[DM-02-014] Vector Data Models

Abstract

Vector data models are a subset of geographic information through which data are encoded as a geometric definition of a feature, organized within a framework that relates this geometry to the spatial location, context, and proximal relationship.   Vector data models are often contrasted to raster data models, encoded in a regular grid spanning an extent with a fixed cell size. Vector data rely on geometric primitives, such as points, lines (sequences of linked points), and polygons (closed geometric forms), with variations on these and additional cases tailored to specific applications as they have emerged during the development of GIS. With the ability to define the location, placement, and interval of points that describe each feature, the resolution of the data are determined by the precision of placement of vertices from the sampled reality or when encoded in a processed form of the data. Due to their finite shapes, defined by their exact placement (point), the segment or overall length (line, polyline), or their area (polygon) vector data are commonly used to encode discrete, rather than continuous, features, with associated characteristics stored in an accompanying table of values corresponding to each shape feature.

Keywords: raster data, resolution, scale, spatial data, vector and object data models, vector data

Author & citation

Christman, Z. (2026). Vector Data Models. The Geographic Information Science and Technology Body of Knowledge (Issue 1, 2026 Edition), John P. Wilson (Ed.).
DOI: 10.22224/gistbok/2026.1.1.

Explanation

1. Vector Characterization of Space

Vector data formats are spatial features, described by geometric characteristics of a single point, a connected set of linked points (a line or polyline), or a closed shape (polygon), typically associated with a set of information in a tabular or relational database to offer additional context, metrics, or reference information for a set of features. For information about how vector features are encoded, their formats, and example uses, please see the GIS&T Body of Knowledge entry on Vector Formats and Sources, with added related entries linked therein.

Vector data models rely on a characterization of space and spatial features defined by forms or composites thereof, based on these geometric primitives of points, lines, and polygons. Spatial relationships among the represented areas can be characterized based on their topological relationships or the modeling of the connectivity of the segments in a spatial network, in contrast to other types of entity-based models that could be used to represent spatial features and relationships in a Geographic Information System framework.

Vector data best describe spatial features with discrete limits, such that a point indicates the the precise location; a line follows an exact path; a polygon circumscribes a defined area. As many real-world phenomena exhibit a continuum of values, a threshold for these ranges must be defined to draw the precise geometric shapes. Thus, choices are made about what is inside or outside of a feature or along a path. The scale of vector data is determined by the ratio of distance/area of the vector feature to the landscape or feature it represents. Additionally, the sampling interval of the vertices of these shapes influences the degree precision with which they are rendered in the geographic information system (GIS) and the ensuing products generated with these data. Further, a choice may be made as to which type of vector data is used to represent a feature dependent on the level of generalization, the ultimate use, and the spatial characteristics of the data. For example, a feature like a river may be encoded as a polygon of the area of the water or as a (poly-)line of the path, and a landmark may be indicated by a polygon of its area or a single point of its centroid or entrance.

2. Frameworks of Vector Data Models

Early and basic forms of vector data models involved creating digital representations of analog map features on a Cartesian grid. In these, each point was defined with a unique combination of x- and y-coordinates, with a line defined by a series of these encoded points, and a polygon defined by a closed loop of coordinates with the same start and end point. In this schema, each feature was saved separately and associated with one set of attributes. Limitations of this model include redundancy of shared points or lines, and a lack of the encoding of any spatial relationships among the features. Regardless of these limitations, this method was adopted by initial GIS software programs, bolstered by the speed and agility at which users were able to reference and visualize the data.

Vector data models are enhanced by the encoding of topology, which defines the relationships among the features, enabling considerations of adjacency, connectivity, overlap, intersection, and logical consistency. This encoding has several advantages that minimize redundancy among information and enable new uses for both querying and spatial analysis.

Here are some key conceptual vector data models that have been utilized over time:

- Spaghetti Model: the simplest vector framework, in which features are sorted separately and without defined relationships. It is suitable for visualization and simple querying and operations like overlay, but limited in other capacities. This is commonly used to represent independent, discrete, features, such as locations of interest, linear features like pathways, and independent boundaries.
- Topological model: incorporates interrelationships among spatial features, enabling routing and querying. This requires additional encoding and computational processing, and can incorporate boundaries that share borders, such as a map of voting districts within a municipality.
- Object-based models: features are defined as unique objects with associated attribute data, usually in a table, enabling each object to be referenced independently, as well

as the use of object-oriented programming principles and complex data structures and operations.
- Network modeling utilizes the topology of connected linear features to encode transportation networks, through which analyses of access, flow, coverage, and optimized distances can be rendered.
- Triangular Irregular Network (TIN) models utilize an array of triangles whose adjacent edges create a continuous array of triangles with the same slope and aspect across the represented facet. They are used to encode continuous surfaces that vary in multiple dimensions, which might otherwise be represented by a raster data model.
- Geodatabase model: a modern vector framework incorporated within a relational database system with a centralize structure, enabling verifications of data integrity and compatibility across large sets of information.

For an extensive comparison of spatial data models, including their uses and development, please see Peuquet (1984).

3. Specific Cases of Vector Data Models and Formats

These cases exist to encode complicated spatial information in data formats that leverage the geometric constraints of the vector data model with a specific application.

- The Dual Independent Map Encoding (DIME) data model was developed for US Census data to encode street networks and geographic boundaries for urban areas, creating a framework for geocoding and mapping, though without incorporating topological relationships (US Census, 2025). This was adapted to the Geographic Base File (GBF/DIME) model by the New York City government for further urban planning applications and updated into the Topologically Integrated Geographic Encoding and Referencing (TIGER) model in the 1980s to improve upon this initial structure and expand applications (Marx 1986), as well as the National Map Digital (NMD) dataset utilized by federal agencies for national mapping.
- The POLYVRT hierarchical model, developed by the Harvard Laboratory for Computer Graphics and Spatial Analysis was an early vector data format that enabled basic overlay operations and spatial queries (Peuker and Christman, 1975), by referencing each “chain” of line segments by start and end nodes and noting the adjacent features.
- Linear referencing enables the association of attributes as locations or spans along a continuous line feature without segmenting into separate features. These can include routes, measures (specific intervals with reference characteristics), and Events (entities that are linearly referenced along the route)
- Hexagonal models are an example of a tessellated model that use an array of regular hexagons encoded in a vector format to craft a complete coverage of a variable over space, with the advantages of enabling an irregularly shaped coverage to incorporate a continuous phenomenon, such as elevation, within a vector data format.

References

Marx, R.W. (1986). The TIGER system: Automating the geographic structure of the United States census. Government Publications Review 13(2): 181-201.

Peucker T. K., & Chrisman, N. (1975). Cartographic Data Structures. The American Cartographer, 2(1), 55-69.

Peuquet, D. J. (1984). A Conceptual Framework and Comparison of Spatial Data Models. Cartographica 21(4), 66-113.

United States Census Bureau. (n.d). Dual Independent Map Encoding. Last accessed April 13, 2026.