Topic Series 14

GIS Organization and Basic Functions/Outputs

I. Introduction to GIS

A **GIS** is an **information system** designed to work with **data** that are **referenced to geographic coordinates**.

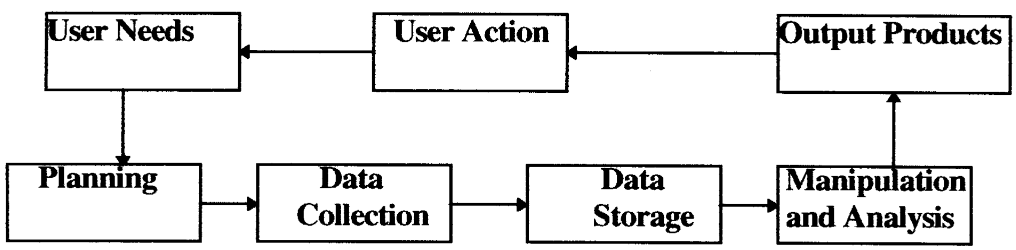
An **information system** is a set of operations that provide for:

data collection,
storage, and
analysis

in order to assist in decision making.

A **map is an information system** in that it is a collection of information, compiled and stored on paper, and it's interpretation can be used in decision making. (e.g. **Which way do I go to find a stand**? **Is the terrain too steep for my truck?**)

Figure 14-1. Flow diagram of an general information system.



An information system and therefore, a GIS, constantly evolves based on new inputs and information needs.

II. Manual and Automated GIS

Manual GIS

Manual GIS is comprised of several data elements including **maps**, **overlays**, **photographs** (aerial or ground), **reports**, **field data**.

In manual GIS, data are often compiled and analyzed by use of instruments such as **ZTS (zoom transfer scope)**, **projection systems**, **sketchmasters**, **planimeters**. The ZTS, projectors and sketchmasters are used to transfer detail from aerial photos or maps to other maps at a different scale.

Different datasets that comprise a manual GIS (such as maps, photographs) are individually referred to as **data layers or planes**.

Each **data layer** can have several levels of information. For example, a map can have **elevations** (contours and points), locations of **highways and roads**, **railroads**, **power lines**, **structures**, etc.

Each of these levels of information is called a **theme--** combination of layers of similar/defined layers; e.g. forest types, soil types, roads, streams, etc.

Some data themes used in a GIS for forest management might be: **forest type**, **soils**, **topography**, **transportation**, **administration boundaries**, **RCW colony locations**.

Tabular data (attributes) that might be related to forest type and boundaries (stands) would be: **age**, **density (BA)**, **volume**, **management history** (thinnings, other treatments).

Automated GIS

Many large corporations now utilize GIS in their administration of forest lands. This state has a system called **Mississippi Automated Resource Information System** (MARIS). The Federal government is using GIS in many agencies: **DOD**, **Dept. Interior** (USGS, FWS, National Parks), **BLM**, **EPA**, **Dept. of Agric**. (**Nat. Res. Cons. Service**, **USFS**).

615 - Forest Service GIS nation-wide. Currently in pilot phase but will eventually be nationwide, networked.

III. Elements of a GIS

1. Data Acquisition - process of **identifying and gathering the data** needed for the GIS. Gather maps, existing datasets, aerial photography, field data.

Accuracy of decisions reached through use of GIS depend on the accuracy and scale of the input data. It is not a good idea to develop stand management plans based on data developed from 1:100,000-scale maps.

2. Preprocessing - this includes **format conversion** of existing digital data and/or **encoding** (digitizing) map data to put it into digital form.

3. Data Management - this is **creation of the data base** from the encoded data. It also includes **making it available** to users and **maintaining** (updates). Here is where we make sure we have clean links between the spatial and attribute data.

4. Manipulation and Analysis - query by attributes, **recombining data sets** for analysis, **spatial operators** such as **buffering, proximity**. This is where we develop information to support decisions for forest management.

5. Product Generation - can be as simple as drawing maps on a **computer screen** to development of complex hardcopy **map products** with tabular reports.

IV. GIS Outputs

A Picture is Worth a Thousand Words (**Overhead**)

The statement is extremely important in how we deal with accurate presentation of and understanding of the information contained in or derived from GIS analysis. The visual sense is our most powerful sense and, as such, can give us good information or badly mislead us depending on the presentation of data derived from GIS. Our objective here is to discuss the types, plan and design, and tools are used to create output from GIS.

Output Types

Outputs can be characterized as either human (**hardcopy**) or machine-readable (**digital**). Here we will concentrate primarily on the former but first a note on digital data.

Digital Outputs

Except for output to **computer screen**, which is both **human and machine readable**, we generally think of digital outputs as some digital form of data and as such, can also be considered in the **storage and retrieval** GIS subsystem (**overhead of subsystems**). Today, more than ever, digital data are the currency of the internet world. Given this, a significant amount of attention has to be paid toward development of digital outputs that can be consumed by your organization or customers. Some things to consider in digital outputs include (**overhead**):

1) data format (generic or software specific)
2) coordinate / projection system
3) metadata records
4) media (CD, tape, internet published)

Hardcopy Outputs

The importance of careful planning can not be underemphasized in developing hardcopy outputs.

Things that must be considered in map design include: (**overhead**)

1) Audience (what is background knowledge of audience)
2) Venue of presentation (publication, separate map, poster)
3) Importance of elements (what to include, how to emphasize)
4) Arrangement of elements (present in meaningful manner)
5) Required information (the map must stand on its own)

There is no perfect way to present material in a map but some standard conventions for layout should be observed: **(overhead**)

1) Portray only those elements essential to the map meaning (clutter adds confusion)

2) Critical elements should be emphasized but without use of garish color or outrageous symbology (don't make an ugly map)

3) Map elements should be balanced on the page (don't leave large blank spaces nor compress important information into small areas)

4) Always include a standard set of information including (**overhead**):
    a) Title
    b) North arrow
    c) Scale (Always a scale bar and if possible a RF)
    d) Source of data (include original scale, date, etc.)
    e) Projection (coordinate reference incl. Datum)
    f) Who made it (person, organization, date produced, etc.)
    g) Legend
    h) Location reference (optional, depends on area covered)
    i) Coordinate grid (optional, depends on audience, etc.)

The following overheads are examples of different map layouts with problems. How many can you spot? **Have the class spend a minute to list then compare.** (**overhead**)

V. Hardware/Software

(overhead)

GIS hardware/software environments have been evolving at a steady and very rapid pace. I've put this discussion into the Outputs topic because after it is all said and done, the hardware/software environment is what will make or break our ability to produce effective output or distribute data for use in GIS decision processes.

Digital mapping programs were first implemented on "mainframe" computers that would easily fill this room. These were simple programs that took a long time to run and produced simple line maps (i.e. contours) or raster maps based on line printer technology. That was the story 30 years ago. Today we are in a world where everyone is connected and everyone has more power than the old mainframes on their desktops. Lets first take a brief look at the evolution of the hardware environment that drives GIS.

Mainframes / Minicomputers

The big buzz words of 20-40 years ago were mainframes and minicomputers. People with heavy duty processing needs would set up jobs on **mainframes** by **printing card decks** or typing on a **text terminal** and submit them to run (sometimes took hours or even days). These tasks had to compete with all other tasks in a queue. **Everyone had to use the same CPU** and related resources.

When **minicomputers** came along, researchers and departments in large organizations obtained a little bit of **independence** from the big centralized processors. Sometimes tasks ran faster (not always) because there were fewer tasks competing for the CPU and peripheral resources.

PC's

The biggest boom to getting computers and applications to the masses was the advent of the **personal computer (PC)**. The early PC's (Apple, IBM, etc.) were **memory and disc poor** so primarily ran small software applications and databases. **Early GIS** applications on these platforms were often **raster-based** due to ease of management of the data structure and **computational simplicity** of handling rasters. Some PC's were set up to **communicate with mainframes or mini's** to off-load big tasks to the faster CPU's.

Today's PC is a real workhorse machine that in many respects, rivals the fastest large computers of 10 years ago. The **primary drawback** to PC technology until recently was the **inability to handle multiple tasks** at the same time (limitations of DOS in memory and resource use). New windows-based operating systems (such as Windows 2000/NT) provide solutions to these problems. These sorts of limitations were the impetus for the boom of **workstation technology**. Now, however, PC's are capable of running high-end operating systems under windowing environments and can also be used as servers for other PC's on **integrated networks**. So the term "PC" may be short lived in the new millennium.

Price of a high-end PC with large, high-resolution color screen, floppy disc, CD-ROM, and large hard disc and sometimes a printer is generally around $2000-3000.

Workstations

A typical workstation might have 500+ **Mbytes** of memory, 80**+ Gbytes** of disc storage, large **20" full color monitor** and be connected to a network supporting multiple workstations. The rule more than the exception is distributed networks where workstations share peripherals and resources. **SoRSC**/**SITL** use such a network. It supports 19 workstations, 400+ Gbytes of storage, and networked printers and plotters.

Low-end workstations run about as much as a high-end PC ($2000-3,000). The performance difference between PC's and workstations is greatly reduced although high-end workstations are still far superior in multitasking and raw compute power.

Networks

Most large organizations maintain very sophisticated computer networks consisting of many PC's, workstations and file servers. File servers provide mass disk storage and often serve as a central point for software access to multiple machines in the net. Workstations and PC's can be set up to directly share files or share them through the file server. Connections to the outside world (a very important thing) are provided by high-speed modems or through direct lines that are secured for computer data traffic only.

The Internet was originally set up for government research but is now the standard for communications. This system is used to have computers sharing resources around the world. We will look at a few Internet world wide web sites in lab.

Input/Output

With current networking technologies, these are not always tied to one machine. There are printers and plotters on the market that can be set up with **their own network address** and are not connected to any machine. Most workstations and servers can be set up to share other devices like tape drives for loading or backing up data. The more labor intensive devices such as scanners and digitizers are still tied to specific machines and as such, are usually more effectively used on PC's rather than high-end workstations or servers.

Plotted products (the glitz in GIS) have gone from simple line printer output to color-ink-jet and laser printers and plotters

SHOW EXAMPLES of Plots.

Large format plotters run around $5,000 as opposed to over $50,000 of the former generation of color electrostatic devices. They are less expensive to run and maintain.