**Why GIS?**


Many organizations now spend large amounts of money on geographic information systems (GIS) and on geographic databases. Predictions suggest billions of dollars will be spent on these items over the next decade. Why should this be true now when only a few years ago GIS was a rarity?

Rapidly declining computer hardware costs have made GIS affordable to an increasingly wider audience. More importantly, we have come to realize that geography (and the data describing it) is part of our everyday world; almost every decision we make is constrained, influenced, or dictated by some fact of geography. We send fire trucks to fires by the fastest available routes. Our central government often awards grants to local governments based on population. We study disease by identifying areas of prevalence and rate of spread. This demand for geographic information parallels the need for GIS, explaining its rapidly growing popularity.

Such generalizations do not, however, explain why and how a GIS can help you. First, you must know what a GIS is and what it can be used for. This lesson addresses topics to help you understand GIS; specifically,

- What is a GIS?
- Questions a GIS can answer
- Some applications of GIS
- The components of a GIS
- Using a GIS

The estimated time needed to complete this lesson is 2 hours.

---

**What is a GIS?**

The use of geographic information systems (GIS) grew dramatically in the 1980s. It is now commonplace for business, government, and academia to use GIS for many diverse applications. Consequently, many definitions of GIS have developed. For example, turning to the Glossary of this workbook, you’ll find a GIS described as:

An organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information.

While accurate, comprehensive, and widely accepted, this definition doesn’t help the newcomer to GIS much. Its meaning will become clear as you progress through the workbook, but for the moment, consider a simpler definition:

A computer system capable of holding and using data describing places on the earth’s surface.

A number of related data layers can represent the many geographies of the real world.
spatial operations

Many computer programs, such as spreadsheets (e.g., Lotus 1-2-3), statistics packages (e.g., SAS; Minitab), or drafting packages (e.g., AutoCAD) can handle simple geographic or spatial data. Why, then, are they not usually thought of as a GIS? The generally accepted answer is that a GIS is only a GIS if it permits spatial operations on the data. As an example, consider the table below.

<table>
<thead>
<tr>
<th>Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>GIS Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>51N</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td>Zurich</td>
<td>47N</td>
<td>8E</td>
<td>25</td>
</tr>
<tr>
<td>Utrecht</td>
<td>52N</td>
<td>5E</td>
<td>40</td>
</tr>
<tr>
<td>Santa Barbara</td>
<td>34N</td>
<td>119W</td>
<td>50</td>
</tr>
<tr>
<td>Orono</td>
<td>45N</td>
<td>69W</td>
<td>30</td>
</tr>
<tr>
<td>Buffalo</td>
<td>42N</td>
<td>78W</td>
<td>30</td>
</tr>
</tbody>
</table>

This table shows the approximate number of people working with certain aspects of GIS in selected centers of activity.

aspatial queries

Asking “What’s the average number of people working with GIS in each location?” is an aspatial query—the answer doesn’t require the stored value of latitude and longitude; nor does it describe where the places are in relation to each other.

spatial queries

“How many people work in GIS in the major centers of Western Europe?” “Which centers lie within 1,000 miles of each other?” “What’s the shortest route passing through all these centers?” These are spatial queries that can only be answered using latitude and longitude data and other information, such as the radius of the earth. A geographic information system can readily answer such questions.

data linkage

A GIS typically links data from different sets. As an example, suppose you need to know what percentage of each country’s total food production is grown for export. You’ve located the data you need, but your total food production for each country is stored in one computer file, and the food export data is contained in a separate file. You must combine these files to solve the problem. Once the files are combined, it’s a simple process to have the computer perform the arithmetic to produce your answer.

If this seems trivial—hardly needing a GIS—consider the different ways in which data sets may need to be linked.

- **Exact Matching**
- **Nonexact Matching**
  - **Hierarchical Matching**
  - **Fuzzy Matching**

**Exact matching** occurs when you have information in one computer file about many geographic features (e.g., counties) and additional information in another file about the same set of features. The operation to bring them together is easy, achieved by using a key common to both files—in this case, the county name. So, the record in each file with the same county name is extracted and the two are joined and stored in another file.

<table>
<thead>
<tr>
<th>COUNTY</th>
<th>POPULATION</th>
<th>AVG. HOUSING COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAKE</td>
<td>108,000</td>
<td>89,000</td>
</tr>
<tr>
<td>LINCOLN</td>
<td>45,000</td>
<td>77,000</td>
</tr>
<tr>
<td>MADISON</td>
<td>213,000</td>
<td>104,000</td>
</tr>
<tr>
<td>ORANGE</td>
<td>1,145,000</td>
<td>167,000</td>
</tr>
<tr>
<td>PENN</td>
<td>22,000</td>
<td>75,000</td>
</tr>
</tbody>
</table>

The top two tables can be joined by performing an exact match on the name of the county.

<table>
<thead>
<tr>
<th>COUNTY</th>
<th>POPULATION</th>
<th>AVG. HOUSING COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAKE</td>
<td>108,000</td>
<td>89,000</td>
</tr>
<tr>
<td>LINCOLN</td>
<td>45,000</td>
<td>77,000</td>
</tr>
<tr>
<td>MADISON</td>
<td>213,000</td>
<td>104,000</td>
</tr>
<tr>
<td>ORANGE</td>
<td>1,145,000</td>
<td>167,000</td>
</tr>
<tr>
<td>PENN</td>
<td>22,000</td>
<td>75,000</td>
</tr>
</tbody>
</table>

The top two tables can be joined by performing an exact match on the name of the county.
Some types of information, however, are collected in more detail or more frequently than other types of information. For example, finance and unemployment data covering large areas is collected frequently. On the other hand, population data is collected for small areas, but at less frequent intervals. If the smaller areas nest (i.e., fit exactly) within the larger ones, then the solution for matching these data is to use hierarchical matching. Group the small areas together until they cover the same area as the larger area, total their data, and then perform an exact match.

### Soil Types

<table>
<thead>
<tr>
<th>Tract</th>
<th>County</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Lake</td>
<td>14,000</td>
</tr>
<tr>
<td>102</td>
<td>Lake</td>
<td>20,000</td>
</tr>
<tr>
<td>103</td>
<td>Lake</td>
<td>5,000</td>
</tr>
<tr>
<td>104</td>
<td>Lake</td>
<td>20,000</td>
</tr>
<tr>
<td>105</td>
<td>Lake</td>
<td>18,000</td>
</tr>
<tr>
<td>106</td>
<td>Lake</td>
<td>22,000</td>
</tr>
<tr>
<td>107</td>
<td>Lake</td>
<td>78,000</td>
</tr>
</tbody>
</table>

The hierarchical structure illustrated in this diagram shows that Lake County is composed of several tracts. To obtain meaningful values for the county, the values for each tract must be added.

On many occasions, the boundaries of the smaller areas do not match those of the larger ones. This is especially true when dealing with environmental data. For example, crop boundaries, usually defined by field edges, rarely match the boundary between types of soil. If you want to determine the most productive soil for a particular crop, you need to overlay the two data sets and compute crop productivity for each and every soil type. In principle, this is like laying one map over another and noting the combinations of soil and crop productivity. (Lesson 8 describes this overlay process more thoroughly.)

A GIS can perform all these operations because it uses geography, or space, as the common key between the data sets. Information is linked only if it relates to the same geographic area.

Why is data linkage so important? Consider a situation where you have two data sets for the same area, such as yearly income by county and average cost of housing. Each data set might be analyzed and mapped individually. Alternatively, they can be combined to produce one valid combination. If, however, you have 20 data sets for the county, you have over one million possible combinations. Although not all combinations are meaningful (e.g., unemployment and soil type), you can answer many more questions than if the data sets are kept separate. Combining them adds value to the database. To do this, you need a GIS.
Questions a GIS can answer

So far a GIS has been described in two ways: 1) through formal definitions, and 2) through its ability to carry out spatial operations, linking data sets using location as the common key. You can, however, also distinguish a GIS by listing the types of questions it can (or should be able to) answer. For any application there are five generic questions that a sophisticated GIS can answer.

1. location What is at...?
   The first of these questions seeks to find out what exists at a particular location. A location can be described in many ways using, for example, a place name, a post or zip code, or a geographic reference, such as latitude and longitude.

2. condition Where is it?
   The second question is the converse of the first and requires spatial analysis to answer. Instead of identifying what exists at a given location, you want to find a location where certain conditions are satisfied (e.g., an unforested section of land at least 2,000 square meters in size, within 100 meters of a road, and with soils suitable for supporting buildings).

3. trends What has changed since...?
   The third question might involve both of the first two and seeks to find the differences within an area over time.

4. patterns What spatial patterns exist?
   This question is more sophisticated. You might ask this question to determine whether cancer is a major cause of death among residents near a nuclear power station. Just as important, you might want to know how many anomalies there are that don’t fit the pattern and where they are located.

5. modeling What if...?
   “What if...” questions are posed to determine what happens, for example, if a new road is added to a network, or if a toxic substance seeps into the local groundwater supply. Answering this type of question requires geographic as well as other information.

Sample GIS applications

Applications for GIS technology developed around the world. Many of the early applications in Europe built land registration systems and environmental databases. However, Britain’s largest GIS expenditure in the 1980s was for developing utility systems and creating a comprehensive topographic database for the country.

Canada developed an important forestry application to plan the volume of timber to cut, identify access to the timber, and report the results to the Provincial governments. Applications in China and Japan emphasized monitoring and modeling possible environmental changes.

In the United States, the U.S. Bureau of Census and the U.S. Geological Survey used GIS technology for their Topologically Integrated Geographic Encoding and Referencing (TIGER) project. They produced a computerized description of the U.S. transportation network—at a cost of about $170 million—to facilitate taking and reporting the 1990 census.

Today, the number and variety of applications for GIS are impressive. The amount of geographic data that has been gathered is staggering and includes volumes of satellite imagery collected from space. Local governments use GIS for planning and zoning, property assessment and land records, parcel mapping, public safety, and environmental planning. Resource managers rely on GIS for fish and wildlife planning; management of forested, agricultural, and coastal lands; and energy and mineral resource management.

GIS supports the daily activities of automated mapping and facilities management with applications for electricity, water, sewer, gas, telecommunications, and cable television utilities, using capabilities such as load management, trouble call analysis, voltage drop, basemap generation and maintenance, line system analysis, siting, network pressure and flow analysis, leak detection, and inventory. Demographers use GIS for target market analysis, facility siting, address matching and geocoding, as well as product profiles, forecasting, and planning. GIS also has an increasing role in supporting education and research in the classroom, the computer lab, the research institute, and the public library.

The most important point to note is that these diverse applications are carried out using similar software and techniques—a GIS is truly a general-purpose tool. Appendix E contains several maps illustrating a few of the many GIS applications.
Components of a GIS

Several components constitute a GIS:

- Software Tools
- Database
- Abstraction or Simplification
- The Real World

Results

The user becomes part of the GIS whenever complicated analyses, such as spatial analyses and modeling, have to be carried out. These usually require skill in selecting and using tools from the GIS toolbox and intimate knowledge of the data being used. At present and for years to come general-purpose GIS will rely on users to know what they are doing—pressing a button is not enough.

A GIS links spatial data with geographic information about a particular feature on a map. The information is stored as attributes of the graphically represented feature. For example, the centerline that represents a road on a map doesn’t tell you much about the road except its location. To find out the road’s width or pavement type, you must query the database. Using the information stored in the database, you could create a display symbolizing the roads according to the type of information that needs to be shown.

A GIS also uses the stored feature attributes to compute new information about map features; for example, to calculate the length of a particular road segment or to determine the total area of a particular soil type.

In short, a GIS doesn’t hold maps or pictures—it holds a database. The database concept is central to a GIS and is the main difference between a GIS and drafting or computer mapping systems, which can only produce good graphic output. All contemporary geographic information systems incorporate a database management system.

If you want to go beyond just making pictures, you need to know three pieces things about every feature stored in the computer: what it is, where it is, and how it relates to other features (e.g., which roads link to form a network). Database systems provide the means of storing a wide range of such information and updating it without the need to rewrite programs. In ARC/INFO, ARC handles where the features are, while the INFO component handles the feature descriptions and how each feature is related to others.

A GIS does not store a map in any conventional sense; nor does it store a particular image or view of a geographic area. Instead, a GIS stores the data from which you can draw a desired view to suit a particular purpose.

A GIS is not simply a computer system for making maps, although it can create maps at different scales, in different projections, and with different colors. A GIS is an analytical tool. The major advantage of a GIS is that it allows you to identify the spatial relationships between map features.

In short, a GIS doesn’t hold maps or pictures—it holds a database. The database concept is central to a GIS and is the main difference between a GIS and drafting or computer mapping systems, which can only produce good graphic output. All contemporary geographic information systems incorporate a database management system.

If you want to go beyond just making pictures, you need to know three pieces things about every feature stored in the computer: what it is, where it is, and how it relates to other features (e.g., which roads link to form a network). Database systems provide the means of storing a wide range of such information and updating it without the need to rewrite programs. In ARC/INFO, ARC handles where the features are, while the INFO component handles the feature descriptions and how each feature is related to others.

Similarly, a GIS gives you the ability to associate information with a feature on a map and to create new relationships that can determine the suitability of various sites for development, evaluate environmental impacts, calculate harvest volumes, identify the best location for a new facility, and so on.

**What a GIS is not**

A GIS does not simply store a map in any conventional sense; nor does it store a particular image or view of a geographic area. Instead, a GIS stores the data from which you can draw a desired view to suit a particular purpose.