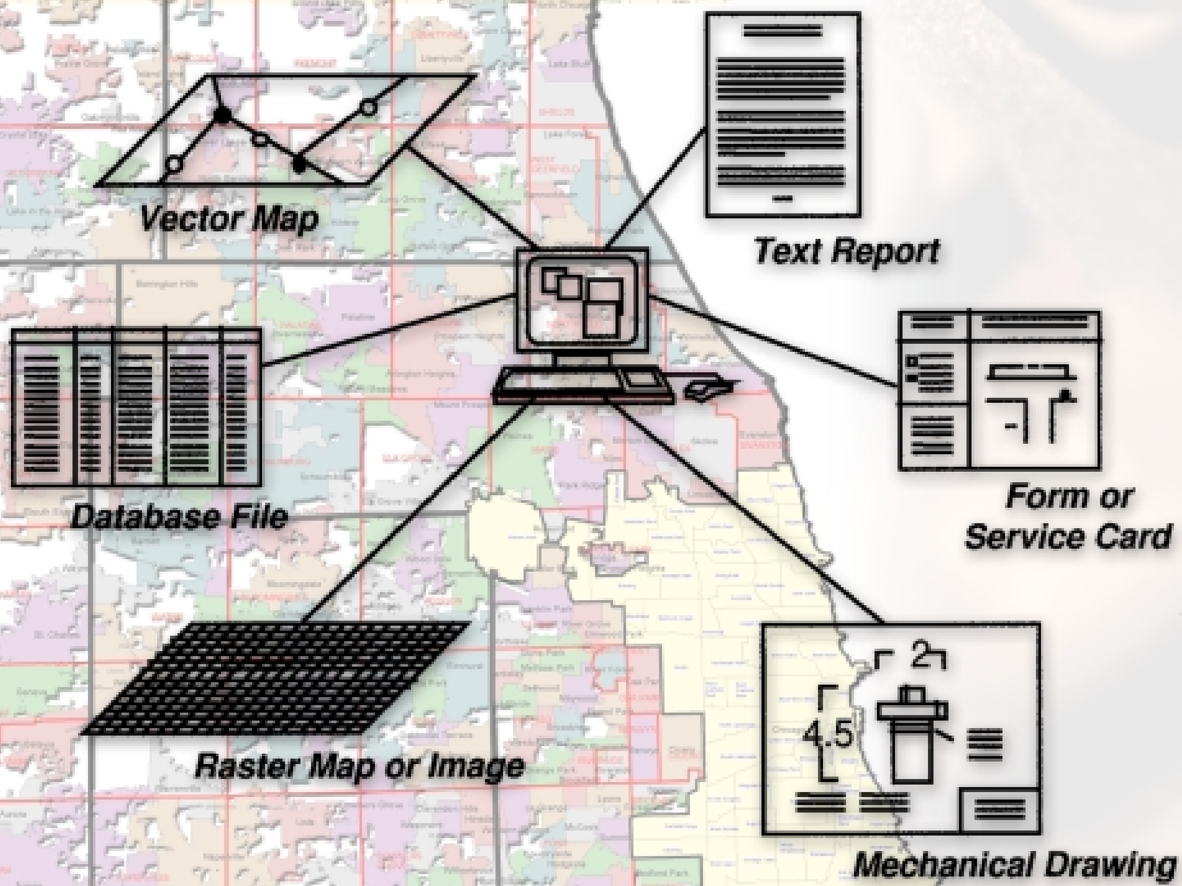
# GIS Database Concepts

## A Tutorial



# GIS DATABASE CONCEPTS

## A Tutorial

URISA
*QUICK STUDY SERIES*

**Prepared with the support of**
**PlanGraphics, Inc.**
**April, 1998**

**The Urban and Regional Information Systems Association**
**1460 Renaissance Drive, Suite 305**
**Park Ridge, IL 60068**

Printed in the United States

ISBN #: 0-916848-09-4

# Table of Contents



## List of Figures

# 1.

## Types and Formats of GIS Data

GIS data storage techniques actually format graphic and nongraphic data in a variety of different formats to facilitate processing and retrieval. This sub-section briefly explains major concepts about GIS data structures.

### Database Layering Concept

It is convenient to describe a GIS database as a series of map layers that are geographically referenced and registered. While specific approaches for physically designing and storing GIS data vary among software packages, most provide a means to organize data by layers each of which contains a set of map information which is logically related. The GIS layering concept is illustrated in Figure 1.

### Raster and Vector Data

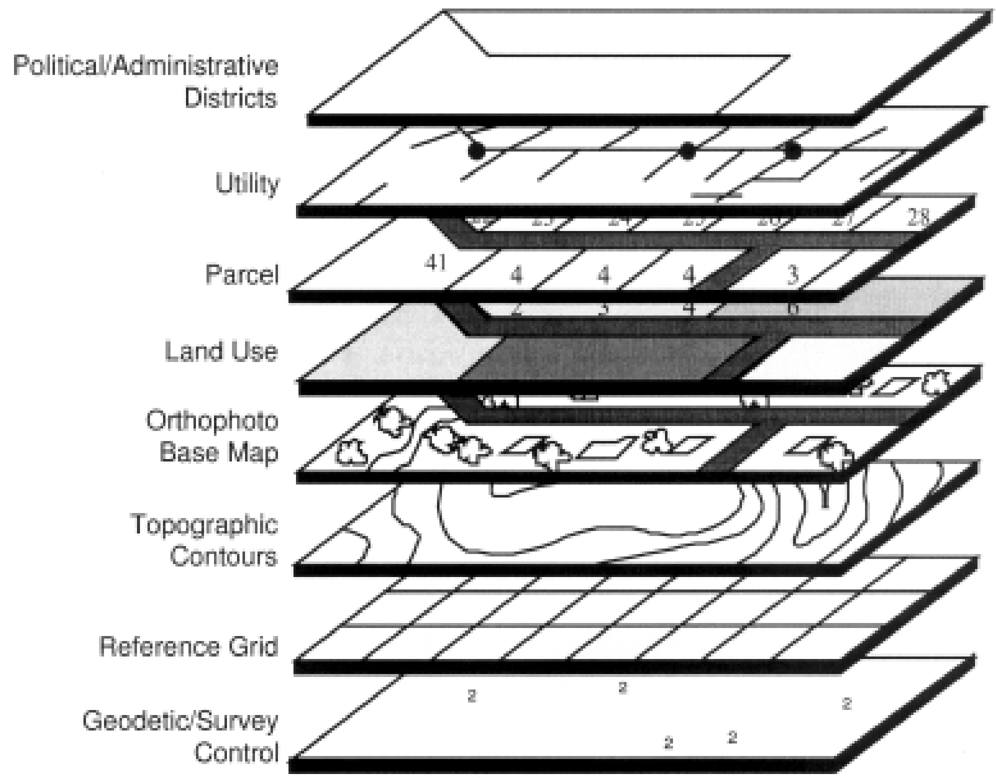
Spatial features in a GIS database are stored in either vector or raster form. GIS data structures adhering to a "vector" format store the position of map features as sequences of x, y (and sometimes z) coordinates. A vector format represents the location and shape of features and boundaries precisely. Only the accuracy and scale of the map compilation process; the resolution of input devices; and the skill of the operator inputting data limit the precision.

In contrast, the "raster" or "grid-based" format (shown in Figure 2) generalizes map features as cells or pixels in a grid matrix. The fineness of the grid or, in other words, the size of the cells in the grid matrix, will determine the level of detail at which map features are represented. There are advantages to the raster format for storing and processing some types of data in GIS. Four principal cases where the raster format is very useful are explained below:

- Map scanning as a step in map automation
- Storage and manipulation of images
- Analysis of grid-based maps
- Map display and plotting.

Some data that may be important for a GIS does not require "intelligence" at the map feature level and may be stored in raster form. This is appropriate for data that is simply an image such as an aerial photograph or orthophotograph that may be overlaid with vector map layers. Figure 3 is an example of a digital ortho image. This type of product is a "map" in the sense that it represents surface features for a specific geographic area, but it is different from vector GIS

**Figure 1: Database Layering Concept**




databases, because individual map features are not uniquely defined or linked with attribute databases. This type of image can be encoded in raster form to create a digital image that can become part of the GIS database. Many GIS software vendors have developed capabilities to overlay digital images of this type with digital layers in vector format.

## Types of Map Features

Map features can be represented in the ways described below. These map feature formats, illustrated in Figure 4, govern how the features are stored and how they are manipulated in a GIS:

- **Point Features**
  These are features or geographically-defined occurrences whose location can be represented by a single x, y or x, y, z location. Points have no linear or area dimensions but simply define the location of a physical feature (e.g., control point monument, sign, utility pole) or an occurrence (e.g., accident).
- **Line Features**
  Lines represent features that have a linear extent in the GIS database but no area dimensions. Centerlines of roads, water mains, and sewer mains are examples of line features.

**Figure 2: Vector-Raster Relationship**

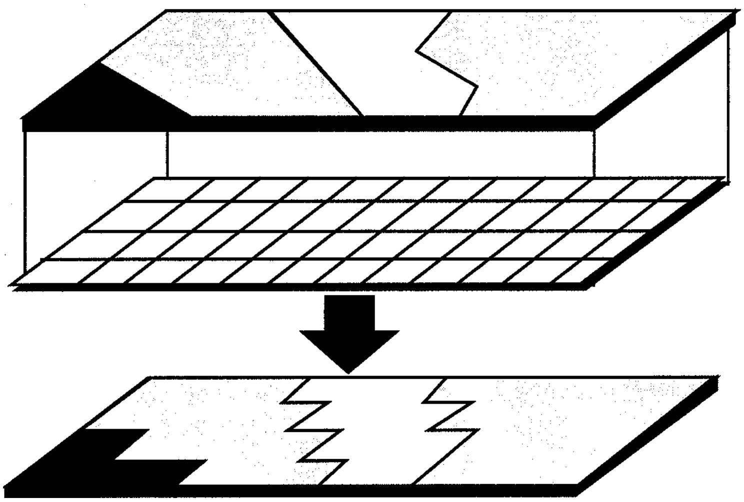



**Figure 3: Orthoimage**

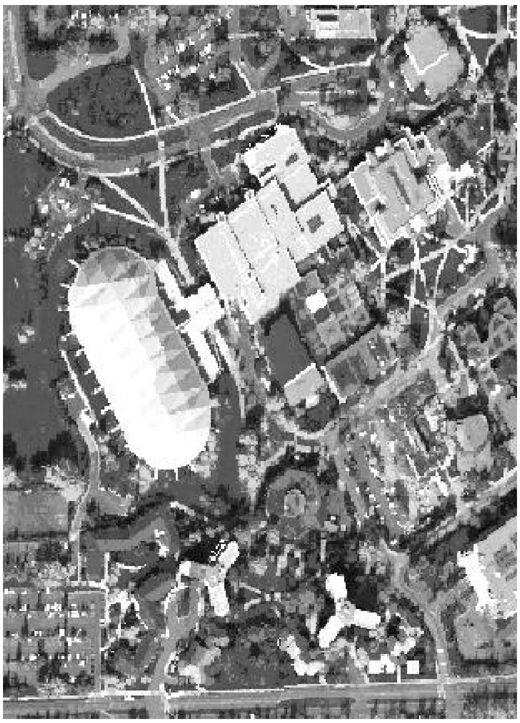


**Figure 4: Types of Map Features**

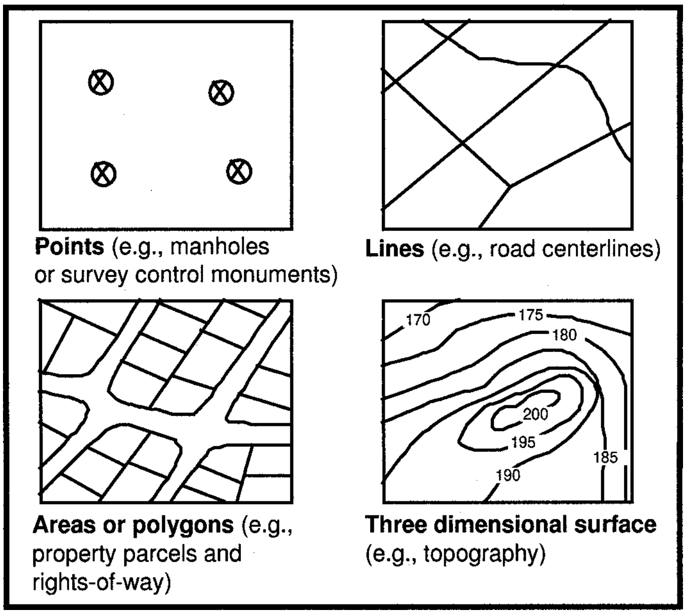
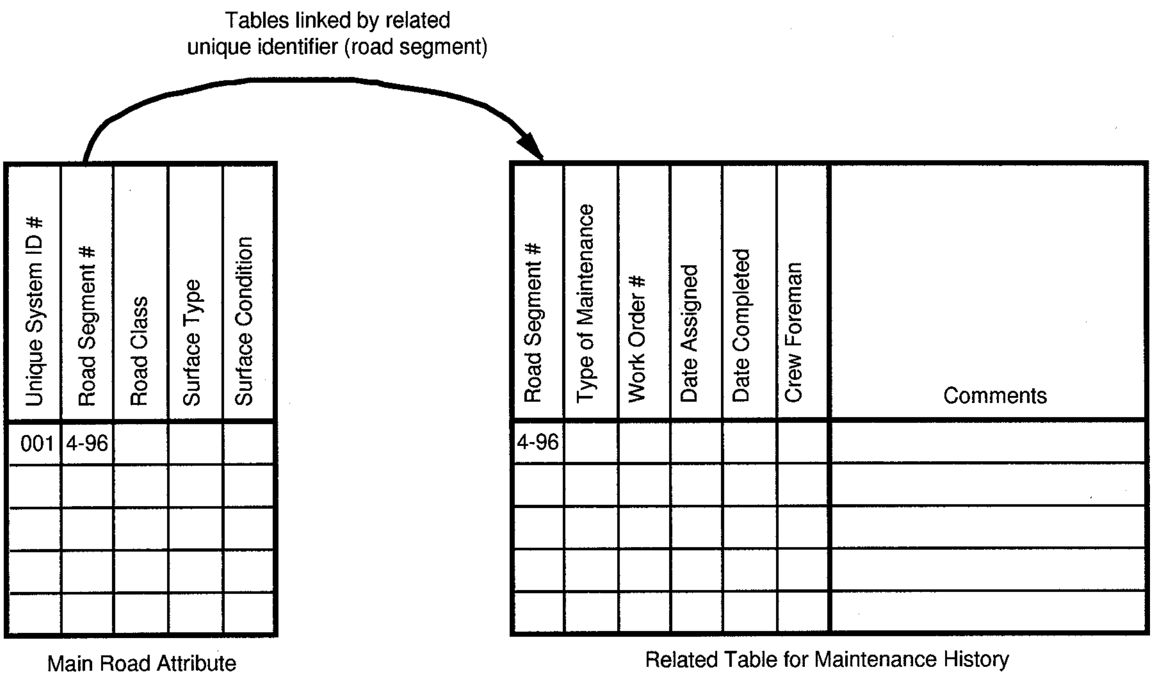



**Figure 5: Relational Database Tables**



- **Area Features**
  Area features, also called "polygons," have a defined two-dimensional extent and are delimited by boundary lines that encompass the area. Typical area features are maintenance districts and soil types.
- **Linear Networks**
  While map features in a GIS may often be stored discretely as points, lines, and polygons, points and lines may be combined to form "networks." A network is an interconnected system of lines (and often point features) which define a pathway for flow (e.g., traffic on a road network). The linear features define road segments, and the point features are locations of changes where flow is being influenced or monitored (e.g., bridges, intersections).
- **Three-Dimensional Surfaces**
  Some geographic information is best suited for representation in three-dimensional form covering an area. The most frequent example is surface terrain often represented by contour lines that have an elevation value. This concept can be applied to other spatially continuous data as well. For instance, population density or income levels could be mapped as a "third dimension" to support demographic analysis or water consumption statistics.

## Tabular Attribute Data

GIS software vendors have employed a variety of techniques to store tabular attribute data and to link these data with map features. All approaches, however, use the concept of a database management system (DBMS) to allow the user to define the specific data element types and formats. A DBMS allows a user to describe the particular contents of a database and the formats of data elements (e.g., integer, decimal, date, character).

The relational DBMS model for storing attributes is, by far, the most popular approach in the GIS software industry. The relational model is based on the storage of attributes as two-dimensional tables. As illustrated in Figure 5, multiple tables can be linked or related based on a common and unique identifying attribute. Columns represent what traditionally have been referred to as data "elements," and rows represent data "records."

In Figure 5, "road segment number" is a key field that uniquely identifies records. Map features in GISs are "tagged" with a unique number (usually an internally generated sequential number) that is used to link the feature to a database as depicted in Figure 6. The capabilities of GISs to manipulate both map features and tabular data allow virtually any automated tabular files to be linked for GIS analysis as long as there are unique identifying fields that can be matched to map features.

## Digital Terrain Models

Digital terrain models (DTMs) are computer files that store elevation information in a way that provides a model of the Earth's terrain. Figure 7 shows that elevation information may be initially gathered as spot elevations and breaklines that represent important breaks in the slope of the land (e.g., streams, ridgetops, etc.). A digital terrain model stores detailed elevation data in two main formats—1) triangular irregular network (TIN) format consisting of a mesh of triangles with defined elevations at their vertices, or b) grid format in which an elevation value is assigned to each grid cell. These DTMs may be used with appropriate software to perform terrain analysis and to generate a variety of products.

**Figure 6: Linkage of Tabular Attributes to Map Feature**

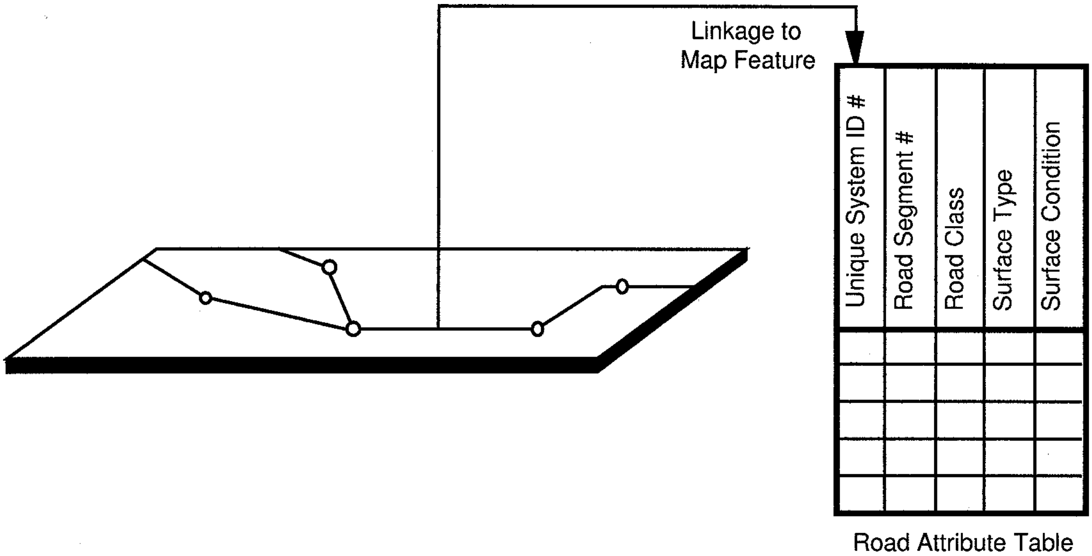
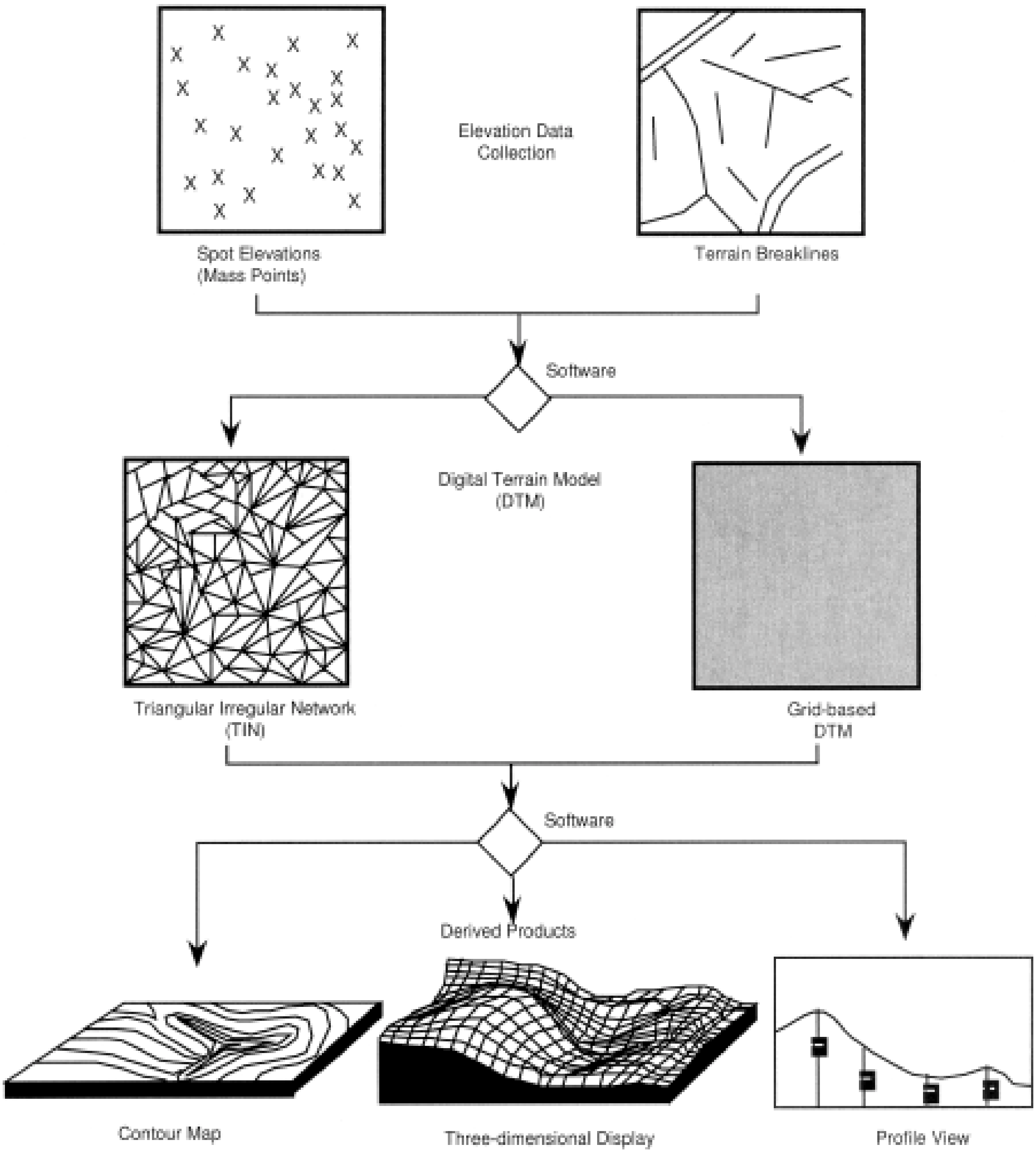


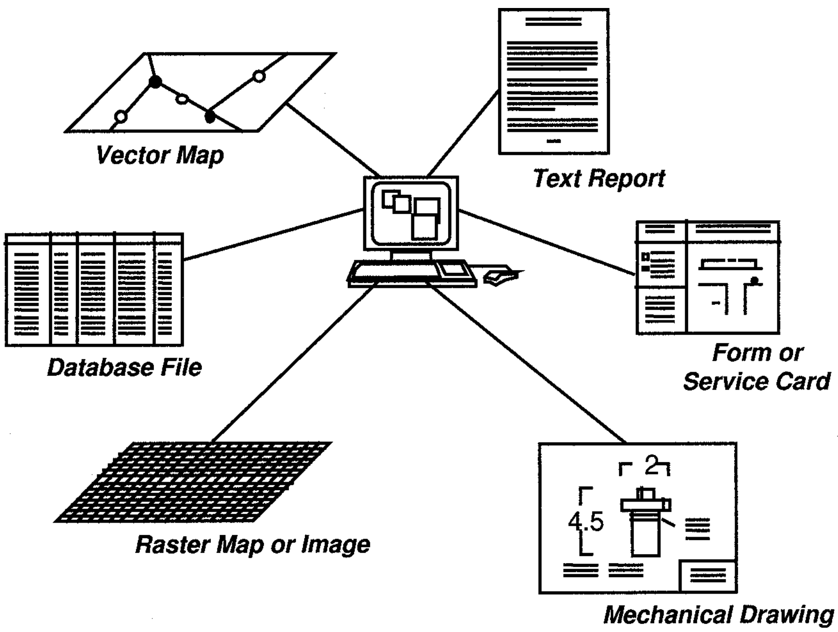
**Figure 7: Digital Terrain Models**



# 2. Multimedia and GIS

Some GIS applications may effectively apply some of the concepts of "multimedia access," which is a recently coined term in the data processing community. Multimedia describes an information system environment that provides access to geographic information that exists in a variety of forms. In GISs, multimedia tools can provide a way to integrate GIS data of different types and formats (other than maps and traditional tabular databases) in a manner that is transparent to the user. Some examples of other data types include engineering or mechanical drawings, architectural plans, aerial photographs or orthophotographs, service or maintenance cards, field sketches, site photographs, or textual documents (see Figure 8).

**Figure 8: Multimedia Access**



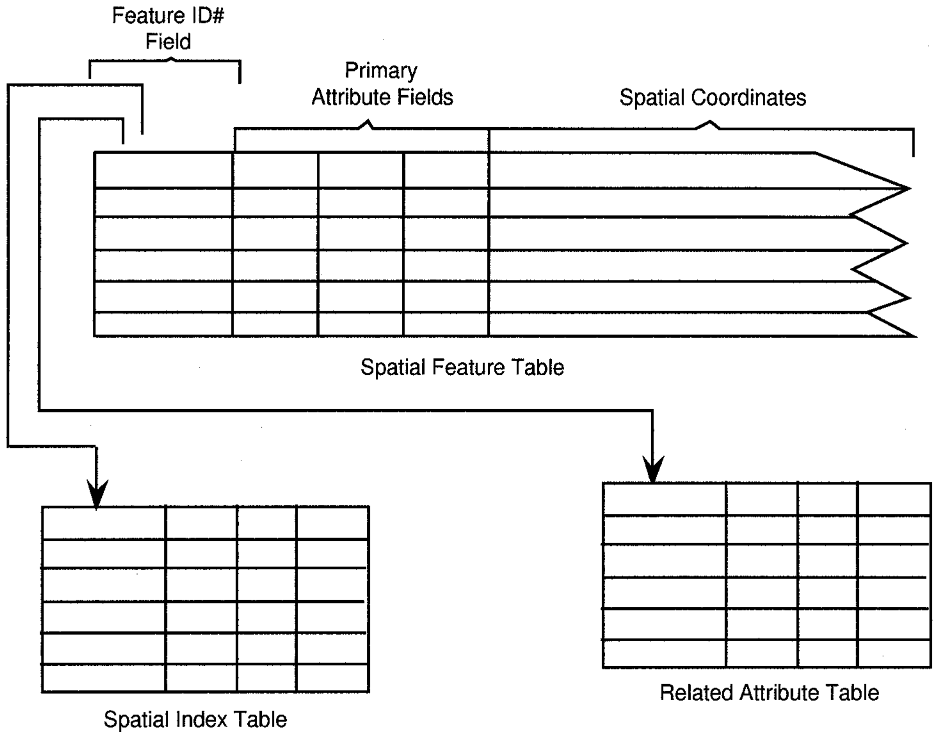
# 3.

## Spatial Data Repositories

Many GIS software companies and relational database software vendors have developed software products, generally referred to as a "spatial database repository" that allow the storage and management of all spatial data (map features, attributes, and even raster data) in a relational database. This architecture is a move away from the traditional concept of a GIS database in which spatial data is stored in a proprietary graphic format and tabular attributes are stored in a separate database table. The general concept for this spatial data repository is presented in Figure 9.

The major advantages of these spatial data repository products is that they are suited for performance with large databases and they can use the robust data administration tools offered with the relational database package. They also provide better opportunities to integrate spatial data with other information that organizations may store in relational databases. These packages provide a more open architecture for managing GIS data and allow access with multiple client software for data viewing, update and analysis.

**Figure 9: General Concept of RDBMS-based Spatial Data Repository Architecture**



# 4.

## Scale, Accuracy, and Data Quality

The concepts, discussed below, relate to the overall format and quality of the GIS database. Each of these factors should be examined in development procedures for creating and maintaining the GIS database.

### Coordinate Systems

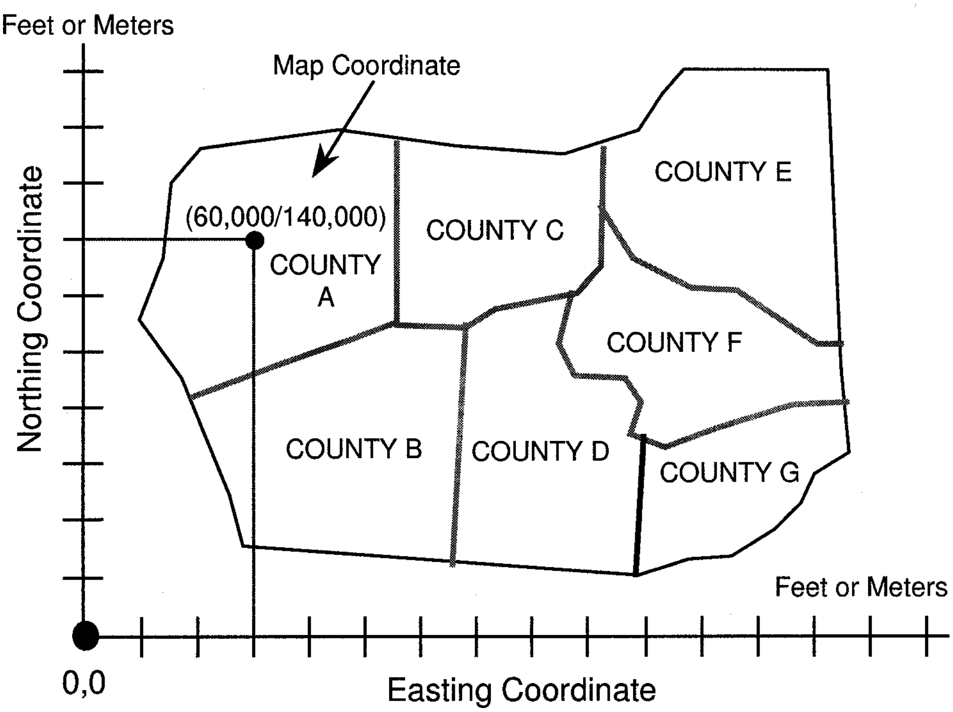
Which mathematically "projects" the curved Earth surface onto a flat surface. Some distortion always results but map projections seek to minimize that distortion. Most GISs store map locations in planar coordinate systems in which distances may be measured in regular grid east and west. The State Plane Coordinate System (SPCS) used for many GISs for the United States defines an origin for a specific geographic region or "State Plane Zone" in which x,y coordinates are defined from the origin of the zone. Figure 10 illustrates the State Plane Coordinate concept.

### Spatial Accuracy

One of the most critical factors in GIS database design and compilation is spatial accuracy. Spatial accuracy specifies how well the position of features or boundaries, as plotted on a map, conforms to their actual position on the ground. Accuracy is normally stated as maximum error between mapped position and actual ground position, both measured relative to a standard grid.

Accuracy is an important concern in building GIS databases because it has a direct and significant impact on the cost of the effort and the applications for which the database can be used. "Absolute accuracy" describes errors in comparison to the Earth's geodetically described reference system. The geodetic reference is based on a very precisely defined mathematical model of the shape of the Earth. A geographic coordinate scheme, such as Latitude-Longitude, or a planar coordinate system such as Universal Transverse Mercator (UTM), or State Plane, allows map positions to be determined horizontally (x, y coordinates) and vertically (z coordinate or elevation). These coordinate systems are "absolute," because positions can be located uniquely relative to all other positions on the Earth.

**Figure 10: Concept of State Plane Coordinate System**




## Scale

Map scale describes the relationship between mapped size and actual size. It is expressed as a relationship between linear distances on the map and corresponding ground distances. Two methods of notating scale are commonly used:

- Inch-Foot Equivalent—The scale relationship is expressed as "1 inch = x feet" where the map distance of 1 inch is compared to its corresponding ground distance.
- Representative Fraction (RF)—This is a pure fraction that represents the ratio of map distance to ground distance without specifying any measurement unit. The inch-foot equivalent of 1" = 100' is represented in RF form as 1:1,200 or 1/1,200.

Large-scale maps cover small areas, but can include a higher level of detail than small-scale maps which depict larger areas at lower detail. There are no precise definitions of large- or small-scale, but for most GIS users, the following general scale categories apply:

- Large-scale:
  1" = 50' to 1" = 200' (1:240 to 1:1,200)
- Medium-scale:
  1" = 100' to 1" = 1000' (1:1,200 to 1:12,000)
- Small-scale:
  1" = 1000' to 1" = 5,000' (1:6,000 to 1:60,000)
- Very Small-scale:
  1" = 5000' and smaller

A number of factors, discussed below, contribute to the overall quality of the GIS database. Each of these factors should be examined in the development of procedures for initial development and ongoing maintenance of the GIS database.

### Completeness

Completeness relates to both map features and corresponding tabular attributes. In the case of a map "layer" of roads, two questions relating to completeness can be posed:

1. Are all road segments that existed at the time of mapping depicted on the map?
2. Are any features that are not road segments erroneously depicted on the map as roads?

The level of completeness that is appropriate for a geographic database may vary from project to project but in most cases, a completeness level approaching 100% is important for effective use of the GIS. Also, the level of completeness should be documented as one element of metadata.

### Attribute Accuracy

This data quality concern relates to the integrity (and frequency of errors) in the values of attributes entered in the GIS database. The potential for errors in number or text fields during data entry can be reduced by establishing quality control procedures including checks against valid lists and ranges (attribute domains).

### Coding and Classification Integrity

Integrity also relates to correctness of coding or classification. In the case of a land use map, we may ask whether all parcels coded as "high density residential" actually fall into this category. Procedures for creating and updating GIS databases are subject to coding inconsistencies that may result from errors in initial interpretation, map compilation, or data entry. Systematic procedures for database creation and quality control can be applied to reduce these types of errors to a tolerable level.

### Currency

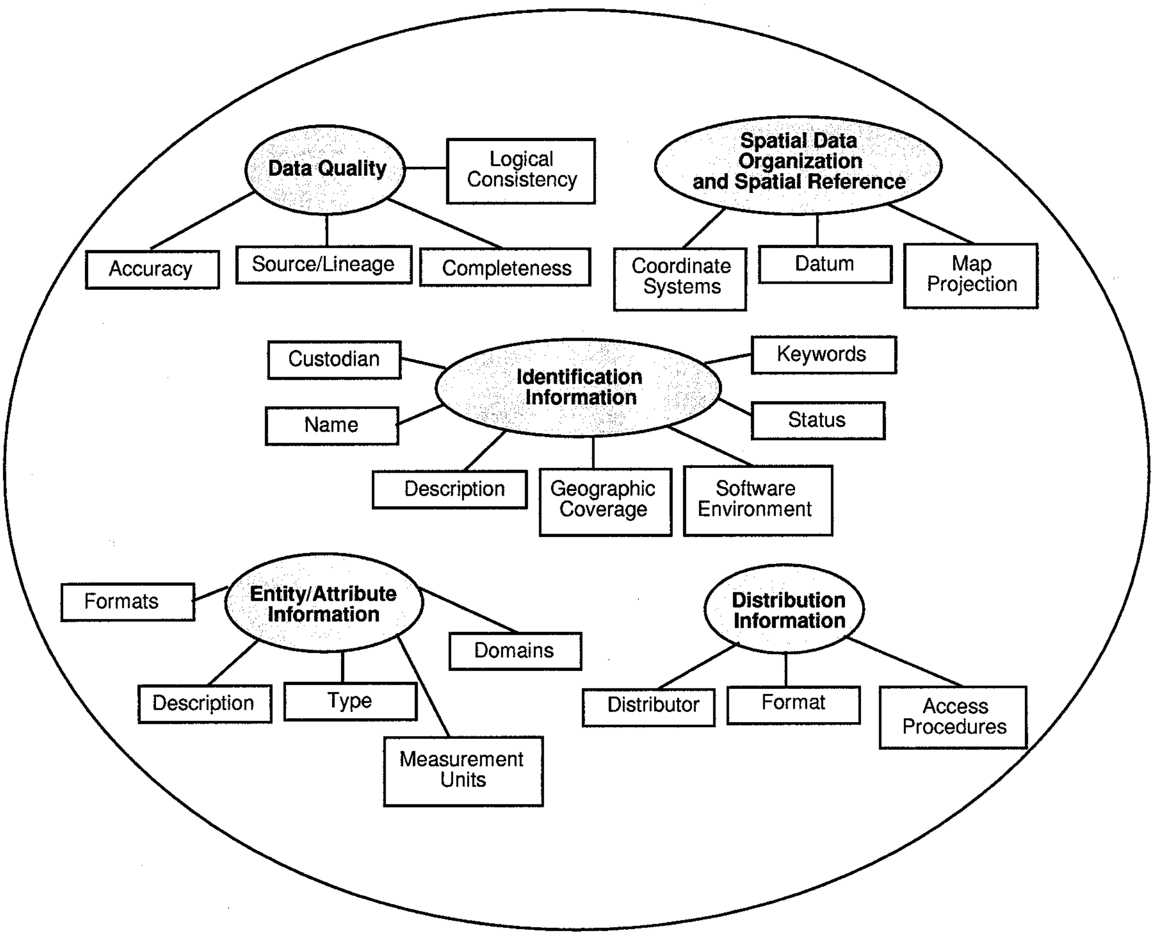
For many components of a GIS database, currency is a critical issue that directly influences the applications for which the data may be used. Some database components change frequently and demand routine update if the data is to be applied effectively. It is a useful approach to keep records as part of the GIS database on the dates and times of updates. In some cases, it may be advisable to include an attribute field for individual map features indicating status or date of change. Currency is an issue that must be addressed as procedures are devised for update of the GIS database "layers."

# 5. Geospatial Metadata

Many organizations that manage geographic databases find it useful to capture and update “metadata” about the geographic data. Metadata includes information about the content, format, quality, availability, and other characteristics of a GIS database. Metadata can generally be defined as “data about data”. Metadata can help answer such questions from database administrators and users as “what data is available?”, “what are its characteristics and quality?”, “how do I access it?”, “will it meet my needs?”. Metadata is designed to facilitate such database activities as a) on-going update, b) browsing and discovery, c) distribution, d) proper use in applications. A metadatabase can be thought of as a robust index to a geographic database which makes the GIS data easier to manage and use.

The Federal Geographic Data Committee (FGDC) has developed a standard for storing metadata called the *Content Standard for GeoSpatial Metadata*. This standard lays out major categories of metadata and specific data items that may be stored in a metadatabase and it provides a model for user organizations to adopt. Figure 11 provides an overview of the types of information that is stored in a GIS metadatabase.

**Figure 11: Overview of the FGDC Metadata Content Standard**



# 6.

## GIS Database Development and Conversion Approaches

Information can be entered into a GIS database using a variety of techniques that depend on the format, content, and condition of source material and on accuracy and quality requirements. This section provides a description of issues surrounding the preparation of records for automation and various approaches for building a GIS database from existing hard copy and automated data.

### Data Preparation

The source documents used in the creation of a GIS database include maps, paper records, card files, automated databases, or aerial photography. Usually some preparation of source documents is required before they are usable for conversion. Source preparation may range from organizing and assembling maps, to a detailed review and revision of their content. Each data layer source material needs to be evaluated to determine the amount of preparation, or "scrubbing," and whether the process should be undertaken by DEC or by a contractor.

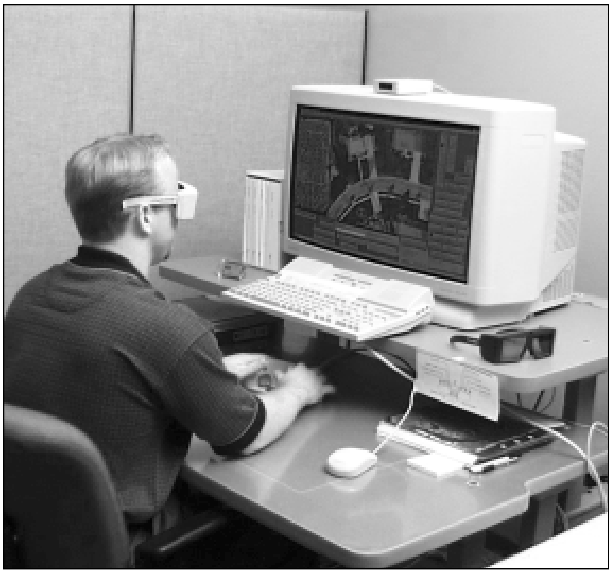
Document scrubbing can be a time-consuming process, but it allows resolution of inconsistencies, provides a source which can more rapidly be digitized or scanned, and ultimately, results in better quality information in the GIS database. Besides redrafting, document preparation tasks may include photographic enlargement, reduction and removal of extraneous marks and notes, or clarifying handwritten notes. Field verification may be required in some cases to verify map features and locations. Depending on their condition, source materials may be used as input into the geographic information system without redrafting.

### Data Automation

#### Photogrammetric Compilation

The primary source used in the process of photogrammetric compilation is aerial photography from aerial cameras. Generally, the process involves using specialized equipment (a stereoplotter) to project overlapping aerial photos so that a viewer can see a three-dimensional picture of the terrain, known as a photogrammetric model. In the photogrammetric model, the location of survey control points can be identified, and the mathematical relationship between the control points, and all other features visible on the photos, can be reestablished. Using the three-dimensional view and the accurate locational relationships between features, the photogrammetrist is able to compile a map by

**Figure 12: Photogrammetric Workstation**



*An operator sitting at and using a photogrammetric workstation (analytical stereoplotter). Photo courtesy of Analytical Surveys, Inc., Colorado Springs, CO.*

digitizing the locations of roads, buildings, and other features on the stereoplotter. The information digitized from the photographs can then be checked and edited, annotation can be added, and the result plotted as a planimetric base map.

The current technological trend in photogrammetry is toward a greater use of digital procedures for map compilation. This trend, referred to in the industry as "soft copy photogrammetry" is resulting in less reliance as photographic film and manual procedures. New photogrammetric workstations will accept scanned photographic images or output from digital cameras and new software will cut down on manual tasks needed for map compilation. Figure 12 shows a state-of-the-art photogrammetric workstation.

**Trace Digitizing**

A digitizing workstation with a digitizing tablet and cursor is typically used to trace digitize. Both the tablet and cursor are connected to a computer that controls their functions. Most digitizing tablets come in standard sizes that relate to engineering drawing sizes ("A" through "E," and larger). Digitizing involves tracing features on a source map, taped to the digitizing tablet, with a precise cross hair in the digitizing cursor and instructing the computer to accept the location and type of feature. The person performing the digitizing may separate features into map layers, or attach an attribute to identify the feature. Figure 13 shows an individual performing trace digitizing.

**Coordinate Geometry (COGO) Entry**

COGO is a technique for entering boundary information to a GIS database by key entering distances, bearings, and curve calculations from field surveys. When distances and bearings define the position of boundary lines based on a coordinate grid, such as the State Plane Coordinate System, GIS software can use this information to create a graphic representation of the lines. This technique is most

**Figure 13: Trace Digitizing**

*An operator hand digitizing a map at a large format digitizing tablet. Photo courtesy of Analytical Surveys, Inc., Colorado Springs, CO.*

commonly used for entering real property boundaries, although some detailed locations for features in the field may be collected in automated form with positions "downloaded" to the GIS from field computers.

**Map Scanning**

Optical scanning systems automatically capture map features, text, and symbols as individual cells, or pixels, and produce an automated product in raster format as described earlier in this section. Scanning outputs files in raster form, usually in one of several compressed formats to save storage space (e.g., TIFF 4, JPEG). Most scanning systems provide software to convert raster data to a vector format differentiating point, line, and area features. Scanning systems and software is becoming more sophisticated with some abilities to interpret symbols and text, and store this information in databases. However, such "intelligence" in scanning systems is still very limited. Most scanners are still very sensitive to variations in line types, widths, and stray marks or creases on maps, and post-scanning clean-up and editing is normally a very time-consuming process. Creating an intelligent GIS database from a scanned map will require vectorizing the raster data and manual time for entering attribute data from a scanned annotation.

**Document Scanning**

Smaller-format scanners can also be used to create raster files of documents such as permit forms, service cards, site photographs, etc. These documents can be indexed in a relational database by number, type, date, engineering drawings, etc., and queried and displayed by users. GIS applications can be built which allows users to interactively point to (e.g., tax parcel) and retrieve for display a scanned document (e.g., a deed). They can also be referenced to map locations or features in the GIS graphic database. As in the case of map scanning, scanned documents sometimes require clean-up and image enhancement after scanning.

**Heads-up Digitizing**

Heads-up digitizing capabilities work with the scanning and editing procedures described above to provide a semi-automated environment to efficiently convert hard copy maps to vector format suitable for the GIS. After a map has been scanned, the raster image, with the scanned line work and annotation, is displayed on a workstation monitor. The operator uses a mouse to invoke batch routines and/or interactively edit and clean the raster image to remove stray marks or line gaps picked up in the scanning process. The operator may then use heads-up digitizing tools to perform vector conversion and to enter annotation and attributes. Tools may allow the user to select individual raster features for vector conversion, invoke automatic line-following and thinning vector conversion, direct keying of attribute data, and other tools to speed the process of vector conversion. Figure 14 depicts an operator performing heads-up digitizing.

**Figure 14: Heads-up Digitizing**

*An operator doing heads up digitizing. Photo courtesy of Analytical Surveys, Inc., Colorado Springs, CO.*

## Tabular Data Entry

Some of the tabular attribute data that is normally in a GIS database exists on maps as annotation and or can be found in paper files. Information from these sources will be required for GIS applications and will have to be converted to digital form through keyboard entry.

This kind of data entry is commonplace and relatively easy to accomplish. Furthermore, GIS software can facilitate this process by providing the capability for creating data entry screens that can provide quality control on information being entered. This control involves checking entries against lists of acceptable entries or ranges of acceptable values. Use of automated quality control needs to be evaluated thoroughly for the types of data appropriate for this type of editing. It may be necessary to manually check some data elements, comparing the source document and an edit printout to ensure data quality.

**Figure 15: GPS Field Data Collection**

*In the field, with a GPS receiver pack on his back and a handheld computer for reading coordinates and inputting information. Photo courtesy of Measurement Science, Inc., Englewood, CO.*

### Field Data Collection

Advances in hardware and software have greatly increased opportunities for capture of GIS data in the field (e.g., sign of utility inventory, property surveys, land use inventories). In particular, electronic survey systems and the global positional system (GPS) have revolutionized surveying and field data collection. Electronic distance measurement services allow for survey data to be gathered quickly in automated form for uploading to a GIS. Sophisticated GPS collection units have provided a quick means of capturing the coordinates and attributes of features in the field. Figure 15 shows a typical GPS field data collection activity.

## Translation of Existing Digital Data

Existing automated data may be available from existing tabular files maintained by DEC or outside sources. In some cases, the federal government or other organizations may have automated map data that can fill some of the needs of DEC's GIS database. If the data have been automated using different GIS software than DEC's, a translation must occur to exchange the data and put it into a format for operation on DEC's system. Many programs are available that perform this translation and, in fact, many GIS packages can be acquired with programs that translate data to and from several "standard" formats which are accepted widely by the mapping industry and have been used as intermediate "exchange" formats for moving data between platforms (e.g., Intergraph SIF, USGS DLG, Census Bureau DIME and TIGER, AutoCAD DXF). Some vendors have developed custom exchange programs which bypass this intermediate format and translate data directly between two specific data formats. To facilitate GIS data exchange between multiple formats, the Federal Geographic Data Committee (FGDC) has developed the Spatial Data Transfer Standard (SDTS). SDTS is an intermediate format allowing the import and export of GIS data.

## Image Enhancement and Classification

Digital raster images from satellites or, in some cases, aircraft are often used for map creation. This is particularly true in the case of land cover mapping using satellite data sources such as LANDSAT or SPOT. These raster images can be purchased for specific geographic areas and dates and used with raster GIS software. In their initial form, they consist of raster files with brightness values assigned to each pixel for one or more spectral bands. Several processing steps may be followed in using these images for mapping work:

- Geographic rectification to match these images to a coordinate grid
- Image enhancement to manipulate contrast and other visual parameters to portray greater detail
- Automated classification to analyze the spectral values of all pixels and assign pixels to specific categories representing geographic variables (e.g., land cover categories).

A new generation of high-resolution satellite data that will increase opportunities and options for GIS database development is becoming available from private sources and national governments. These satellite systems will provide panchromatic (black and white) or multi-spectral data in the 1- to 3-meter ranges as compared to the 10- to 30-meter range available from traditional remote sensing satellites.

**Published by the:**

**Urban and Regional Information Systems Association**
**1460 Renaissance Drive, Suite 305**
**Park Ridge, Illinois 60068**
**Phone: (847) 824-6300 • Fax: (847) 824-6363**
**E-mail: info@urisa.org**
**www.urisa.org**