Overview

1. A GIS is NOT a Map!
2. How does a GIS handle its data? Data Formats!

GIS = Map + Database

Vector  Raster  TIN  Table

Point  Line  Polygon  Cell  Triangle  Column  Row
1) A GIS is NOT a Map!

A GIS is not a map, but rather a database that (can) create a map!

The MS Excel Analogy to GIS

In MS Excel the data are stored as a table with columns and rows – essentially a database.

You can enter data, look at data, change data, etc. inside the table. You can also perform calculations and analysis, for example the average of a range of cells or more advanced statistical calculations.

You can also create graphs to display the data in your table. Graphs are helpful to visualize trends, patterns, or correlations.

*In MS Excel, the graph and the underlying data table are separate, but linked:*

If you change a cell value in the table, the linked graph changes accordingly. If you change the graph (graph type, colors, symbols, labels, etc.) the data in the underlying table does NOT change!

There is a one-way street connecting the data and its visualization! Changing the data changes the visualization – which is good! But, changing the visualization does NOT change the underlying data – which is also good!

A GIS is NOT a Map!

The same applies to a GIS: A GIS is geospatial data stored in database tables and you, the user, can decide to create a (linked) map from those data.

- You can create whatever map you want and make the map look the way you want to…the underlying data remain unchanged.
- But, if you change your underlying data…your map will change accordingly.

Maps are great and creating maps is one of the unique capabilities of a GIS. But maps are not the only way sometimes not the best way to present the results of a GIS analysis.

- You do not have to create a map. You can also create graphs, tables, or reports – whatever makes the most sense in a given situation.
- Maps are great and very often you will use your GIS to create a map. But, sometimes it is better to present the results of your GIS analysis as a graph or table.
The Big Difference MS Excel vs. GIS!

You can create a map with a GIS because your data are both spatial and attribute data = geospatial data. MS Excel handles only attribute data – MS Excel cannot create maps. Well…at least not in the way a GIS can. In fact, you can create maps with MS Excel, but that’s a topic for another class.

Consider this Example:

This ‘map’ shows the results of a land use change analysis for West Springfield (MA). This is not really a map, but rather a canvas or poster compiling various ways you can present data and analysis.

- 5 frames with maps.
- 1 table summarizing the results, created in MS Excel.
- 2 pie charts illustrating one important aspect of the results, created in MS Excel.
- Text explaining the analysis and the results.
- A written report documenting the analysis, presenting the results, and discussing the implications.

The entire poster with the 5 map frames and associated legends was created in ArcGIS. The table and pie charts were created in MS Excel and copied/pasted into ArcGIS.

➤ That’s a GIS!
2) How does a GIS handle its data? Data Formats!

Storing geospatial data (= spatial and attribute data) is a bit trickier than storing just attribute data (which is all that MS Excel has to do).

**Spatial Data**

- Where is it? (= Location, latitude and longitude)
- What is it? (= Geometry, point, line, or polygon)
- What’s next to it? (= Topology, adjacency, etc.)

- Examples of spatial data: the location and geometry of a wetland (polygon), a river (polyline), or utility pole (point).
- Spatial data tells us where something is located, but not what it is.

**Attribute Data**

- What are its characteristics?

Attribute data tells us what a feature is (i.e. its characteristics), but not where it is located. Attribute data are stored in a database table; hence the term attribute table is used in GIS

- Examples of attribute data: the height of a mountain and its rock type, the biodiversity of a wetland, and whether a road is paved, dirt, private, etc.

⇒ The power of a GIS is that it can handle, display, and analyze both types of data together. You can ask “Where is something?” and “What does it do?”

⇒ The GIS stores and link the spatial and the attribute data. All of that happens automatically and you do not really need to think about it…most of the time.

---

**Carsten’s House: Spatial Data**

- Street Address: 110 Haywood St
- Greenfield, MA 01301, USA
- Geographic Coordinates
  - Latitude: 42.598631°N
  - Longitude: -72.586614°W
- Now you know where my house is, but nothing more.

**Carsten’s House: Attribute Data**

- 4 bedrooms, 2 baths
- red roof
- 2-car garage
- 2,000 square feet
- natural gas heat
- etc.
- Now you know what my house is, but not where it is located.
You can imagine two different approaches for storing geospatial data!

**Plan A: The ‘Keep-Everything-Separate’ Approach**
Here, everything (spatial data, attribute data, projection files, photographs, tables, raster data, etc.) is kept separate and stored as individual files, grouped as families, and linked by so-called index files. That makes sense, because it neatly separates, for example, the spatial data from the attribute data.

- That’s what GIS users are (still) doing right now with shapefiles and we have a special software program called ArcCatalog to help us organize the many individual parts of shapefiles.
- Shapefiles have a disadvantage…they are difficult to keep organized. On the other hand, each file itself is simple and you can always figure out what is going on.

**Plan B: The Kitchen Sink Approach**
Here, everything is stored in a single, very complicated database table. That sounds appealing, since your GIS is located inside one single file and all you need is a GIS software to read that file.

- In fact, that’s where we are with ArcGIS10. Esri introduced a new data format called a geodatabase to serve effectively as the kitchen sink. GIS users are gradually changing over to this new data format.
- Geodatabases have some disadvantages: they tend to use up a lot of space on your computer and are so complex that it is almost impossible to figure out what is going on inside the geodatabase.

➤ **Geodatabases are the future…get used to them!**

However, shapefiles are still very common in the GIS world and will not go away in the foreseeable future…thus we are dealing with a bit of a messy situation in terms of GIS data formats.

You can easily import shapefiles as feature classes into a geodatabase and vice-versa. The same is true for raster data, networks, etc.

### 2.1) File Geodatabases and Feature Classes

You can think of a file geodatabase as a container holding all kinds of GIS data sets. For example, vector data are stored inside geodatabases as so-called feature classes.

- However, the structure of geodatabases is so complicated and complex that you need ArcCatalog to look inside a geodatabase.
- Try it in Windows Explorer for fun…all you see incomprehensible chaos!
The figure below shows the feature classes contained in a file geodatabase called United States.gdb.

- Each feature class is one vector data set (point, line, or polygon).
- In ArcCatalog you can preview the geography and attribute table of each feature class to get an idea of what the feature class contains before adding it to ArcMap.
- Using Add Data…you can add each feature class to your map as a new map layer.

Features Classes contained in the file geodatabase United States.gdb

There are three types of geodatabases:

- **File geodatabases**: Stored as folders in a file system. Each dataset is held as a file that can scale up to 1 TB in size. The file geodatabase is recommended over personal geodatabases.
- **Personal geodatabases**: All datasets are stored within a Microsoft Access data file, which is limited in size to 2 GB.
- **ArcSDE geodatabases**: Stored in a relational database using Oracle, Microsoft SQL Server, IBM DB2, IBM Informix, or PostgreSQL. These multiuser geodatabases require the use of ArcSDE and can be unlimited in size and numbers of users.

➤ Always use a file geodatabase!

Additional Information

2.2) Shapefiles

Most GIS data from MassGIS and others are still distributed as shapefiles...thus we need to learn how to handle shapefiles.

A shapefile has to include (at minimum):

- The geographic data of the spatial feature.
- The attribute data of the spatial feature.
- An index, linking the geographic data to the attribute data.

Therefore – you can:

1) Look at a shapefile as a map layer (= map the geographic data).
2) Look at a shapefile as a table (= view the attribute data).
3) Use the attribute data to change how the associated geographic data are displayed on the map.

Example

The file format for MS Word is *.doc = a MS Word document is stored on your computer as a *.doc file. The file format for GIS data is a shapefile, meaning that GIS data are saved on your computer as shapefiles.

Here is a screen-shot of a folder called C: \ GIS\tutorial \ UnitedStates. There are 28 files visible in Windows Explorer...but only actual 4 shapefiles: States, Counties, Cities_dtl, and Cities.

Here is a close-up and you see 7 individual files, but only 1 actual shapefile: States. That's a bit weird...but that's the way it is.
A shapefile is actually not a single file, but a family of 3 to 7 individual files.

**There are three mandatory files for each shapefile:**
- States.shp: The shapefile ‘proper’
  - The geographic data: shape + location = geometry of the features
- States.dbf: A data base table with associated attribute data of the features
- States.shx: A spatial index file to link the geometry of features to their attributes

**There are several optional files for each shapefile:**
- .sbn/.sbx: Additional index files as needed
- .prj: The projection file
- .shp.xml: Metadata for the shapefile (= information about the data)

We are used to a different situation!

For example, you type a paper in MS Word and save it as a single *.doc file. This one single file contains all your text, all your figures, all your tables – everything. That single file you can attach to an Email and, for example, send your professor.

Well, GIS data are handled differently. The States shapefile is not a single file, but a family of seven files – even though we refer to it simply as the States shapefile. Therefore you need to combine or zip the individual files together as a single archive for sharing or downloading.

**Example with ArcCatalog**

Below is a screen-shot of the C:\GIStutorial\United States folder, but this time viewed in ArcCatalog.

- ArcCatalog manages and organizes the GIS data.
- ArcCatalog only shows you the States.shp file and ‘hides’ the rest of the files that make up the shapefile family.

This makes it easier to organize your files. You only deal with 1 file and ArcCatalog takes care of the rest.
GIS Analysis Part 1 - A GIS is NOT a Map!
2.3) Who Cares?

*Good point – we have to!*

GIS means using and mapping geospatial data and therefore we need to deal with these different file types. In a few years shapefiles will be obsolete, but in the meantime we need to be able to use geodatabases/feature classes and shapefiles.

- Most MassGIS data is distributed as shapefiles.
- You can easily import shapefiles as features classes into a geodatabase in ArcCatalog.
- You can easily export a feature class as a shapefile using ArcCatalog.

*Fun with your new language!*

- GIS data = geospatial data, stored as feature classes inside file geodatabases or as shapefiles.
- Feature classes (or shapefiles) are organized in ArcCatalog.
- Feature classes (or shapefiles) are added as map layers to a map in ArcMap.
- The map you create in ArcMap does NOT contain the feature classes (or shapefiles) underlying each map layer, but a link back to the location where the actual feature classes (or shapefiles).

➤ *GIS map and GIS data are independent, but linked!*

<table>
<thead>
<tr>
<th>GIS Data</th>
<th>Geodatabase Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>Feature dataset containing feature classes</td>
</tr>
<tr>
<td>Shapefile</td>
<td>Feature class</td>
</tr>
<tr>
<td>Raster data (e.g., satellite images, air photos, scanned maps, and digital pictures)</td>
<td>Raster dataset and/or raster catalog</td>
</tr>
<tr>
<td>CAD data</td>
<td>Feature dataset containing feature classes</td>
</tr>
<tr>
<td>Surface modeling or 3D data</td>
<td>Terrain</td>
</tr>
<tr>
<td>Utility network data (e.g., water systems, gas pipelines, and telecommunication networks)</td>
<td>Geometric network</td>
</tr>
<tr>
<td>Transportation network data (e.g., street networks)</td>
<td>Network dataset</td>
</tr>
<tr>
<td>GPS coordinates</td>
<td>Table of x,y coordinates that can be generated into a feature class</td>
</tr>
<tr>
<td>Survey measurements</td>
<td>Cadastral fabric</td>
</tr>
</tbody>
</table>
The Geodatabase Offers A Comprehensive Approach to Modeling and Managing Spatial Data

With the geodatabase, all of an individual user’s or organization’s GIS data can be stored in a uniform format, in one central location, for easy access and management.

The geodatabase (GDB) is designed to make full use of the capabilities of ArcGIS Desktop and ArcGIS Server. It is not just another spatial data format that can be used by ArcGIS – it is an integral part of the ArcGIS system.

GIS Data in the Geodatabase

- Attribute Table
- Feature Class
- Cartographic Representation
- Annotation
- Dimension
- Relationship Class
- Raster Dataset
- Raster Catalog
- Topology
- Geometric Network
- Network Dataset
- Terrain
- Locator
- Survey Dataset
- Toolbox

Multiuser GDB

RDBMS and ArcSDE Technology

Single-user GDB

File GDB

Personal GDB

Enterprise GDB

Workgroup GDB

Desktop GDB

Key Benefits of the Geodatabase:

- Store a rich collection of data types in a centralized location.
- Apply sophisticated rules and relationships to the data.
- Define advanced geospatial relational models (e.g., topologies, terrains, networks).
- Maintain integrity of spatial data.
- Work within a multi-user access and editing environment.
- Integrate spatial data with other IT databases.
- Easily scale your storage solution.
- Support custom features and behavior.

Functionality

Versioning

Editor 1

DEFAULT

Project 1

Editor 2

Project 2

Geodatabase Replication

Parent Replica

Synchronize

Child Replica

Enables GIS data to be shared across two or more geodatabases. Data changes can be made in each geodatabase, then synchronized. Two-way, one-way, and check out/check in replication workflows are supported.

Geodatabase Archiving

January

February

March

Historical Version

When working on a dataset, archiving captures any and all changes made to the dataset in the DEFAULT version of the multiuser geodatabase.

www.esri.com/geodatabase