Geographic data model

The core data component of a GIS is often represented by a *geographic data model,* which is an industry or discipline-specific template for geographic data. A geographic data model offers the user flexibility in the design of the file management and database hierarchy. Geographic data models typically utilize a grid-based structure (known as raster) or a coordinate point structure (known as vector).

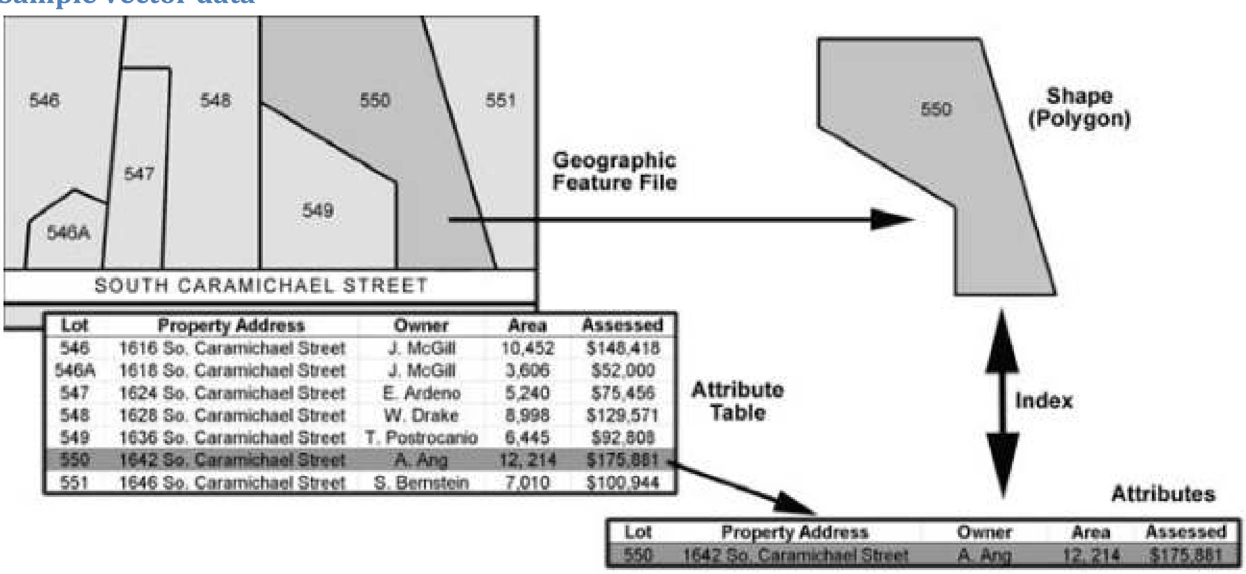
Vector data

Many times, geographic data are modeled in what is known as vector space. A vector space is simply a platform for geographic vector data, which use *x-y* coordinates with lines and shapes to depict Earth features. Geographic vector data store non-topological coordinate geometry and attribute information for spatial features.

Most standard GIS vector file formats consist of a feature file, an index file, and a linked attribute table. A *feature file* contains geographic object feature information, such as representative point, line, and polygon information. An *index file* contains unique identifiers that comprise more detailed information and help speed spatial feature queries. A linked *attribute table* is a matrixed table that contains explicative attributes for a group of spatial features.

The index file links the feature file data to the attribute table. Consequently, attributes and features exist in a strict one-to-many relationship, whereby geographic features can have multiple attributes. The vector model is readily recognized by those who use computer-aided design (CAD) environments, using feature geometry in a GIS vector model to instantiate points, lines, and polygons as objects.

Sample vector data

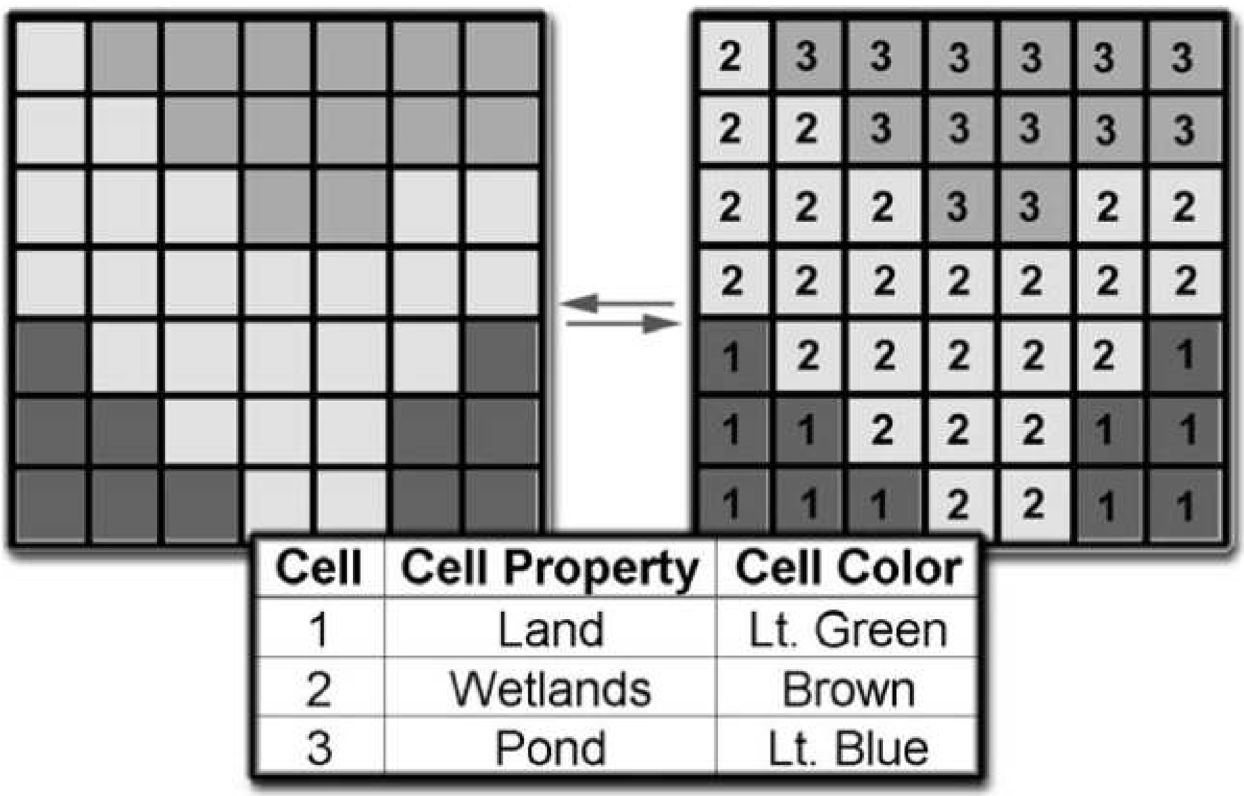


In most cases geographic vector data are *discrete data* occurring in cases where there are well-defined boundaries for physical representations or limited data values. The above figure involves tax lots that have visibly evident and discrete boundaries. Vector data can be created or modified directly by digitizing features using GIS development environments.

Raster data

Raster data are digital images represented by a grid of valued pixels, or *cells*. The image type and the number of colors represented determine the properties and appearances of these pixels. In most cases, raster data are more suited for representing *continuous data* than vector data. Continuous data are a numeric form of data usually associated with the physical measurement of boundaries that are not well defined. Additionally, the surfaces represented are, in many cases, estimated through statistical analysis. Raster imagery is perfectly suited for GIS sensor webs, a current industry trend that harnesses real-time observation and measurements over large, even global, regions.

Sample Raster data



Geodatabase

A *geodatabase* is a collection of geographic data sets, real-world object definitions, and relationships. Comparable to a Microsoft Access file, a geodatabase is a collection of geographic data sets and geometric features. A geodatabase furnishes the data organizational structure and workflow process model for the creation and maintenance of the core data product. In essence, the geodatabase is the heart of a GIS's management capability.

Georeferencing

Many data sources lack formal spatial referencing. Some CAD and GIS data sets are developed in a generic "design" space and have unique, often proprietary, types of referencing that simply need reinterpretation to be spatially integrated into a GIS environment. However, many of these sources are scanned raster data (digital imagery) that have only the coordinates of a raw pixel grid from an original scan. While these raster sources are often times unique and critical to a GIS project, images also need to be referenced from scratch, spatially transformed into a defined coordinate referencing system, then integrated and overlaid in a GIS environment. This process is known **as *Georeferencing.***

Georeferencing is the art of selecting common point locations in the real world using at least two data sources: an unreferenced source (such as a raster map) and a referenced source of the same area providing positional information. Basic georeferencing procedures involve point selection and transformation. For example, when a hardcopy map is scanned to an electronic file, it has no relationship to any real-world coordinate system. The georeferencing process establishes (or in some cases reestablishes) the relationship between image pixel locations and real-world locations. Georeferencing is accomplished by first selecting points on a source image (scanned raster map) with known coordinates for the real-world surface location (benchmarks, grid ticks, road intersections, and so on). These real-world coordinates are then linked to the corresponding pixel grid coordinates in the raster source image. After the image is georeferenced, each pixel has a real-world coordinate value assigned to it.

Queries: Locations and Attributes

Once data are adequately georeferenced and resident in a GIS, the user can then create a query expression to find the relevant data for a specific application. A query enables the user to search geographic data to collect location, feature, and attribute information from a relational database management system or geodatabase. The most fundamental test of geographic coincidence between data sources is a query by location in tandem with a targeting expression, such as *Intersection, Within distance of, Contained by, Share features with, Touch the boundary of, Are crossed by the outline of,* and *Completely in.*

Queries can also be executed on feature attribute tables and performed on specific fields. Records are selected in this manner; each record (or "row") in the table corresponds and is linked to a distinct geographic feature data set. Features conforming to certain attribute criteria can be selected and extracted using this method.

Quite simply, queries are the user's refining tool for taking the massive available data and selecting only those pieces that pertain to the application at hand.

Geoprocessing

Geoprocessing is the fundamental process of creating a derived set of geographic data from various existing data sets using operations such as feature overlay and data conversion. In a typical geoprocessing environment, the user applies GIS functions to a group of geographic (input) data to yield a precise output data set suitable for a particular application.

Most professional GIS software environments include a mission-specific geoprocessing interface, or "workbench," of geospatial dialogs and tools. These software environments usually include extensible scripting tools and compilers to automate, customize, and document geoprocessing workflows. The most important contribution of geoprocessing to the GIS big picture is the automation of repetitive tasks. Geoprocessing is an elaborate turnkey for efficient and clean geographic output.

Geoprocessors come in different forms. Many geoprocessing functions are embedded in a GIS environment. A GIS environment is a package of integrated GIS components: a geographic map control, a map layout designer, a data tree catalog, and so forth, of which geoprocessing is a member. However, many powerful stand-alone software applications offer specific, related subsets of geoprocessing functionality. Some include file format translators or spatial referencing transformation tools. Many professional GIS efforts actually require licensing these stand-alone accessories in order to reap the often-advantageous outputs of these stand-alone software applications.

There are eight categories of predominant geoprocessing operations, or

families of operations:

1. *Conversion.* Conversion is completely an issue of formatting. File format conversions (translations) and coordinate system referencing conversions are the most common geoprocessing conversion operations, and serve to characterize the conversion family.
2. *Overlay (union, intersecting).* Overlay involves superimposing two or more geographic data layers to discover relationships. In fact, overlay is intimately associated with the discipline of set topology, which defines the rules for valid spatial relationships between features in a geographic data layer.
3. *Intersect.* Geoprocessing computes a geometric intersection of the input features. The resultant features or portion of features common to all layers or assigned groups of same shape type (called a feature class) will be written to the output.
4. *Union.* Like intersect, union computes a geometric intersection of the input features. All features with the overlapping attributes from the input features will be written to the output feature class.
5. *Extraction (clip, query).* Like overlay, extraction is also intimately associated with the discipline of set topology. Queries help select the geographic data to be clipped or extracted, subject to a specific group of topology rules.
6. *Proximity (buffer).* Proximity is initiated through a query that selects geographic features based on their distance or proximity from other features. Geographic features include lines, points, and polygons.
7. *Management (copy, create).* GIS data management software is generally designed to facilitate the organization of a user's unique personal catalog (or collection) of geographic data. The intrinsic forms of all types of geographic data are accommodated by these applications.

Spatial Analysis

By their very nature, geographic data are intimately related to locations and feature attributes. Spatial analysis harnesses this duplicity through the study of geographic feature locations and shapes. Spatial analysis offers the user a range of procedures, tools, and interfaces varying in application and complexity. For example, creating a simple map in a GIS environment is a basic form of spatial analysis. Spatial analysis is not necessarily complex, but it is a process of reducing complex relationships to something simpler, possibly bringing to attention things that otherwise would have remained hidden to the user.