**Geographic Data**

***Geographic data*** originate from actual locations and physical characteristics of features on/or near the surface of the Earth.

**Geography and Geographic Data**

***Geography*** is the study of the Earth’s surface and climate, and is the founding science to a ***GIS***. ***Geography*** provides information about the Earth and distinguishes how features upon the Earth correlate with one another.

**Georeferencing**

**Georeferencing** is the process of assigning real-world coordinates to each pixel of the raster. Many times these coordinates are obtained by doing field surveys - collecting coordinates with a ***GPS*** device for few easily identifiable features in the image or map.

**Geoprocessing**

***Geoprocessing*** is a fundamental process of creating a derived set of geographic data from various existing data sets using operations such as feature overlay and data conversion.

**Geographical Features**

***Geographical features*** are the components of the Earth. There are two types of geographical features, namely 1) natural geographical features 2) artificial geographical features. Natural geographical features include but are not limited to landforms and ecosystems.

**GIS Attribute Table**

It is an ***information*** about a geographic feature in a GIS, usually stored in a ***table*** and linked to the feature by a unique identifier. For example, attributes of a river might include its name, length, and sediment load at a gauging station. There are various attribute field data types to handle the multitude of data, such as 1)numerical data (integer, real), 2)text, date, 3)Boolean (Y/N), 4)BLOB (Binary Large Object) such as any type of image, video, sound, geometry.

**The Structure of Geographic Data**

***Geographic data*** are two type, ***Raster*** and ***Vector***.

* + - 1. ***Raster*** data are digital images represented by a grid of valued pixels, or cells. Raster data structures characterize continuous data (such as imagery). Raster data provide data as a pixel grid or a cell. A raster image can vary in file format, color representation, resolution (size of pixels/number of pixels per set area).
			2. ***Vector*** is A coordinate-based data model that represents geographic features as ***points***, ***lines***, and ***polygons***. Each ***point*** feature is represented as a single coordinate pair, while ***line*** and ***polygon*** features are represented as ordered lists of vertices. ***Attributes*** are associated with each vector feature, as opposed to a raster data model, which associates attributes with grid cells. Vector data structures characterize discrete data (such as roads, pipelines and topographic features). Features are represented as geometric shapes defined through single or grouped coordinates on a set grid.

**Geographic data source**

***Geographic data*** come from a variety of sources, such as digitized maps, aerial photography, GPS, and field data. Geographic information systems utilize two primary data models to manipulate and structure geographic data: the ***raster data model*** and the ***vector data model***.

**GIS Data Structure Design**

**Purposes of GIS Data Structure**

1. Efficiently access and retrieve data from database;
2. Minimize the required storage space; and
3. Support the desired GIS operations.

**Concept of Vector**

1. The basic elements of vector data model are ***points***;
2. The directional linear connection between two points;
3. The root of vector data model lies in cartography;
4. Complex spatial objects are created by connecting primitive objects;

**Primitive Objects**

**Points**

* Represent geographic features too small to be depicted as lines or areas.
* A ***point*** is recorded as a pair of x, y coordinates.
* ***Node*** is a topological point at which two or more arcs connect each other.

**Lines (Arc)**

* Represent geographic features too narrow to depict as areas.
* Lines are recorded as a series of ordered x,y coordinates.

**Polygons (areas)**

* Represent homogeneous geographic features. The term ***polygon***, meaning‘ many-sided figure’.
* ***Polygons*** are recorded as a series of x,y coordinates defining line segments that enclose a polygon.

**Advantages of Vector**

* Data can be represented at its original resolution and form without generalization.
* Graphic output is usually more aesthetically pleasing (traditional cartographic representation);
* Since most data, e.g. hard copy maps, is in vector form no data conversion is required.
* Accurate geographic location of data is maintained.

**Disadvantages of Vector**

* The location of each vertex needs to be stored explicitly.
* Algorithms for manipulative and analysis functions are complex and may be processing intensive.
* Continuous data, such as elevation data, is not effectively represented in vector form.
* Spatial analysis and filtering within polygons is impossible.

**Advantages of Raster**

* The geographic location of each cell is implied by its position in the cell matrix.
* is usually easy to program and quick to perform.
* Grid-cell systems are very compatible with raster-based output devices, e.g. plotters, graphic terminals.

**Disadvantages of Raster**

* The cell size determines the resolution at which the data is represented.
* It is especially difficult to represent linear features depending on the cell resolution.
* Raster maps inherently reflect only one attribute or characteristic for an area.
* Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

