorem of Pythagoras in three dimensions. Above, this is done for satellite number 1. The two ways can be combined, and we get an equation with 6 known (c , t_{R1} , t_{S1} , x_{S1} , y_{S1} , and z_{S1}) and 4 unknown (b , x_U , y_U and z_U) factors.

TO ESTIMATE 4 UNKNOWN FACTORS
WE NEED 4 EQUATIONS IN AN
EQUATION SYSTEM

THIS MEANS

WE NEED 4 SATELLITES
TO ESTIMATE OUR POSITION
IN X, Y AND Z



Copyright 2004. Lund University GIS Centre. All Rights reserved.

Slide 21:

If we want to estimate all the 4 unknown factors, we need data, and consequently equations, from 4 different satellites. These equations then create an equation system that can be easily solved by the software in the receiver.

ONE UNKNOWN

E.G. $x = 17 + 5$
 $x = 22$

WE NEED ONE
EQUATION

TWO UNKNOWNNS

E.G. EQUATION SYSTEM $\begin{cases} x = 15 + y \\ x = 3y \end{cases}$

$3y = 15 + y$
 $2y = 15$
 $y = 7.5$
 $x = 15 + 7.5$
 $x = 22.5$

WE NEED TWO
EQUATIONS

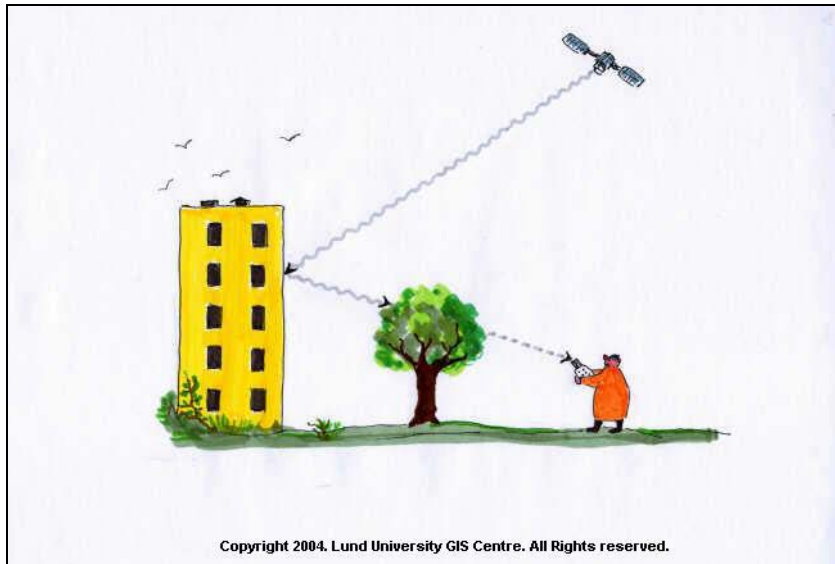


AND SO ON...

Copyright 2004. Lund University GIS Centre. All Rights reserved.

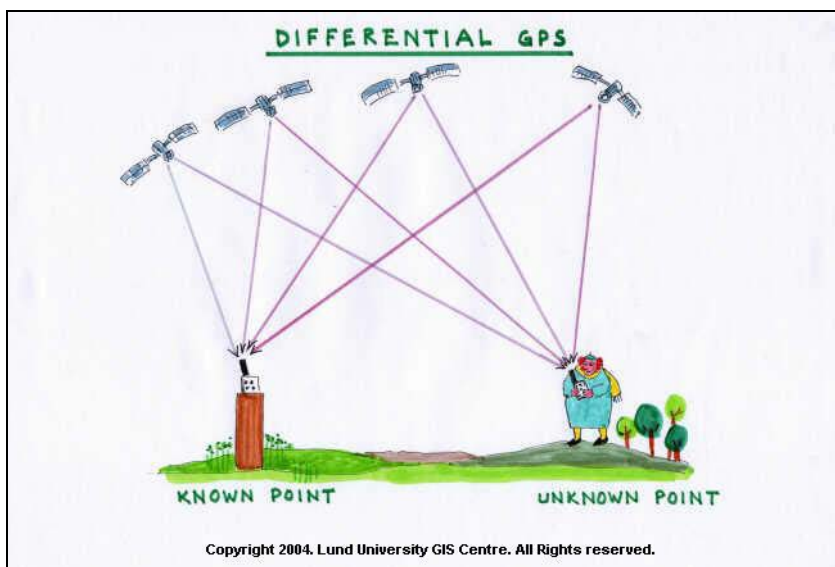
Slide 22:

Of course, an equation with one unknown factor only needs one observation to be solved. An equation with two unknown factors needs two observations to be solved (see above), an equation with three unknowns needs three observations and so on. This is obviously the reason why at least four satellites have to be in view when using a GPS receiver.



Slide 23:

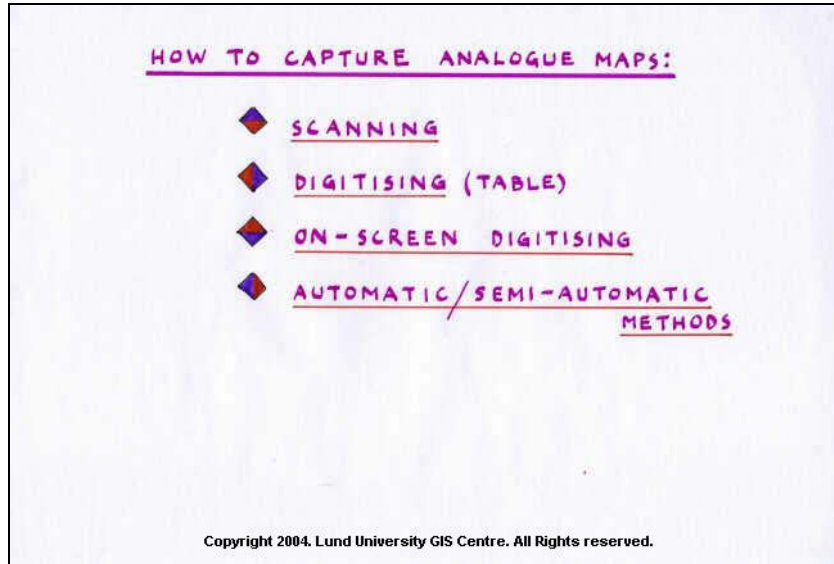
Note that the satellite signal can be blocked, e.g. by roofs and dense forest. It is also worth mentioning that the more satellites that are used simultaneously, the better the precision in the estimation of the position of the receiver. Apart from the number of satellites, the quality of the estimation depends on e.g. errors in satellite positions, variation in velocity of the signal, and reflection of satellite signals. The mismatch between a satellite's predicted position (which is the one that is sent down to the receiver) and the actual one can be up to 80 metres. This in combination with disturbance related to deviations from the speed of light for the signal when it passes the troposphere and the ionosphere can decrease the accuracy by approximately 10 metres. Reflections of the signals against, for example buildings, can decrease it even more.



Slide 24:

One way to increase the quality of the estimation is the use of differential GPS. The idea of differential measurements is that a GPS receiver located on a KNOWN position is recording satellite data, from the same satellites, at the same time as another receiver is performing the “real” measurements near by. The error in the estimation can then be estimated by examining the difference between the known position and the position that has been obtained by using GPS. The differences can then be applied to the real measurements. This methods yields position with an accuracy of approximately 1 metre or better. If even better estimations are wanted, it is possible to

use carrier-wave measurements. This method demands more advanced (and thus more expensive) receivers, but yields better accuracy. The code measurement is based on phase differences of the carrier wave (with a wave length of approximately 20 cm) and generally demands longer measurement time and post processing of the data. Another way to (to less extent) improve the measurement is to combine GPS and other, similar, systems like the Russian GLOSSNAS (Global Navigation Satellite System) or the planned European GALILEO system.



Slide 25:

Analogue maps are normally entered into the GIS either by scanning or manual digitising using a digitising table. An alternative is a combination where the map is scanned and then digitised manually on the screen. Also, automatic and semi-automatic methods can be used. These are often based on methods that follow lines automatically.

<u>SCANNING</u>	<u>DIGITISING TABLE</u>
+ FAST	+ CONTROL
+ CAPTURES EVERYTHING	- SLOW
- CAPTURES EVERYTHING	- EXPENSIVE

<u>ON-SCREEN DIGITISING</u>
+ CONTROL
+ POSSIBLE TO ZOOM
+ WE "SEE" WHAT IS DONE
+ CHEAP
- SLOW

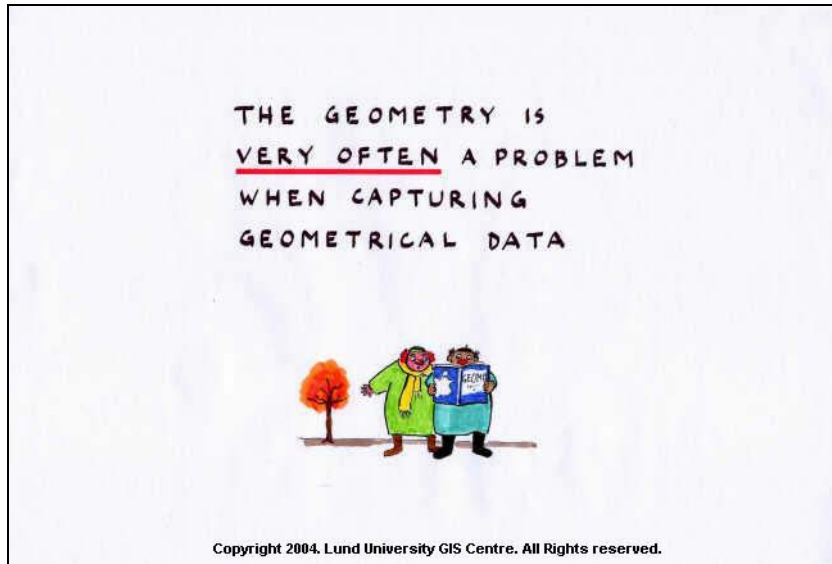
Copyright 2004. Lund University GIS Centre. All Rights reserved.

Slide 26:

When scanning a map, the data are automatically transformed into raster form, either as one data layer or, if preferred, into three layers representing the colours red, green and blue. After scanning additional documentation (metadata) have to be added. These are e.g. coordinates of lower left corner, cell size, number of rows and number of columns. The main advantages of scanning are that the process is fast and that the scanner does not miss any data that is present on the map. However, the later advantage can also be seen as a disadvantage. The user might not want to capture EVERYTHING on the map. Things like text, dirt and geometrical objects of no interest then have to be manually removed. This editing procedure can be very time consuming. Actually, scanning paper maps is rarely cost efficient. However, scanners are excellent when converting attribute data, e.g. old drawings and photographs, into digital form.

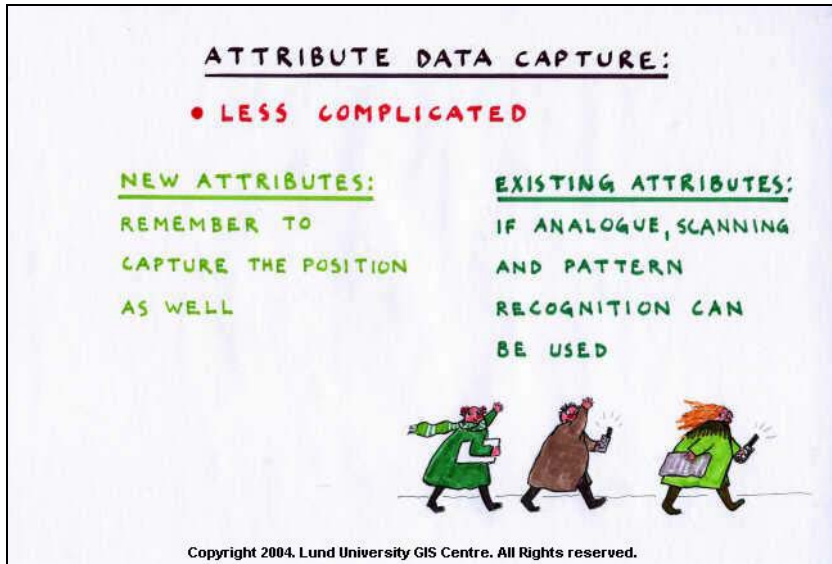
A data capture method often preferred is manual digitising by the use of a digitising table. The paper map is then fixed on the surface of the table and a cursor is used to register the positions of the geometrical objects of interest. First of all, a number (minimum 4) of known points are digitised on the table and their true coordinates are typed in from the keyboard. The purpose of this is to establish the relationship between the coordinate system of the digitising table (often in mm from a lower left origin) and the coordinate system used on the map. Then "new" objects can be registered on the digitising table and their map coordinates can be calculated. Even if manual digitising is a relatively slow process it is often preferred because of the possibility for the user to decide what data to capture.

Since it is often faster and easier to digitise directly on the screen compared to using a digitising table, a combination of scanning and manual digitising has evolved. This, on-screen digitising, uses a scanned image in raster format that is displayed on the screen, either alone or combined with other layers. User-friendly software can then be used to digitise points, lines and polygons by the use of the mouse and the cursor.



Slide 27:

Independently of selected method, the main concern when capturing existing analogue geometrical data (as well as all other geometrical data) is always the geometry. Even if the projection of the map is known, tension in the paper has often changed the geometry enough to reach an unacceptable level of matching to other data in the same projection. Additionally, inaccuracies in the scanner optics and in the digitising table often lead to the need of a more comprehensive geometrical correction of the data. One alternative of doing this is so called “rubber sheeting”, where corresponding points are identified, in the new (digitised and/or scanned) dataset as well as in the geometrically correct one (often an existing layer in the GIS database). By comparing the coordinates, a polynomial that transforms the new dataset into the desired projection can be established. This procedure demands advanced GIS/image processing software.



Slide 28:

Capture of attribute data is often less technically complicated than geometrical data capture. We are often more familiar with the methodologies, and huge amounts of data is already available in digital format. The main problem is then to convert the data into the right format and link it to the geometrical objects in the GIS. For analogue data, there are possibilities to, instead of “typing it in” manually, scan the data and then use text/pattern recognition to create e.g. text. If used in a GIS, all data have to be linked to a geographical position.