A Multidimensional Representational Model of Geographic Features

Open-File Report 2015–1241

U.S. Department of the Interior
U.S. Geological Survey
A Multidimensional Representational Model of Geographic Features

By E. Lynn Usery, George Timson, and Mark Coletti

Open-File Report 2015–1241

U.S. Department of the Interior
U.S. Geological Survey
U.S. Geological Survey
Suzette M. Kimball, Director


For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod/.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

ISSN 2331-1258 (online)
Contents

Abstract ...........................................................................................................................................................1
Introduction ....................................................................................................................................................1
Theory of Geographic Feature Representation ..........................................................................................2
System Design and Implementation ...............................................................................................................3
  System Design Overview ..........................................................................................................................3
  Representation of Changes in Spatial Configuration with Time .............................................................4
  Representation of Thematic Change with Time .......................................................................................4
  Representing Relationships ......................................................................................................................7
  Representing Features in Raster Data .....................................................................................................8
Populating the Feature Library ..................................................................................................................8
Conclusions ..................................................................................................................................................8
References Cited ..........................................................................................................................................8

Figures

1. Diagram showing representation of geographic entities as computer objects with attributes and relationships of space, theme, and time .................................................................2
2. Diagram showing the implementation of the conceptual model as a Feature Library provides support for multiple representations and multiple applications ..................3
3. Diagram showing the top level diagram for the Feature Library design ...........................................5
4. Diagram showing coastline features for St. Simons and Cumberland Islands in Georgia from two different time periods represented in the Feature Library ......................6
5. Diagram showing representation of time attributes in the Feature Library ...................................7
6. In this example of relationship implementation in the Feature Library, four relationship instances are shown by the octagons and five feature instances are shown by the double ellipses .................................................................8
A Multidimensional Representational Model of Geographic Features

By E. Lynn Usery,1 George Timson,1 and Mark Coletti2

Abstract

A multidimensional model of geographic features has been developed and implemented with data from The National Map of the U.S. Geological Survey. The model, programmed in C++ and implemented as a feature library, was tested with data from the National Hydrography Dataset demonstrating the capability to handle changes in feature attributes, such as increases in chlorine concentration in a stream, and feature geometry, such as the changing shoreline of barrier islands over time. Data can be entered directly, from a comma separated file, or features with attributes and relationships can be automatically populated in the model from data in the Spatial Data Transfer Standard format.

Introduction

Geographical phenomena, termed features, exist in the real world as individual, unique entities and can be modeled as objects in a computer representation. To support human abilities to conceptualize features at a variety of resolutions, scales, and with various attributes and relationships, there are many representations (objects) for any one geographical entity. This conceptualization of geographic phenomena as single entities with multiple representations can be effectively represented within a framework of feature objects that have spatial, temporal, and thematic attribute types that, in turn, are interconnected by way of explicit semantic relationship objects (Berry, 1964; Sinton, 1978; Usery, 1996a; Yuan, 1996).

Traditionally, geographic representation has been based on a space-dominant approach in which entities in the real world are modeled geometrically, as vector coordinates or raster matrices, with thematic attributes attached to basic geometric elements of points, lines, areas, or pixels. Berry (1964) established this approach with the concept of a geographical matrix. Current commercial geographic information system (GIS) software implements this concept. The structure of temporal attributes and of thematic and temporal relationships has received little attention until recently, but even these recent efforts use the spatial (geometric) characteristics of entities as the primary basis of features (Molenaar, 1991). Many spatio-temporal data models have been developed by researchers since the late 1980s, with each model representing a single type of geographical phenomenon. For example, Armstrong (1988) proposed snapshots for polygonal layers with invariant borders and structures were developed for cadastral by Al-Taha (1992) and Chen and Le (1996), for wildfire and thunderstorms by Yuan (1997, 2001), for the atmosphere by Peuquet (1994, 2002), for public boundaries by Wachowicz (1999), and for transportation by Koncz and Adams (2002). In many of these models, an object representation is used for the geographic phenomena, but usually the object is defined by geometry using a space-dominant view as defined by Wachowicz (1999). Extension to data with fields of objects (Cova and Goodchild, 2002) and to objects with field-like properties (Yuan, 1996, 1999) have also been added to the basic geometric and object models. Again, in these cases, specific applications and data types are the primary purpose of the representational model. The purpose of this report is to present a generic model that can be used to represent objects, fields, or features with both types of characteristics and that is capable of serving a variety of applications.

The objectives of this research are to present a theory based on the dimensions of space, theme, and time in which geographic features exist in the real world, and to provide an appropriate representation of this theory that can be used to implement geographic entities and processes. The theory is grounded firmly in geographic and cartographic abstraction and modeling principles (Peuquet, 1984; Peuquet, 1988; Guptill and others, 1990) and in cognitive category theory (Mark, 1993; Usery, 1993, 1996; Frank, 1998). The theory builds from basic research into the three aspects of geographical entities defined by Berry (1964) and used by many researchers since that time (Sinton, 1978; Usery, 1993,1996; Peuquet, 1994; Peuquet and Duan, 1995; Yuan, 1996, 1999). This research is a unique contribution because the developed theory supports a variety of data types including objects, fields, and objects with field-like properties and multiple applications from a single theoretical structure. This structure has been implemented in a completely object-oriented environment.

A design and an implementation of the theory developed to support extraction of geographic features from multimodal sources for the U.S. Geological Survey’s (USGS) The National Map are included in this report. In the report, the section titled

1U.S. Geological Survey.
2Oak Ridge National Labs.
“The Theory” presents the basic theory and further places this work in the context of other similar research. The basis of the system design and implementation is discussed in the “System Design and Implementation” section. Methods of populating the feature instances in the system with data are then presented in the “Populating the Feature Library” section, detailing the application of building a feature database for The National Map. The “Conclusions” section draws some conclusions from this work and addresses future research needs.

**Theory of Geographic Feature Representation**

To support multiple representations including different spatial geometries and resolutions, multiple thematic attributes and relationships, and multiple temporal attributes and relationships for single instances of geographical entities, a separation of the entity in the real world from the representation is required. This separation follows the definition of a feature as specified in the Spatial Data Transfer Standard (SDTS); that is, a feature consists of two parts, the entity that exists in the real world and the object that is the computer representation of that entity (U.S. Geological Survey, 2006). A framework for the space, theme, and time dimensions of geographic features is presented in Usery (1993, 1996). A similar model with three domains was proposed and implemented by Yuan (1996, 1999). Peuquet (2002) attempted to extend the framework by positing existence as a fourth element in a pyramidal structure. The resulting model assumes existence is equivalent to the other three dimensions of space, theme, and time. In reality, existence is determined by the presence of attributes and relationships of space, theme, and time, which is the representation used in this report (fig. 1). Thus, this theory of geographic feature representation is that unique entities exist in the real world and are represented as single objects with multiple associations of attributes and relationships along three basic dimensions of space, theme, and time.

Attributes are characteristics that are associated with single objects and thus apply to individual geographical entities (for example, a road has two lanes). Relationships are linkages or interactions between objects and thus involve two or more geographical entities (for example, one road, Independence Road, intersects another road, Tenth Street). Egenhofer and Franzosa (1991) developed the nine-intersection model of spatial relationships to account for the possible combinations of spatial objects. This seminal work by Egenhofer and Franzosa (1991) does not include other spatial relationships, such as distance and direction, but accounts well for topological combinations of spatial objects. A different model is needed to include other spatial relationships and relationships along thematic and temporal dimensions.

Historically, representations of geographical phenomena, implemented in GIS, were designed to model the spatial characteristics and relationships (topology) of entities with thematic attributes attached to the spatial representation. This space-dominant representation is a logical progression from

![Figure 1](image-url).

**Figure 1.** Representation of geographic entities as computer objects with attributes and relationships of space, theme, and time.
converting analog map representations, also a space-dominant approach, into digital form. Sinton (1978) concluded that in an analog map, of the three dimensions, time is held constant, theme is controlled, and space is measured. Galton (2001) agreed that spatial entities have constant time and a controlled theme. This space dominant approach was copied directly in the implementation of digital systems for cartography and GIS. Research in geographic representation during the past 15 years has struggled to remove the constraints of the space-dominant model and incorporate theme and time as equivalent dimensions (Peuquet, 2002).

With the space-dominant model of conventional GIS, thematic relationships and temporal attributes and relationships are not handled. For example, the layer approach in GIS includes a rudimentary capability to handle the “is_a” thematic relationship. The school ‘is_a’ building relationship is included in the layer approach by including all buildings in a single layer, which is an indirect method to handle a relationship resulting from categorizing geographical phenomena into arbitrary layers. The layer approach does not handle other thematic relationships, such as “contains” or “part_of.” For example, a lake may be a “part_of” a river system, and although the two entities may exist on the same hydrography layer, the “part_of” relationship must be determined from geometry or topology; it is not explicitly modeled as a thematic relationship.

Temporal aspects of geographical entities received early examination by Langran (1992), and since that time many alternatives have been proposed for discrete phenomena (Al-Taha and others, 1994; Chen and Le, 1996; Chen and Jiang, 1998) and continuous geographical processes (Peuquet, 1994; Yuan, 1997; Mark and others, 2003). Each of these approaches handles one aspect, such as event-based temporal modeling (Peuquet and Duan, 1995). The representation developed in this research includes the temporality of geographical entities.

System Design and Implementation

System Design Overview

Because the theory captures human cognitive abilities for multiple representation, a conceptual model has been developed to support attributes and relationships of geographical phenomena, including geometric representations such as vector coordinate lists and topology (Tang and others, 1996) and raster matrices (Usery, 1994a, 1994b; Yuan, 1996), spectral response curves, and mathematical formulas, which can be used to depict change in geometric and thematic attributes with time. The conceptual model structures a three-dimensional framework of space, theme, and time with associated attributes of individual features and relationships among features as shown in figure 1.

The objective of the implementation is to capture the conceptual model and support multiple representations of individual features including attributes and semantic and temporal relationships. For example, a road may be represented as a vector string of coordinates with attributes of width, date of construction, and others; topological relationships including “connected-to” other roads; and semantic relationships such as “composed-of” other roads and “is_a” interstate highway. Alternatively, the same road can be represented simply as a series of transportation pixels in a raster representation of land cover or as a series of spectral reflectances from a digital image. These multiple representations allow support for a variety of applications as shown in figure 2. The implementation of the Feature Library from figure 2 with a sufficiently encompassing data model allows support of the multiple representations and multiple applications.

Figure 2. The implementation of the conceptual model as a Feature Library provides support for multiple representations and multiple applications.
This conceptual model is comprehensive and flexible enough to handle all types of geographic entities, their attributes, and relationships. Single or multiple feature, attribute, and relationship objects are instantiated, as needed, to represent real-world entities. Attribute objects are connected to entity objects or to other attributes to adequately describe their respective entities. Relationship objects explicitly interconnect entities and they accommodate one-to-one, one-to-many, many-to-one, and many-to-many semantics. The model can be extended to add new operations and attribute types, such as those for geographic operations. Currently, this “feature space” model has been implemented in C++; its design is shown in figure 3.

The unique feature identifier (“id” in fig. 3) is the item in the library that identifies a particular occurrence of a geographical entity and separates one feature instance from other instances in the feature space. Although the identifier is arbitrarily assigned, it provides a convenient handle, currently implemented with pointers, for accessing characteristics of a feature instance. Each feature instance has attributes associated with it that can exist in the variety of forms specified in the conceptual model (for instance, numbers, characters, time stamps and ranges, equations, and spatial and geospatial graphic elements; Usery, 2000). These types of attributes potentially allow structuring of all characteristics of space, theme, and time of a geographical entity with the feature identifier. It is important to note that in this design, the spatial dimension is of no more relevance than the thematic or temporal dimensions; that is, the entity exists in geographic reality. The spatial configuration of the entity is simply an attribute of the entity at a particular time and in this representation, becomes an attribute of the feature. The representation of the spatial configuration of an entity can be as vector objects of points, lines, or polygons (Tang and others, 1996); or raster GIS and field or field-like objects (Usery, 1994a, 1994b; Usery and Pape, 1995; Yuan, 1996, 1999).

The top-level diagram (fig. 3) gives an overview of the Feature Library design and provides a glimpse of the potential power of such a structure. Feature is the core class, represents a single geographical entity, and has an arbitrary but unique identifier. The feature is shown to have zero or more attributes which may, in turn, have other attributes; thus, it is possible for a feature object to have a complex tree of attribute objects that describe the corresponding entity. Each attribute is a wrapper for a value that could be a number, an event, a time range, a string, a function, or a geospatial value, among other things; the design gives a subset of possible value types and can be readily expanded to include new ones. Each feature can also be in a relationship with one more other features. The relationship object represents semantics between features; each feature is also aware of each relationship to which it belongs. Features, attributes, and relationships are typed objects; that is, they have a corresponding type. Types have a name and possibly an associated, terse description. For example, a type object could be for a “ROAD,” “AIRPORT,” or “SAMPLING STATION.” Types themselves can be part of a type hierarchy. Each type can have an optional parent type as well as a possible set of “children,” or subordinate, types. These type parent/child relationships reflect the semantic relationships present in source data that would otherwise be lost. For example, one dataset might specify that “road” types have “number of lanes” and a “name,” whereas another dataset might also specify that “roads” have “surface composition.”

In the following sections, we examine the details of specific parts of the Feature Library design, including structure of attributes and relationships and the ability to handle multiple instances of attributes and relationships for a single instance of the entity or feature. These multiple instances of attributes and relationships may be from different time periods, with different spatial configurations and different thematic attributes. In all cases, a single feature instance is used; only the attributes and relationships (spatial, thematic, and temporal) change. Real-world data from applications in GIS analysis and modeling are used in the examples. Shorelines are used as the geographical entities and their spatial configurations and other characteristics at two different points in time are compared.

Representation of Changes in Spatial Configuration with Time

An example of two shoreline features and their associated attributes are provided in figure 4. Note that each shoreline has a different geometric configuration in 1990 from that in 1958. Each shoreline feature is represented by a double ellipses. Saint Simons Island shoreline is feature ID 18 and Cumberland Island shoreline is feature id 6. Both shorelines have two time attributes, one valued 1958 and one valued 1990. For each time attribute value, there are further attributes of Position, Accuracy, and F_code. In this example, the Position attribute values have changed from 1958 to 1990 for both shoreline features. The Accuracy and F_code values have not changed during this time. From an analysis and modeling perspective, a tool now exists that provides multiple instantiations of the attributes and relationships of the same geographic entity in various times and with various characteristics.

Representation of Thematic Change with Time

Time is implemented in the system as attributes of time stamps, which provides a basis for multiple representations of the same feature at different times (fig. 5). Change with time then can be modeled as difference in attributes with time stamp dates. In this example from a simulated water quality modeling study, feature id P1 is of type “Sampling Station” and has two attributes of type “Sampling Period.” From November 11, 1996, to November 17, 1996, the average chlorine (Cl) reading was 13.00 milligrams per liter (mg/L), and the average phosphate (PO_{4-P}) reading was 0.473 mg/L.
Attribute values can be arbitrary types including those not listed here.

Figure 3. The top level diagram for the Feature Library design.
Figure 4. Coastline features for St. Simons and Cumberland Islands in Georgia from two different time periods represented in the Feature Library.
From November 18, 1996, to November 24, 1996, the average Cl reading was 12.87 mg/L and the average PO$_4$-P reading was 0.461 mg/L. Note that the position and geometry of the feature have not changed, but the thematic attributes of chemical concentration have and are modeled in the Feature Library.

Dynamic implementation of time as an equation also is supported in the model. In this implementation method, the spatial and thematic attributes (characteristics) change as a function of a mathematical equation. For example, we model the change in PO$_4$-P concentration as a function of time, assuming that the changes between two readings at different times occur at a constant dilution rate, 0.002 mg/L per day. With a starting reading on November 18, 1996, of 0.473 mg/L, we model the reading on November 24 as follows:

$$\text{PO}_4^-\text{P}_{Nov24} = \text{PO}_4^-\text{P}_{Nov18} - (24-18)*0.002$$ (1)

where $\text{PO}_4^-\text{P}_{Nov18}$ and $\text{PO}_4^-\text{P}_{Nov24}$ indicate the phosphate values for the dates of November 18 and November 24, respectively.

Evaluation of the equation yields 0.461 mg/L, corresponding to the readings in figure 5.

Once the equation is implemented, the attribute for phosphate concentration is changed to the computed value. Because the rate is daily, the equation can be implemented for any or all days and provide a dynamic computation of the change in phosphate concentration on a daily basis. The series of values provide a direct method to implement visualization in the form of animation. This simple example indicates the potential available from the ability to include an equation as an attribute of the feature directly in the Feature Library.

### Representing Relationships

Relationships represent interfeature semantics. Typically, relationships represent connections such as connected_to or flows_into, and containment, such as is_a or part_of. The Feature Library design has a single relationship class that allows for any type of relationship. Relationship objects can be topological, thematic, or temporal in nature and may be deleted if the relationship changes. Relationships may be of any cardinality with regard to their respective features: one-to-one, one-to-many, many-to-one, and many-to-many. An example of relationships in the system design is shown in figure 6; the feature instances were populated from the National Hydrography Dataset (NHD) of the USGS automatically from SDTS format files. The diagram represents how a REACH feature in the NHD consists of other features using relationship objects (U.S. Geological Survey, 2015). In this case, the REACH feature consists of two STREAM/RIVER features and two ARTIFICIAL PATH features.

There are four relationship instances represented by octagons and five feature instances represented by double ellipses in figure 6. The relationships are all of the composed_of type. The diagram can be read as REACH feature id 2069058 is composed_of ARTIFICIAL PATH feature id 2064800, ARTIFICIAL PATH feature id 2064802, STREAM/RIVER feature id 2063420, and STREAM/RIVER feature id 2064454. The dataset used to create this example, NHD, models feature relationships using one-to-one cardinality; therefore, four relationship objects are used to model the REACH feature composition. Because the Feature Library design allows for multiple cardinality relationships, one composed_of relationship could have modeled this situation if desired. With time, if...
additional features become part_of the REACH, these can be depicted as new relationship objects. Likewise, if a feature is no longer part_of the REACH, the relationship can be deleted, but the deletion is recorded so the entire history of the feature is always available.

Representing Features in Raster Data

A feature in a raster dataset, whether continuous (such as elevation and temperature) or categorical (such as land cover and soils), can be a single pixel, a line of pixels, a contiguous set of pixels (polygon), or a set of non-contiguous individual or aggregations of pixels (Usery, 1994a). For example, a water well might be represented as a single pixel, a road by a line of pixels, a lake by a contiguous set of pixels, and a tank farm as a set of non-contiguous pixels. Storage of feature geometry may be as a listing of pixels with association of geographic coordinates, either implicitly or explicitly; a raster image mask defining the feature pixel set; or through a formula that, when applied to a stored raster layer, yields the appropriate set of pixels representing the feature. Because many geographic features have fuzzy boundaries, some pixels need to be represented with partial membership when specifying the geometry of a feature.

Methods to extract features from raster datasets have been developed for geographic entities with crisp (Usery, 1994b; Usery and Pape, 1995) and fuzzy boundaries (Usery, 1996b). Once extracted, these features can be stored in the Feature Library as individual pixel locations, lists of pixels, a binary raster mask, or as a vector outline after conversion of the individual pixels or pixel boundary along with the associated attributes and relationships along the three dimensions: space, theme, and time.

Populating the Feature Library

Populating the Feature Library requires a list of features with associated attributes and relationships; that is, a comprehensive ontology of geographic features. There are a variety of formats of geographic data in use, and many include feature lists and associated attributes. Some formats, such as SDTS and NHD, incorporate relationships of specific types. Because one target application of the Feature Library is The National Map, and no comprehensive ontology for it exists at this point, our approach is to use existing formats and datasets, such as Digital Line graph–Enhanced (DLG–E), Digital Line Graph–Feature (DLG–F), and NHD, that provide needed elements of the ontology and can be loaded with file import capability (U.S. Geological Survey, 2000).

Creating actual instances of features with all the associated attributes and relationships is a tedious, time-consuming, and expensive process. To facilitate this process, we have included capabilities in the implementation to import spatial data from standard formats such as SDTS and GeoTIFF; import tabular information from comma-separated-value (CSV) files; directly enter instance data, including all attributes and relationships through an interactive session; and interface to existing commercial software systems, such as ERDAS Imagine and Quantum GIS. For cases such as NHD in SDTS format, the import function automatically populates appropriate attribute and relationship classes as shown in figure 6.

Conclusions

A multidimensional representational model for geographic features was developed and implemented as a Feature Library. The feature model has significant capability for storing attributes and relationships of geographic features, allowing multiple representations of geometry and thematic characteristics at differing times or scales. The implementation of the model uses object constructs and was programmed in C++. Testing of the model for multiple representation and multiple applications was performed in several ways. First, the automatic import and storage of attributes and relationships from available data in SDTS and NHD formats was achieved. The capability for handling temporal change was tested in two ways: (1) spatially with a coastline change example including maintenance of appropriate relationships and (2) as thematic change over time with an example from water quality analysis and change in concentration of chlorine and phosphate over time.
time. In all cases, the model and implementation successfully met the requirements of multiple representations and multiple applications. Limitations of the implemented methodology include persistence in a database sense, good links to traditional geographic processing systems such as commercial GIS software, and query capabilities.

References Cited


A Multidimensional Representational Model of Geographic Features


