# Table of contents

- Learning objectives........................................................................................................2
- Geographic Information Systems................................................................................2
- GIS Software................................................................................................................3
- GIS architecture and functionality................................................................................4
- Spatial data infrastructure..........................................................................................4
- Database management systems...................................................................................5
- Reasons for using a DBMS ..........................................................................................6
- Alternatives for data management..............................................................................7
- Strengths and weaknesses of GIS and DBMS..............................................................8
- GIS and database.........................................................................................................9
- Future of GIS...............................................................................................................10
- Summary....................................................................................................................11
Learning objectives

At the end of this lesson, you will be able to:

- indicate the main functional components of a GIS;
- recognise why database are used in a GIS; and
- describe interrelations between databases and GIS.

Geographic Information Systems

For many years, analog data sources were used, processing was done manually, and paper maps were produced. The introduction of modern techniques has led to an increased use of computers and digital information in all aspects of spatial data handling.

The software technology used in this domain is centered around geographic information systems (GIS).

A geographic information system provides a range of capabilities to handle georeferenced data including:

1. Data capture and preparation
   Typical planning projects require data sources, both spatial and non-spatial, from different national institutes. The data sources obtained may be from different time periods, and the spatial data may be in different scales or projections.

2. Data management
   With the help of a GIS, the spatial data can be stored in digital form in world coordinates, allowing proper integration. This makes scale transformations unnecessary, and the conversion between map projections can be done easily with the software.

3. Data manipulation & analysis
   With the spatial data thus prepared, spatial analysis functions of the GIS can be applied to perform the planning tasks.

4. Data presentation
   What remains is to pay careful attention to the quality level of the different datasets, to ensure that during data presentation unnecessary error is not introduced.
GIS Software

The main characteristics of a GIS software package are its analytical functions, which provide means for deriving new geoinformation from existing spatial and attribute data.

There are many GIS packages available on the market and all of them have their strengths and weaknesses. Some GISs have traditionally focused more on support for raster-based functionality, others more on vector-based spatial objects. We can safely state that any package that provides support for only rasters or only vector data is not a complete GIS.

Well-known, full-fledged GIS packages include ILWIS, QuantumGIS - an open-source GIS -, Intergraph’s GeoMedia, and ESRI’s ArcGIS.

We will attempt to describe the field of GIS independently from such packages, as ‘principles’ should be useful to users of any package.

Characteristics of the main GIS software packages

There is no single GIS package that is necessarily better than another: it all depends on factors such as intended application, and the expertise of its user.

ILWIS’s traditional strengths are in raster processing and scientific spatial data analysis, especially in project-based GIS applications.

Intergraph, ESRI and MapInfo products have been known better for their support of vector-based spatial data and their operations, user interface and map production (a bit more typical of institutional GIS applications).

QuantumGIS supports both raster and vector data, provides a collection of plugins for geoprocessing, analysis and visualization. In addition, QuantumGIS is multiplatform (can be used with Windows, Linux, Mac operating systems), is free of charge and open-source (i.e., it can be modified by adding functionality).

Any such brief characterization, however, fails to do justice to any of these packages, and it is only after extended use that their strengths, and sometimes weaknesses, will become clear.
GIS architecture and functionality

A geographic information system in the wider sense consists of software, data, people, and an organization in which it is used. All these elements enable several functional components that support key GIS functions.

We should consider that, in addition to functional components, geographic information systems are impacted by organizational factors. These define the context and rules for the capture, processing and sharing of geoinformation, as well as the role that GIS plays in the organization as a whole.

For a particular GIS, each of the functional components may provide many or only a few functions. Arguably, the system should not be called a GIS if any one of these functional components is missing.

It is important to note, however, that the same function may be offered by different components of the GIS: for instance, data capture and data storage may have functions in common, and the same holds for data preparation and data analysis.

Spatial data infrastructure

Later in this lesson we will be focusing on storage and maintenance and the role of spatial databases, but before we move on to that, let’s review how organizational components impact on a GIS.

Many organizations are forced to work in a cooperative setting, in which geographic information is obtained from, and provided to, partner organizations and the general public. The sharing of spatial data between the various GISs in those organizations is important, and aspects of data dissemination, security, copyright and pricing require special attention.

The design and maintenance of a Spatial Data Infrastructure (SDI) addresses these issues.

SDI - “In (Nebert, 2004) an SDI is defined as the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data”.

Fundamental to those arrangements are, in a wider sense, the agreements between organizations and, in the narrow sense, the agreements between software systems on how to share and hand over the geographic information. Standards are often the starting point for those agreements.
Standards exist for all facets of GIS, ranging from data capture, data description to data presentation.

They have been developed by different organizations, of which the most prominent are the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC).

The collection of technologies, policies and institutional arrangements that compose an SDI provides the different facilities for finding, viewing, downloading and processing data.

Because the organizations in an SDI are normally geographically distributed, computer networks are used as the means of communication. With the development of the internet, the functional components of GIS have gradually become available as web-based applications.

Much of the functionality is provided by so-called geoservices, software programs that act as an intermediate between geographic datasets and the users of the web. For their spatial data handling, these services commonly use standardized raster and vector representations following the main standards.

Database management systems

Data storage and maintenance in a GIS are implemented as a database and executed with a database management system.

A database is a large, computerized collection of structured data.

In the non-spatial domain, databases have been in use since the 1960s, for various purposes such as bank account administration, stock monitoring, salary administration, order bookkeeping, and flight reservation systems.

These applications have all in common that:

- The amount of data is usually quite large;
- the data itself has a simple and regular structure; and
- there is a large end-user community.

Designing a database is not an easy task. First of all, one has to consider carefully what the database purpose is, and who its users will be.

Secondly, one needs to identify the available data sources and define the format in which the data will be organized within the database. This format is usually called the database structure.
Lastly, data needs to be entered into the database in a consistent way. It is important to keep the data up-to-date, and it is therefore wise to set up the processes for this, and make someone responsible for regular maintenance of the database.

A database management system (DBMS) is a software package that allows the user to set up, use and maintain a database. Like a GIS allows the set-up of a GIS application, a DBMS offers generic functionality for database organization and data handling.

**An example of a DBMS** - Many standard PCs are equipped with a DBMS, albeit of the toy type. One of the most common examples is part of the Microsoft Office package, and is called MS Access. MS Access offers a useful set of functions, and the capacity to store and operate on data stored in tables. It is not fit for enterprise-wide applications, however, and for these many other DBMSs are available.

### Reasons for using a DBMS

There are various reasons for using a DBMS for data storage and processing.

- **A DBMS supports the storage and manipulation of very large data sets**

  Some data sets are so big that storing them in text files or spreadsheet files becomes too awkward for use in practice. The result may be that finding simple facts takes minutes, and performing simple calculations perhaps even hours. A DBMS is specifically designed for doing this efficiently.

- **A DBMS supports the concurrent use of the same data set by many users**

  Many people are usually involved in the data collection, maintenance and processing. These data sets are often considered to be of a high strategic value to the owner(s). Moreover, for different users of the database, different views on the data can be defined. In this way, users will be under the impression that they operate on their personal database, and not on one shared by many people. They may all be using the database at the same time, without affecting each other’s activities. This DBMS function is called **concurrency control**.

- **A DBMS supports the use of a data model**

  A data model is a language with which one can define a database structure and manipulate the data stored in it. The most prominent data model is the relational data model. Its primitives are
**tuples** \(^1\) (also known as records, or rows) with attribute values, and relations, being sets of similarly formed tuples.

- **A DBMS provides a high-level, declarative query language**

In a database, a **query** is a computer program that extracts from the database that data that meets some specified conditions. A query language can be used to define queries and updates.

- **A DBMS can be instructed to guard over data correctness**

We can ensure that the data that is entered into the database does not contain errors. This is generally known as **integrity constraints**. More complex integrity constraints are certainly possible, and their definition is part of the design of a database.

- **A DBMS includes data backup and recovery functions and the control of data redundancy**

Regular back-ups of the dataset and automatic recovery schemes provide an insurance against loss of data. Storing a fact multiple times, a phenomenon known as **data redundancy**, can lead to situations in which stored facts may contradict each other, causing reduced usefulness of the data. A well-designed database takes care to store single facts only once.

Redundancy, however, is not necessarily always problematic, as long as we specify where it occurs so that it can be controlled.

**Alternatives for data management**

The decision to use a DBMS will depend, among other things, on how much data there is, what type of use will be made of it, and how many users might be involved. Let’s look at an example. Amina is a statistician working for an international NGO.

She keeps all useful addresses in a personal address book on her computer. This list contains all her contacts. When the data set is small, its use relatively simple, and with just one user, simple text files, and a text processor may be used.

Text files offer no support for data analysis whatsoever, except perhaps in alphabetical sorting.

---

\(^1\) Tuple - in the relational model is formally defined as a finite function that maps attributes to values. For example: \((\text{player} : "Harry", \text{score} : 25)\)

*(From Wikipedia)*
Amina has recently collected a number of field observations with measurements that she wants to prepare for statistical analysis.

In this case, since Amina's data set is still small and numeric by nature, and she has a single type of use in mind, a spreadsheet program will suffice. Spreadsheets do support some data analysis, like averages, sums, minimum and maximum values. All such computations are usually restricted to just a single table of data.

However, when data become a large number, spreadsheets may not be enough.

Amina has carried out a nationwide census, with many observation stations and/or field observers and all sorts of different measurements.

She needs to analyse, manipulate and prepare these data for visualization on a map.

In this case, spreadsheets do not accommodate concurrent use of the data set well, although they do support some data analysis, especially when it comes to calculations over a single table, like averages, sums, minimum and maximum values.

When one wants to relate the values in the table with values of another nature in some other table, some expertise and significant amounts of time are usually required to make this happen inside the spreadsheet, whereas a database enables to quickly keep track of all the data.

**Strengths and weaknesses of GIS and DBMS**

How do GIS and databases interact? Let’s consider their different characteristics.

<table>
<thead>
<tr>
<th>GIS</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIS software packages provide support for both spatial and attribute data. They accommodate spatial data storage using the vector and raster approaches and attribute data using tables.</td>
<td>Database management systems (DBMS) have been based on the notion of tables for data storage.</td>
</tr>
</tbody>
</table>

Currently, all major GIS packages provide links to a database that stores attribute data with and make use of its superior data management functions.

Spatial (vector or raster) and attribute data are still sometimes stored in separate structures, although they can now be stored directly in a spatial database.
A GIS has spatial data represented with rasters or vectors, and the attribute data stored in an external DBMS.

The DBMS serves as a centralized data repository for all users, while each user runs her/his own GIS software, obtaining its data from the DBMS.

With raster representations, each raster cell stores a characteristic value. This value can be used to look up attribute data in an accompanying database table.

With vector representations, our spatial objects - whether they are points, lines or polygons, or multipoints, multilines, or multipolygons - are automatically given a unique identifier by the system. This identifier is usually just called the ‘objectID’ or ‘featureID’ and is used to link the spatial object (as represented in vectors) with its attribute data. The principle applied here is similar to that in raster settings, but in this case each object has its own identifier.
Future of GIS

While database systems originated from a very early need for sharing data amongst distributed users, GIS (as software) originally was developed as a single-user system.

In that context, since the internet did not even really exist as we know it nowadays, it was not strange at all that GIS developed its own rather special file formats to hold data and GIS project information.

Over time, though, the need to share spatial data and GIS project information amongst multiple users became more and more apparent, first within project teams, later within enterprises, and then across enterprises, and nowadays even amongst large communities of often mobile users, who hardly know each other.

The consequence of this natural development is that GIS has matured with the variety of software architectures, and there are no GIS software "solutions" to all actual architectures: Enterprise GIS, server GIS, web GIS, mobile GIS are all terms that demonstrate these developments.

The modern GIS user, should be aware of the sharing facilities that GIS-as-software nowadays offers through these respective architectures. In the near future, many more possibilities appear to arise.

We see:

- developments in volunteered geographic information (VGI), in which data is obtained from non-authoritative sources, sometimes ordinary citizens;
- developments in deepening the semantic content of metadata, through the use of semantic web tools;
- developments in using cloud computing for spatial application purposes, allowing building large-scale systems without immediate large investments in hardware and software.

And more is yet to come, in this century of the digital age.
Summary

Data management and processing functions are central to GIS. In this lesson we have provided an overview of DBMS and geodatabase technology.

Traditionally, GIS is more suited for data management and processing and DBMS better for organizing spatial and attribute data. As a result, a GIS was often linked to an external DBMS for substantial applications or projects requiring more powerful attribute data management capabilities. Spatial databases are a marriage of GIS and traditional DBMS.

They support storage and manipulation of both geometry and attribute data, and allow spatial queries. The functions and capabilities of spatial databases are constantly improving. In the near future, it is likely that we will use spatial databases exclusively for storage of all geometric and attribute data.