[FC-04-010] GIS Data Properties

Abstract

Data properties are characteristics of GIS attribute systems and values whose design and format impacts analytical and computational processing. Geospatial data are expressed at conceptual, logical, and physical levels of database abstraction intended to represent geographical information. The appropriate design of attribute systems and selection of properties should be logically consistent and support appropriate scales of measurement for representation and analysis. Geospatial concepts such as object-field views and dimensional space for relating objects and qualities form data models based on a geographic matrix and feature geometry. Three GIS approaches and their attribute system design are described: tessellations, vectors, and graphs.

Keywords: data property, database design, geographic concepts, GIS attribute, logic, spatial data model

Author & citation

Varanka, D. E. (2021). Data Properties. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2021 Edition), John P. Wilson (Ed.).
DOI: 10.22224/gistbok/2021.1.15.

Explanation

1. Introduction
2. Data Properties
3. Attributes of GIS Data Models
4. Discussion

1. Introduction

Geographic information systems (GIS) object properties refer to characteristics of digital data that represent entities and are stored in a database model. Properties are the technical basis for GIS data attributes that represent geographic qualities or relations that are abstracted from environmental interactions and modeled in computational systems. For example, decimal degrees are often used for geographic coordinates and integers are appropriate for coding choropleth maps. The blended implementation of computer science technology and intuitive recognition of geographic qualities support spatial information understanding and data handling.

There are a variety of usages of the terms ‘properties’ and ‘attributes’ as data structures in model concepts and software languages. Sometimes they are used synonymously. One distinction can be made that ‘properties’ mean intrinsic characteristics of GIS objects, meaning characteristics without which the object would not exist, while ‘attributes’ are ascribed characteristics whose alteration would not fundamentally change the object. The intended mental model for these terms used in this text is that properties, attributes, and their systems are constructs of the relation between representational and semantic

elements in databases that refer to real, including physical and conceptual, entities. An attribute system is related to the representation of an entity or class of entities. A property system is related to a representational entity or class of entities. The two systems integrate. Attributes contain data properties for their representation. Properties are characteristics of the attributes and their structure extends to allowable values.

Data models in GIS attempt to describe widely understood and agreed views of the world at three widely agreed levels of data abstraction: conceptual, logical, and physical. The conceptual level of data abstraction demonstrates the information content of the system as seen through a user's viewpoint. A user can understand what can be done with the system supported by but without specifying the technical implementation. The logical abstraction of conceptual information is implemented through a database design. Data model constraints and rules will generalize information as logical structures. The storage of data is determined at the physical level of abstraction. This paper reviews the character and significance of properties at these levels as they interact with attribute values in predominant geometric models used in GIS, including vector, tessellation, and their hybrid as Triangulated Irregular Networks (TIN). Other GIS logical models, such as object-oriented and graph, use properties that are related to vector and tessellation structures.

At the conceptual level of data abstraction, the capture of environmental attributes that persist for some time creates the semantic stability of a GIS but is complicated by varying individual perceptions and interpretations. Perspectives on geospatial feature categories, spatial references, and temporal variability are shaped differently by attention to details such as geographic scale, the users' cultural environment, and appropriate technical data structures. Communications between individualized perspectives succeed despite their variability because they are based on certain concepts that are universal, meaning widely agreed, and because contexts can be aligned or related. Some aspects of human cognitive processing involved in the perception and construction of world views may be relatively similar between people (Peuquet, 2002). For example, categorization is a basic construct of human thought (Rosch 1978).

The following sections discuss the logical and physical properties of data at the core of GIS attributes. Levels of abstraction are interdependent, so that logical abstraction may be thought of as a schema that organizes physical levels of data instances. This paper ends with a summary of key concepts about these points.

2. Data Properties

Database logical consistency, scales of measurement, and data types have significant impact on data storage, precision, and computation for analysis. The expression of concepts as formal logic can help avoid errors and self-contradictions in geographic information, but such documentation is rarely created. The Stevens (1946) system of measurement scales is frequently cited in data science, but spatial concepts introduce complexity beyond that system. In their role as descriptive and analytical representations, attribute systems may require a range of data types.

2.1 Logical consistency

Data properties in GIS attributes support the logical consistency of spatial databases. Some

characteristics of logical consistency of geographical information are listed below.

- Consistency of the data model accuracy to the real world. The selected data model is appropriate for the application if the attribute structure aligns with important qualities of the real-world entities they describe.
- Consistency of data types. File formats and processing rules are appropriate to the data model.
- Consistency of positional data. Positions are described with a similar range of locational precision based on data generalization.
- Consistency of database normal form. Undesirable dependencies and need for restructuring that interfere with effective application are reduced.

2.2 Scale of Measurement

Scales of measurement structure attribute values and support data applications ranging from simple description to complex statistical computing. According to the Stevens system, nominal scale data are values that assume no relative order and are represented by a name or other label. Nominal data are sometimes called categorical data because they form discrete categories with little or no overlap. The ordinal scale ranks data within a range that expresses the relative extent of certain object characteristics. Ranking reflects a certain order of non-numeric attributes, but the data values are not uniformly quantified. Interval scale data reflect uniform measurement relative to an arbitrary starting point. A common example of interval data is degrees of temperature measured in Fahrenheit or Celsius scales. The ratio measurement scale is based on an origin of zero and allows mathematical comparisons such as above or below, but more specifically, arithmetical computations such as two times as much, less than half, and others. Interval data scale variables support descriptive statistics, such as central tendency and variability. Ratio data scale variables support arithmetic operations including multiplication and division.

Though the Stevens system is widely relevant to GIS, spatial data uses additional data measurement scales such as raw number counts, absolute scales, cyclical measures, such as angles of a 360-degree circle, and graded category membership such as fuzzy sets (Chrisman 2002). The precision of spatial scales must conform with the method of geographical analysis.

2.3 Data type

A data type is a characteristic of a value that directs the meaning and the way the data can be represented and used. A data type provides a set of values allowed by a programming language expression. These constraints on the values define the operations on the data and the storage of those values. Most programming languages support basic data types used in a GIS such as byte, integer, and floating-point with several levels of precision for numeric measures and characters/strings for textual attribution. Custom data types can be created and added to software systems. If a data type used for spatial purposes is stored in a database management system, their allowable data types may not match exactly and must be mapped to the closest available.

3. Attributes of GIS Data Models

Before the widespread adoption of digital computational systems, maps displayed properties of spatial structures within a planar coordinate grid in an analog manner within certain display resolutions and geographical scales. Map users could see feature shapes, spatial relations, thematic categories, annotations, and other geospatial information. Kuhn (2012) listed ten core conceptual levels of spatial information abstraction rooted in geography and cartography that GIS logically and physically structures and represents, including location, neighborhood, field, object, network, event, granularity, accuracy, meaning, and value. Concepts of continuous vs. discrete and absolute vs. relative space can be used together in complementary and interdependent ways to support the logic of spatial concepts within different types of geographical analysis.

Having adopted predominant data structures developed in computer science, GIS describes concepts and their logic using attributes and properties for the processing of spatial data, including tessellation, relational table, hierarchical, and network structures. Spatial data have other challenges, including the storage and geometric processing of coordinates and determining topology for spatial relation reasoning. The GIS framework is a multi-dimensional matrix that includes spatial and non-spatial attributes.

This section describes two common logical models: tessellations, that correspond to the general geographic concept of a 'field,' and vector data, that prioritize 'object' types of representation (Couclelis 1992). Following these, section 3.3 describes graph data model properties for GIS. Like vector data, graphs resemble networks.

3.1 Tessellation

Tessellations are data models that approximate coordinate grids by partitioning a continuous surface of a geographical area into separate and adjacent geometric cells of a specific basic shape whose attributes are location values determined for each cell. Each cell has its own coordinates, which may be projected plane or geographic, that are defined intrinsically in the grid. For GIS attributes to logically resemble the real world, their associated location must conform to a reference system based on the geographic scale. The boundaries of the cells can be accurately calculated and generalized to a geometric shape with a coordinate representation for each cell. The areas within the cells share a common value of either location or an attribute with no internal detail and controlled by the resolution of the tessellation.

Tessellations can take many forms, but the two most common types are raster and Triangulated Irregular Networks (TIN). Each cell in a raster representation has an attribute value associated with that cell. In a TIN tessellation, each point of the triangle has x/y coordinates and an elevation value, and the slope of the triangle can be determined from those.

3.1.1 Raster

Raster data take the form of rectangular pixels with associated values. Attributes are normally limited to coordinate location values and a cell value. Panchromatic aerial photography attributes involve greyscale values that compose a visual representation of the ground surface. Digital Elevation Models (DEM) are a matrix of elevation values spaced regularly across the raster. Multi-band satellite imagery is a compilation of separate raster layers with values that correspond to different wavelengths in the electromagnetic spectrum. These values are areal, aggregated electromagnetic response from an

instantaneous field of view, not point, of a sensor. DEM values, currently (2021) derived from Light Detection and Ranging (lidar) data can be averages of all ground return points within a cell or some number of ground returns with inverse distance weighted values. Additional attributes for raster cells can be integrated using an identifier assigned to a file and linked to a relational table database.

Multiple rasters can be registered relative to each other and the cell values of each layer can be combined as input for an operation. A computational overlay operation requires that cells have identical resolution because the spatial reference system serves as the control for aligning geometry. If resolution is not identical, the attributes are resampled to align with each other.

Because of the simple basic organization of raster files, attribute data often repeat, especially along rows, columns, and adjacent cells. Data can be compressed without loss (lossless) to store more easily. Another way to save storage by aggregating data having the same or similar values is with the hierarchical data structures of recursive spatial decomposition. Adjacent pixels are subdivided until only the same attributes are common among them (Samet, 1990).

3.1.2 Triangulated Irregular Network (TIN)

TINs form a surface structure by establishing topological relationships between a set of point values. Spot measurements are assigned at meaningful point locations that vary according to characteristics of features to form the surface nodes. TINs connect three neighboring nodes with edges to form adjacent triangular planes. TIN point attributes can be used to calculate slope gradients and aspect measurements for a triangular facet. Similar to a DEM, every point can be associated with an interpolated value because the entire region is covered by triangles.

3.2 Vector data

Vector data take the form of geometric representations of object-like entities delineated as points stored as part of a projected geographic coordinate system. Lines connect points as visual vector representations and form polygons from lines that close at the endpoints. Because coordinates are stored as either integers or floating-point values, the appearance of a straight line from one coordinate pair to another depends on an appropriate resolution for the two-dimensional form of a grid.

Attribute tables are based on relational table database designs and combined with interactive cartographic display to visualize geographical data. Relational tables are a collection of fixed format records. The data are structured as a set of uniquely identified rows with a value for each attribute heading of the table columns. Attributes have domain sets that consist of the allowable values in terms of data types or other constraints for the record. The names of the attributes are mostly composed by database designers for application needs. Tables can be joined or related through ‘key’ fields duplicated between them.

Storing geographic coordinate geometry objects is problematic in the relational table data model because points are related by sequence. Most systems don’t allow table values to take the form of lists, so a single cell cannot contain a set or array of values even if together they represent a geometry object representing a single entity. Approaches called

the extended relational or hierarchical vector model use indexing as the solution. A hierarchical referencing system structures a spatial attribute to reference coordinates associated with a geospatial feature object. In addition to their use for storing coordinates, geographical objects such as land parcel polygons can be organized in relation to each other in this way using unique identifiers instead of by location (Lo and Yeung, 2002). The identifiers implicitly organize spatial referencing for nested entities such as land jurisdiction and administration, or for Census coding of demographic data by employing a lookup table to retrieve information. Such indexing systems are an efficient storage solution because codes are more compact than natural language descriptions. The same lookup table can be used for cartographic styling such as color ramping or another GIS operation. Geocoding is a hierarchical referencing system primarily used for linear features such as street addresses and adjacent parcel property ownership. This approach must ensure consistent coupling between segments of complex objects and of feature geometry to non-spatial attributes.

Topological spatial relations between entities are stored for vector data by identifying feature IDs as attribute values. The topological values of adjacent polygons or intersecting lines helps ensure data quality control, reduce data storage requirements by reducing duplication, and represent calculable spatial relations. Intersecting geometry objects that are required for performing complex spatial operations also do not easily adapt to the relational table model (Worboys, 1999), but relational algebra operations on layer-based and topological sets are possible, including union, intersection, and difference. As with resampling operations for raster resolutions, attribute reordering occurs after geometry overlay operations are complete (Tomlin, 1990). Spatial queries are supported on vector data attributes by some structured query languages that allow the specification of user-defined data types such as for geospatial data.

3.3 Graph Properties

Graph data model properties form a triple as an edge representing an attribute between two nodes that represent instances, classes or sets of instances, or literal strings. Nodes have any number of properties that connect between them to form graphs. The emerging study of geospatial semantics and knowledge graphs aims to develop a model that reflects an applied ontology whose object properties are formal logic axioms specifying the relation between those classes or instances. The logical axiom of properties may support inference or other types of reasoning to create a subgraph that forms the semantics of the entity in question. Graph query languages support navigation along property chains in addition to Boolean operations and string matches.

4. Discussion

Data properties structure the relation between geographic information, representational attributes, and the computational storage and application of those values. The appropriate selection of properties aims to maintain a logical relation among geographic information, GIS data models, scales of measurement, and data types. Tessellation and vector data model attributes try to describe real-world features by approximating key concepts in geography. Properties impose technical constraints on these descriptions to limit and support data storage and analytical functions.

Although GIS has adopted predominant computer science forms, some geospatial concepts vary from those rules. Some significant concepts that correspond well for spatial cognition and effective thinking are logical consistency, hierarchies, coordinate geometry, and topology. The dimensional matrix of GIS data is rooted in concepts of fields and cartographic representation forming the background of objects and their spatial relations. Hierarchies are widely applicable to GIS database logical structures as an approach to index data. Geometry objects are optionally stored as fully defined entities or simply arc segments. Topology is an indicator of data correctness, reduces data storage, and supports adjacency analysis in GIS. In graph databases, topology supports spatial relation inference through transitivity.

References

Chrisman, N. R. (2002). Exploring Geographic Information Systems, 2nd edition. New York, NY: Wiley.

Couclelis H. (1992) People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In: Frank A.U., Campari I., Formentini U. (eds) Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science, vol 639. Springer, Berlin, Heidelberg.

Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12), 2267-2276.

Lo, C.P. and Yeung, A.K.W. (2002). Concepts and Techniques of Geographic Information Systems. 1st Edition. Prentice-Hall, Inc.

Peuquet, D. J. (2002). Representations of Space and Time. New York: The Guilford Press.

Rosch, E. (1978). Principles of categorization. In Collins, A. & Smith, E. E. (eds.), Readings in Cognitive Science, a Perspective From Psychology and Artificial Intelligence. Morgan Kaufmann Publishers. pp. 312-22.

Samet, H. (1990). The Design and Analysis of Spatial Data Structures. Addison Wesley Publishing Company, Inc.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. Science 103, 677-680.

Tomlin, C. D. (1990). Geographic Information Systems and Cartographic Modeling,1st Edition. Englewood Cliffs, New Jersey: Prentice-Hall (now Pearson Education).

Worboys, M. F. (1999). Relational databases and beyond. In PA Longley, MF Goodchild, DJ Maguire, and DW Rhind (Eds.), Geographical Information Systems, 2nd ed., Vol. 1 Principles and Technical Issues (pp. 373-384). John Wiley & Sons Inc.