A Multi-scale, Multipurpose GIS Data Model to Add Named Features of the Natural Landscape to Maps

Aileen Buckley
abuckley@esri.com

Charlie Frye
cfrye@esri.com

Cartographic Research and Special Projects Group
Environmental Systems Research Institute, Inc.
Redlands, CA

There is a certain class of features on maps that are difficult to generate from traditional GIS databases — named features of the natural landscape. Physical features, such as mountain ranges, canyons, ridges and valleys, and named water bodies, such as capes, bays and coves, are often not found in GIS databases. This results in their omission on maps or at best their addition to the map as graphic type that is not georeferenced to the data used to make the map. This paper describes an inherently multi-scale GIS data model for physiographic features, and by extension named water bodies and named islands and island chains and groups, that can be used to create many different types of maps. The semantic model (what features to include), the representation (how to define the geometry of the features and their attributes), and the symbology (the specifications for both type properties and type placement) are discussed. In addition, the sensitivity of the representations and symbology to the software used for mapping are described. These issues are reviewed in hopes that others will be better able to use GIS data and software to make maps that include these features. Cartographers know that without the inclusion of the type for these names on maps, the products created are less informationally — and cartographically — rich. If more GIS databases with these features in them were developed, non-cartographers using GIS software to make their maps, as well as cartographers who have not generally had these data at hand, could produce better products.

Keywords: cartographic data modeling, indeterminate boundaries, physiographic features, GIS

INTRODUCTION

Maps produced by cartographic organizations, both public and private, often distinguish themselves from maps produced by organizations using solely GIS methods. The presence of names for natural features is one obvious way that maps produced by cartographers can be differentiated from maps produced by non-cartography savvy GIS users. Natural features are often represented by the type for their names alone and not by any distinct boundary that is delineated on the map. These features are typical of those described as “fuzzy features” in the GIScience literature (Couclelis, 1996; Brändli, 1996, among others) and various interesting characteristics of these types of features have been discussed (Mark and Turk, 2003; Smith and Mark, 2003; Waters and Evans, 2003). Because these types of features are not currently captured in geospatial databases as a common practice, they often do not show up on maps made exclusively with GIS. In this paper, the focus is on these types of features as they relate to their cartographic representations on maps and their GIS representations in databases. On reference maps, this includes named marine water

Initial submission, June 9, 2006; revised submission, August 25, 2006; final acceptance, August 29, 2006

bodies, such as bays, straights and gulfs, as well as named terrestrial physiographic features, such as mountain ranges, deserts, and ridges. The goal is to provide guidelines for how these types of features can be modeled in a database so that they can be included on more maps that are made with GIS software.

The type for natural features that appear on maps is currently preserved in forms not readily applicable to other uses; they are often stored as a type layer for a given finished map, which rarely makes them useful for another product. They are typically stored or archived in proprietary data formats, or sometimes they are stored only in the paper product. Increasingly, names for geographic features are stored in a digital database with point locations that sometimes do not correspond to either their actual geographic position or the best cartographic position for type placement. By and large, geographic names are certainly not at this time linked to mainstream spatial information infrastructures.

To complicate matters further, the way the names of natural features without delineated boundaries are drawn on maps is highly tailored to the shape and nature of the features so as to imply their extent without having to draw debatable demarcating lines on a map. It would be valuable to have a flexible lowest common denominator GIS representation of these features (that is, a primary feature type for these features that can be used on many types of maps at many different scales). In most cases, this will be a polygon within which the type for the feature would appear. Digitized correctly, this polygon would be able to encompass the appropriate location for the type at any scale and at the smallest scales could be treated as a point for type placement. Such a versatile representation would be more useful and as a result would likely become more widely used by more map-makers to achieve a higher level of information quality as well as cartographic quality.

"The goal is to provide guidelines for how these types of features can be modeled in a database so that they can be included on more maps that are made with GIS software."

In order to discuss a practical means of doing this, there is first a need for an understanding of the types of natural named features with indeterminate boundaries that appear on maps. Then guidelines for how those features should be represented in GIS, driven by an understanding of how they have typically been depicted or symbolized on maps, is needed. How the software handles feature type can affect the mapped representations and therefore should be considered in the database development.

TYPES OF FEATURES AND THEIR GIS REPRESENTATIONS

Features with indeterminate boundaries can be organized into general themes based on the category of information they represent. There are several useful taxonomies for organizing thematic information on a map. The basis for our framework is the model discussed by Arctur and Zeiler (2004), which describes a set of themes commonly found on many base or reference maps. Those that are relevant for our purposes are: transportation, cultural, boundaries, hydrography, hypsography and surface overlays. More recent research has identified an additional base map theme, physiography, also relevant to our work (Buckley *et al*., 2005).

Within the themes of physiography and hydrography, one can identify specific feature types. A number of taxonomies already exist which can be useful when determining the types of labeled features that will appear on maps. For some sources, the principal concept behind the taxonomy is map or GIS-based. For example, the Alexandria Digital Library (ADL) Gazetteer contains feature type classifications from three independent typing schemes (ADLP, 2004); all of the names in the database are associated with one or more feature type terms from the 1) *ADL Feature Type Thesau-*

rus, which are drawn from a variety of authoritative sources, including glossaries and government publications, and portions of the names are associated with either 2) the gazetteer type terms from the U.S. Geological Survey or from 3) the U.S. National Geospatial-Intelligence Agency (NGA). These feature types have specific descriptions and are polyhierarchical; however, they are simply descriptions that have no associated geographic feature data.

An alternate feature name and type source is the Digital Geographic Information Exchange Standard (DIGEST) data model used in conjunction with Vector Map (VMAP) data (DIGEST, 2001). The VMAP database consists of textual, attribute, and geographic data, and physiographic features are stored as points, lines or polygons (Figure 1). This is an interesting example of the implementation of a GIS database for named places; however, there are some serious limitations. The VMAP data files have a field for the name of a location (NAM) but it is rarely populated; indeed, the Named Location features in this database do not even have a name field. Instead, it is intended that the names of the feature be derived from the GEOnet Names Server (GNS) for all locations except the United States and Antarctica, which come from Geographic Names Information System (GNIS) database.

"These feature types . . . are simply descriptions that have no associated geographic feature data."

Additionally, as is apparent in Figure 1, there are some problems with the classification of feature types along source map sheet boundaries. To further complicate matters, VMAP data are available in low resolution (Level 0), medium resolution (Level 1) and high resolution (Level 2). Level 0 provides worldwide coverage of geo-spatial data and is equivalent to a small scale of 1:1,000,000; this is a slightly more detailed reiteration of the Digital Chart of the World. Level 1 data is equivalent to a medium scale 1:250,000 resolution. Level 2 data is equivalent to a large scale 1:50,000 resolution. Because of the varying scales, a feature that is represented as

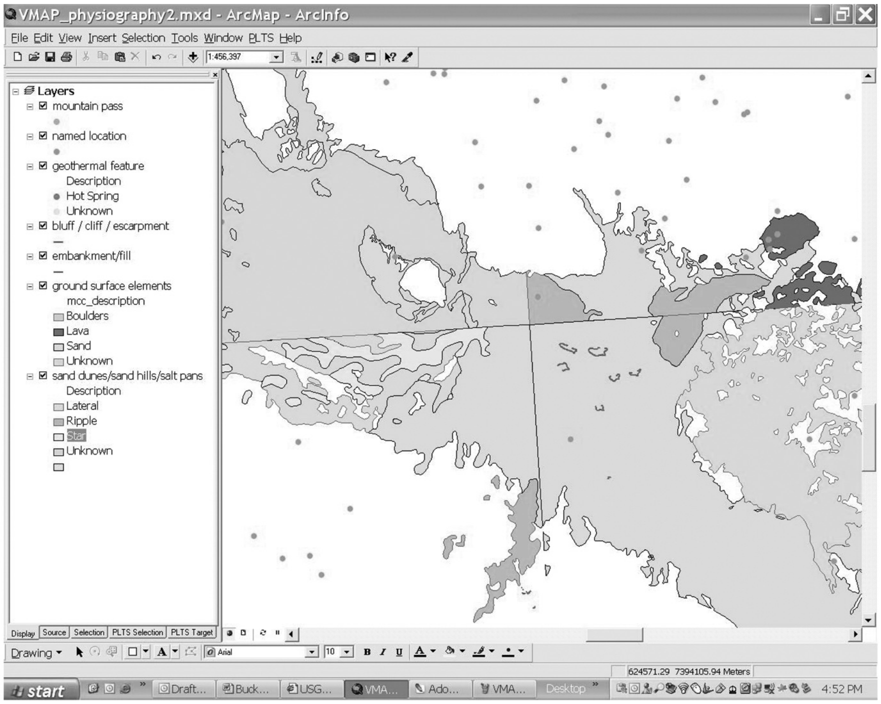


Figure 1. Physiographic features in the VMAP data are designated as points, lines or polygons. Obvious problems can occur at the boundaries of the map sheets that were used as the source documents. Not all names for these features are stored in the GIS dataset. (see page 78 for color version)

a polygon at one scale can become a point at smaller scales. The varying feature representations and the technicalities related to maintaining the names linkages make using these data for multipurpose, multi-scale maps problematic.

A third example of a physiographic feature taxonomy can be derived from the types of features found on the U.S. Geological Survey (USGS) topographic map series (USGS, 2002). Features on the USGS topographic maps have been divided into the following categories: coastal features and shorelines, escarpment features, glacial features formed by alpine glaciation, glacial features resulting from continental glaciation, miscellaneous features, mountain features, plains features, plateau features, solution features, valley features, volcanic features, water features, and wind features. Definitions of many of the features are provided in the National Mapping Program Standards for maps at various scales (USGS, 2005); however, not all of the features are described or defined therein. Furthermore, features that appear only as type for names on USGS maps, such as named landforms, exist only in the GNIS database as point locations that represent the cartographic position for type placement on the original map, and these locations sometimes do not correspond to their actual geographic positions.

Since no ideal GIS databases exist that can be used for mapping features with indeterminate boundaries in order to show their geographic names on maps, the challenge was to determine how that type of database could be specified, compiled and used. It was determined that all these features could be stored as either points (in a very few cases) or polygons. Lines could be used but they do not scale as well as polygons. With polygon features, the label can shift location based on the scale and extent of the area being mapped; with line features the text must be placed relative to the line that likely was drawn for a particular scale and extent. Because the features are only used to label the map, a further categorization of the point and polygon designations was formulated so that they can be used to specify particular placement properties for the feature types based on polygon geometries. These distinctions are clarified further in the "Reclassification of Feature Types" section below.

"Since no ideal GIS databases exist that can be used for mapping features with indeterminate boundaries in order to show their geographic names on maps, the challenge was to determine how that type of database could be specified, compiled and used."

The left column of Table 1 lists some of the most common types of natural features that lack delineated boundaries on maps. The content of the table was extended to show how other types for non-physiographic regionalizations could also be considered; however, in this paper the concentration is only on the first three types of features. The second column

Types of Features with Indeterminate Boundaries	GIS Theme	Required Topology
Named Marine Water Bodies	Hydrography	Must Nest, No Overlaps
Named Physiographic Features	Physiography	May Nest or Have Partial Overlaps
Named Islands and Island Chains and Groups	Physiography and Hydrography	Must Nest, No Overlaps
Neighborhoods and Districts, Vernacular Regions	Cultural and Transportation	May Nest or Have Partial Overlaps
Land Cover, Geology, Soils and Other Overlays	Overlays	No Overlaps

Table 1. Thematic organization of features with interminate boundaries.

shows where, relative to a base map data model, these features should be modeled within the GIS database. The Topology column shows the nature or restrictions of the spatial relationships of features within each theme. Topology is closely related to the semantic model for how these features are defined by our languages and cultures, as are the mereological (that is, the part-to-whole) relationships. It is therefore useful to consider the mereotopological relationships, which marries wholes, parts and boundaries with their topological roles and relations (Smith, 1994 and 1995; Varzi, 1997). These relationships are discussed for each of the three types of features described in this paper.

Many features with proper names exist within each of these types, and they often have indeterminate boundaries. For the most part, only the type for the names of these features are depicted on maps, and map readers draw conclusions about the actual location of the features from the contextual relationship that the type has with other features on the map. For instance, a map may have the names of canyons on it and the context that helps map-readers are the contour lines and a hillshaded representation of the terrain.

The following section describes the semantic model for the types of geographic features considered in this paper. Because the type specifications and the placement rules may vary for different feature types, it is useful to develop a set of valid feature type values that can be used as an attribute to distinguish the features in the GIS dataset. The valid values for the feature types should be based on (1) the requirements to make a certain set of maps at varying scales, and (2) the source documents from which the features were delineated. In the development of other semantic models for typing the features, both these factors should be kept in mind. Other semantic models could be derived from any of those that currently exist (e.g., the ones developed in this research, ADL, DIGEST, or others), and they could be easily modified to meet the mapping requirements for other projects or organizations.

> *"For the most part, only the type for the names of these features are depicted on maps, and map readers draw conclusions about the actual location of the features from the contextual relationship that the type has with other features on the map."*

Named Marine Water Bodies

A suggested set of feature types for named marine water bodies is:

- Bay
- Bight
- Channel
- Firth
- Gulf
- Inlet
- Ocean
- Passage
- Sea
- Sound
- Strait

The mereotopology for named marine water bodies is that they must nest and that they do not overlap. For instance, the Straits of Florida do not overlap the Gulf of Mexico or the Atlantic Ocean.

It may be desirable to include features that are either antiquated, like the Sargasso Sea, or legendary, like the Bermuda Triangle. These features should be managed as exceptions to the rule, unless modeling a database that is devoted to all but contemporary water body names and locations (i.e., those that are currently in use or in use at the same time). If these features are included in the database, an attribute should be used to define the feature type for marine water bodies; valid values may include historical, relic, and legendary types of water bodies.

Since marine water bodies are typically represented with type only on maps, the feature types can be categorized from the semantics for the feature names. That is, these feature types are based solely on the name rather than some taxonomy relating to physical characteristics. Thus, the

Gulf of Mexico is a gulf; the Sea of Japan is a sea, etc. The type is displayed so that the largest water bodies have the largest type sizes. Type is typically aligned along the graticule, or, for protracted water body shapes, along the major trending axis. For reference maps, if the scale of the map is such that the type will not fit roughly into or just slightly overrun the bounds of the feature, it should not be shown on the map. If the water body is too small to contain its type, but is critically important to the purpose of the map (i.e., has notoriety), then a leader line should be used to identify the location of the water body.

Figure 2. Small-scale representation of named physiographic features and regions (Raisz, 1965).

Named Physiographic Features

Identifying and classifying named physiographic features relates to work that others have done in regionalization. Geographers and others (Lobeck, 1932 and 1947; Fenneman, 1938; Fenneman and Johnson, 1946; Raisz, 1957) have sought, at small scales, to regionalize the United States and North America based on broad geologic or geomorphic characteristics. Figures 2 and 3 are excellent examples of such work.

A digital example of physiographic regionalizations such as those illustrated by Raisz is the United States Geological Survey dataset (USGS, 1992) of named physiographic *Divisions, Provinces, and Sections* that regionalizes the United States based on a 1946 map by Fenneman and Johnson (1946) (Figure 4).

These kinds of smaller-scale regionalizations are used as the basis for further definition of medium-scale physiographic features. Currently, individual physiographic features are typically not part of GIS or spatial data repositories. Although some GIS

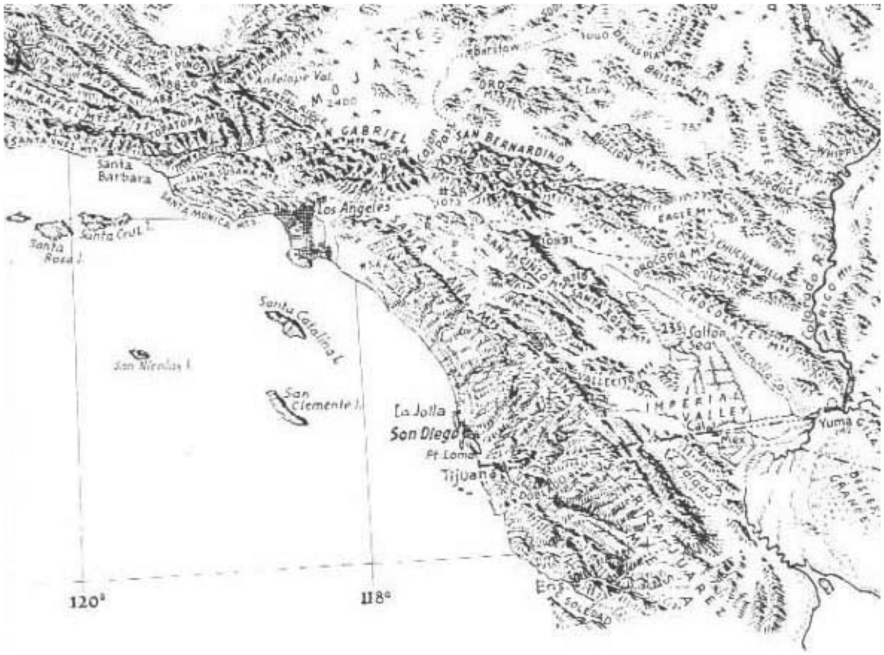
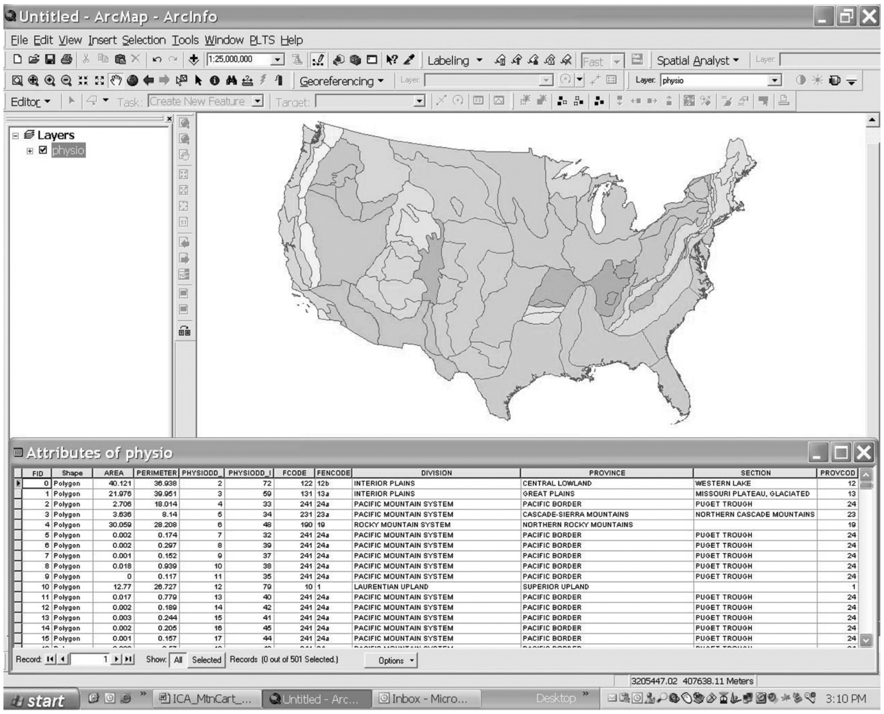


Figure 3. Medium scale representation of named physiographic features and regions (Raisz, 1965).



FID	Shape	AREA	PERIMETER	PHYSIODD_	PHYSIODD_I	FCODE	FENCODE	DIVISION	PROVINCE	SECTION	PROVCOD
0	Polygon	40.121	36.938	2	72	122	12b	INTERIOR PLAINS	CENTRAL LOWLAND	WESTERN LAKE	12
1	Polygon	21.976	39.951	3	59	131	13a	INTERIOR PLAINS	GREAT PLAINS	MISSOURI PLATEAU, GLACIATED	13
2	Polygon	2.706	18.014	4	33	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
3	Polygon	3.636	8.14	5	34	231	23a	PACIFIC MOUNTAIN SYSTEM	CASCADE-SIERRA MOUNTAINS	NORTHERN CASCADE MOUNTAINS	23
4	Polygon	30.059	28.208	6	48	190	19	ROCKY MOUNTAIN SYSTEM	NORTHERN ROCKY MOUNTAINS		19
5	Polygon	0.002	0.174	7	32	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
6	Polygon	0.002	0.297	8	39	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
7	Polygon	0.001	0.152	9	37	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
8	Polygon	0.018	0.939	10	38	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
9	Polygon	0	0.117	11	35	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
10	Polygon	12.77	26.727	12	79	10	1	LAURENTIAN UPLAND	SUPERIOR UPLAND		1
11	Polygon	0.017	0.779	13	40	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
12	Polygon	0.002	0.189	14	42	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
13	Polygon	0.003	0.244	15	41	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
14	Polygon	0.002	0.205	16	45	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24
15	Polygon	0.001	0.157	17	44	241	24a	PACIFIC MOUNTAIN SYSTEM	PACIFIC BORDER	PUGET TROUGH	24

Figure 4. USGS dataset of physiographic regions and provinces for the conterminous United States. (see page 78 for color version)

datasets with limited utility do exist, as described earlier, most digital representations of these features are managed in place names indexes with no associated geometric representation useful for cartography. Data for larger scale maps is even more critically lacking, underscoring the need for a model that envelops more local features such as peaks, ranges, mountains, valleys, deserts, canyons, flats or playas, passes, etc.

As with named marine water bodies, the development of a set of physiographic feature types should relate to the map or map products that will be made from the data, as well as the source documents used to compile the dataset. The set of feature types for named physiographic features developed in this research include:

> *". . . the development of a set of physiographic feature types should relate to the map or map products that will be made from the data . . ."*

- Badlands
- Bar
- Basin
- Bend, Land
- Bend, Water
- Bluff
- Butte
- Canyon
- Cape
- Carolina Bay
- Cliff
- Crater
- Delta
- Desert
- Dunes
- Escarpment
- Fault Zone
- Gap
- Hill
- Hills
- Incline Flow
- Incline Flow, Earthen
- Incline Flow, Lava
- Incline Flow, Rockslide
- Incline Flow, Slope
- Island
- Isthmus
- Landfall
- Lowlands
- Mesa
- Moraine
- Mountain
- Mountain Range
- Natural Arch
- Natural Bight
- Pass
- Peak
- Piedmont
- Pinnacle
- Plains
- Plateau
- Playa
- Promontory
- Ridge
- Saddle
- Terrace
- Uplands
- Valley
- Volcano, Active
- Volcano, Inactive

A few of these feature types warrant further description. Playas include all flats (mud, sand, etc.). Peaks are only the uppermost portion of a named mountain and do not slope upward to other peaks or mountains; peaks may have no logical or semantic link to mountains with “peak” in their name. Islands refer to exposed named land masses within inland water bodies — islands within named marine water bodies are managed in a separate dataset, as discussed below.

The mereotopological relationships for physiographic features vary by type; therefore, the general rule for the topology of named physiographic features is that they may nest, or they may have partial overlaps. Unlike named marine water bodies, some named physiographic features are polyhierarchical with respect to the fact that for different map scales, there are different contemporary regimes of features, sometimes more than one at the same scale. One cannot assume that the mereotopological relationships from one scale will necessarily hold for other scales, at least for the purposes of mapping. For instance, mountain ranges may contain mountains, which contain one or more peaks, but sometimes mountains contain other mountains and a single mountain can contain more than one peak. The impact of cultural and linguistic history is not necessarily logical when it comes to the names of these types of features. Also, unlike named marine water bodies which can be suitably categorized by the feature type indicated in the name, some physiographic features cannot be semantically categorized this way. For example, a flat may be a playa or a mud flat, and a mount may be a peak or a mountain or even a small rise. The name of the feature does not always clarify the distinction.

Named physiographic features, like named marine water bodies, are typically represented on the map with type, unless the feature is too small at a given map scale, in which case the type should not appear at all unless it is determined to be a critically important feature. Unlike named marine water bodies, the type for these features is not aligned to the graticule. Rather, type placement for physiographic features is guided more by the geometric major trending axis. In addition, the representation of the terrain should be used to guide how to drape, nestle, or span the type for a given feature. This is a rather artistic process that requires the map-maker to adequately imply where the feature exists by the positioning of the type.

“This is a rather artistic process that requires the map-maker to adequately imply where the feature exists by the positioning of the type.”

Named Islands and Island Chains and Groups

Islands and island chains and groups generally have these terms in their proper names, and there are only a small set of synonyms, such as archipelago and chain. This makes the semantic model for, and the identification of these features fairly simple and straightforward.

The set of feature types for islands and island chains and groups developed for this research is:

- Archipelago
- Chain
- Islands
- Atoll
- Group
- Isle
- Barrier Island
- Island

The mereotopological relationships for islands and island chains and groups are pretty straightforward. Islands may or may not be part of an island chain or group; therefore, the topological restriction is that they must nest with no overlaps. Island chains and groups, however, are cartographically represented with type that is by necessity drawn outside the geometric bounds of the individual island features.

In the cartographic representation of islands, the extent of terra firma is generally neither fuzzy nor indeterminate. The type for an island is placed inside the island if it fits at the given map scale. If an island is too small to contain its type, the label is placed outside the island using methods for positioning type associated with a point features. The typeface for islands that are countries typically differs from those that are not. If an island is not a country, the country that has dominion over that island may be shown as well.

Island chains and groups are more complicated, as the type placement for the name of the features is somewhat similar cartographically to that of mountain ranges. Ideally the type extends over of geographic space for all islands shown that are officially part of the chain or group. The type for an island chain or group may overlap islands, that is, it does not have to be positioned without exception outside of the geometry of the individual islands. Typically the type for an island chain or group follows the major trending axis of the chain or group. If a group or chain of islands is labeled, the type placement is more similar to that of larger named marine water bodies in that the type is aligned to the graticule, and is positioned large enough within the group or chain to imply virtual ownership of the islands by the group or chain.

GIS REPRESENTATIONS FOR CARTOGRAPHIC USE

"A key factor is that these features will be used expressly to produce maps."

This section discusses what to store in the GIS database for the geometry of features without definite boundaries. A key factor is that these features will be used expressly to produce maps. To successfully store representations that facilitate mapping, the common assumption that the geometric representation of a feature in a GIS is the most accurate representation possible will be contradicted. Typically some process of cartographic abstraction happens prior to representing a GIS feature on a map. Conventional wisdom or logic dictates that the abstracted representation cannot enhance the accuracy or precision of the feature's geometric coordinates. However, if a feature's geometry describes a shape that adequately encompasses that feature's location, rather than the feature itself, an inherently non-specific cartographic representation of that feature can be abstracted and potentially function as a better representation on a given map.

To facilitate the abstraction of GIS data into cartographic representations, cartographic attributes must often be stored with the GIS representation. These attributes, together with an adequately encompassing geometric representation stored in a GIS, are the inputs to the cartographic abstraction process for map production. The following sections provide a discussion of how specific types of features with indeterminate boundaries might be modeled in the GIS so that they may be used as the basis for cartographic representations. As the title of this article indicates, one key aspect is that a single GIS feature must serve as the basis for multiple cartographic representations at varying scales. As suggested, in most cases, this requires a polygon within which the appropriate location for the type could appear at any scale, and that at the smallest scales could be treated as a point for type placement.

Named Marine Water Bodies

The GIS representation of a named marine water body feature is a polygon with the following attributes:

- **PolyID** [Integer]: This is used as a primary or foreign key in a join or relationship to other databases, such as a names database or a revision information table.

- **Name** [Text]: This is the proper name of the feature as it would appear on the map.
- **FeatType** [Integer]: This is used to categorize features based on different type style or placement requirements. In cases where the semantics of the name do not match the actual type of feature, this attribute also stores the actual feature type.
- **SizeClass** [Integer]: This is based on the area of the polygon and is used for two purposes: 1) to determine whether the feature will be represented on a map at a particular scale, and 2) to determine the size of the type that will be used to represent that feature at that scale.
- **Sources** [Integer]: Ideally this is used as a primary or foreign key for a join to a look-up table of sources. Unless an agency has performed the primary research to determine the name of a feature, this field can be used to identify at least three independent sources demonstrating that the feature name exists in the public domain.

"The symbolization and placement of the type is performed by the GIS software using the polygon geometries and attributes in the application of the type style and placement rules."

Figure 5 shows the GIS representation of named marine water bodies between the southern United States and northern Central America. Notice the lines that separate unique named bodies of water.

Figure 6 shows the cartographic representation, which does not indicate the feature outlines but does display type that is appropriately sized and positioned within each water body relative to the chosen map scale. The symbolization and placement of the type is performed by the GIS software using the polygon geometries and attributes in the application of the type style and placement rules. The polygon attributes are used to determine the content of the text string, how to set the type style, and whether to include it at a given map scale. The polygon geometries are used to determine the type placement.

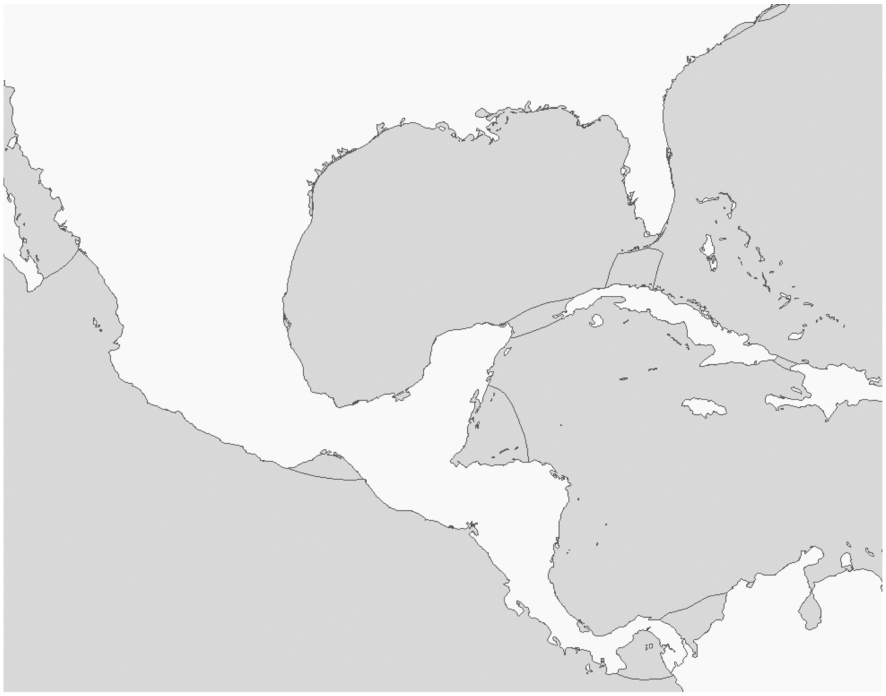


Figure 5. GIS representation of named marine water bodies as non-overlapping polygons with an extent adequate to label the features at multiple scales.

Figure 6. Cartographic representation of named marine water bodies.

Named Physiographic Features

The GIS representation of a named physiographic feature is a polygon, with the exception of named summits, in which case the representation is a point. As with named marine water bodies, the polygon features are never actually drawn on the map. Rather, they are the basis for creating and placing the type on the map at any scale. Along with named marine water bodies, named physiographic feature polygons are excellent examples of multi-purpose datasets because they can be used for any map at any scale. This method of feature representation provides a vast improvement in efficiency over traditional methods of only maintaining the type for the feature, which serves a very narrow range of cartographic purposes (i.e., maps).

> *". . . named physiographic feature polygons are excellent examples of multi-purpose datasets because they can be used for any map at any scale."*

The GIS representation of a named summit is a point with the following attributes:

- **PolyID** [Integer]: This is used as a primary or foreign key in a join or relationship to other databases, such as a names database or a revision information table.
- **Name** [Text]: This is the proper name of the summit as it would appear on the map.
- **FeatType** [Integer]: This is used to categorize features based on different type style or placement requirements. This is also used for filtering out smaller features like hills, or for showing only mountain top elevations.
- **Elevation** [Integer]: This is an integer only because most maps do not require specification of elevations as the sub-foot or -meter level.
- **Units** [Short Integer or Boolean]: This is used to denote whether the elevation is in feet or meters.
- **Sources** [Integer]: Ideally this is used as a primary or foreign key for a join to a look-up table of sources. Unless an agency has performed

the primary research to determine the name of a feature, this field can be used to identify at least three independent sources demonstrating that the feature names exists in the public domain.

The GIS representation of a named physiographic feature is a polygon with the following attributes:

- **PolyID** [Integer]: This is used as a primary or foreign key in a join or relationship to other databases, such as a names database or a revision information table.
- **Name** [Text]: This is the proper name of the summit as it would appear on the map.
- **FeatType** [Integer]: This is used to categorize features based on different type style or placement requirements. As above, this is also used for filtering out smaller features or for showing only selected features.
- **Order** [Integer]: In general, this is a classification of the size of the features, although the values of this attribute for individual features could also be modified to reflect their notoriety.
- **Sources** [Integer]: Ideally this is used as a primary or foreign key for a join to a look-up table of sources. Unless an agency has performed the primary research to determine the name of a feature, this field can be used to identify at least three independent sources demonstrating that the feature name exists in the public domain.

"Unless an agency has performed the primary research to determine the name of a feature, this field can be used to identify at least three independent sources demonstrating that the feature names exists in the public domain."

Figures 7 and 8 give an indication of how the GIS features for named physiographic features look. In Figure 7, the extents of the physiographic features stored in the GIS can be seen. Colors relate to the size of features to emphasize the smaller features in this display. Features for all potential map scales are represented; for example, the canyons shown in Figure 8

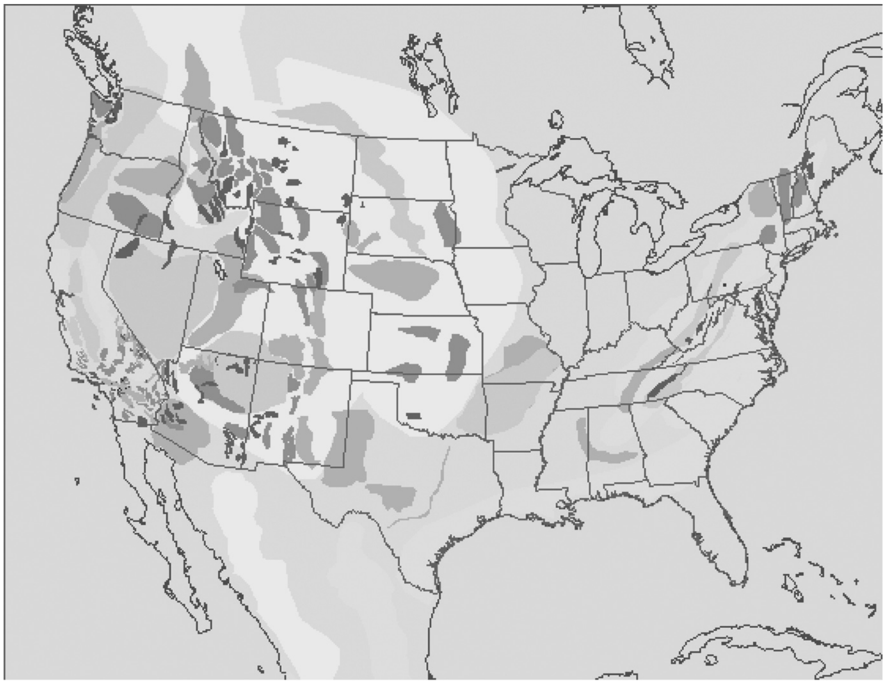


Figure 7. GIS representation of named physiographic features as polygons for a portion of North America. These GIS data do not represent a complete inventory of physiographic features in the area mapped. (see page 79 for color version)

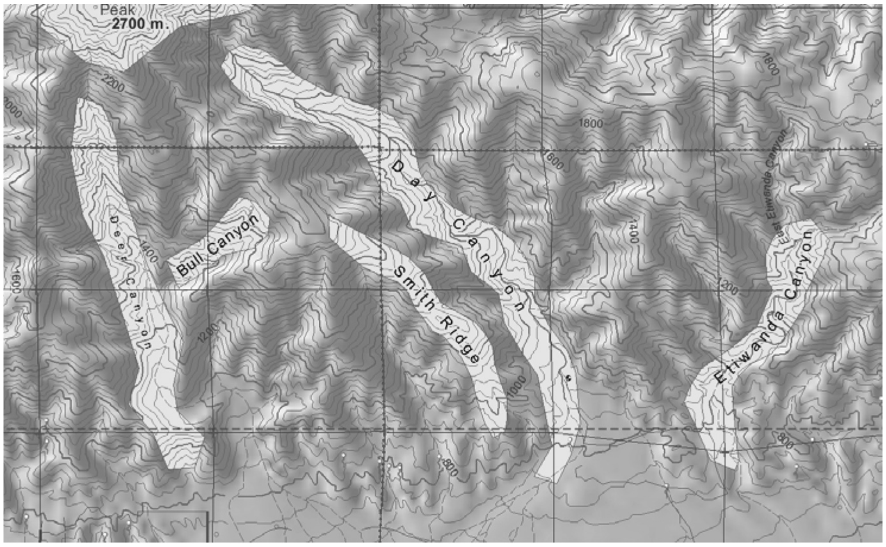


Figure 8. A portion of a 1:100,000 scale topographic map showing the GIS representation (as polygons) and the cartographic representation (as type) of several canyons in Southern California. (see page 79 for color version)

are in the dataset but they are too small to be seen at the scale mapped in Figure 7. In Figure 8, notice that the canyon polygons fit the terrain such that they do not extend too far up the side slopes and they extend far enough up the canyons that if only the upper portion of the feature appears on a map it will be still be represented.

Named Islands and Island Chains and Groups

"For islands that are too small to contain most of the name, the type should be placed as if the island were a point."

The GIS representation of both islands and island chains and groups is a polygon. The polygon for island chains and groups should fully encompass all islands in the chain or group and allow some extra space for the type that should be placed. The island chain or group name should be placed inside the polygon if possible, or it can overrun the polygon just slightly. Additionally, the type for each island should be placed within its polygonal outline if possible, but the type may be allowed to overrun the outline if there is lack of sufficient space. For islands that are too small to contain most of the name, the type should be placed as if the island were a point.

The GIS representation of islands and island chains and groups is a polygon with the following attributes:

- **PolyID** [Integer]: This is used as a primary or foreign key in a join or relationship to other databases, such as a names database or a revision information table.
- **Name** [Text]: This is the proper name of the feature as it would appear on the map.
- **TerritoryOf** [Text]: This is the proper name of the country with sovereignty over the island or island chain or group that would appear on the map.
- **FeatType** [Integer]: This is used to identify which features will be labeled as islands and which features will be labeled as chains or groups.
- **InGroup** [Short Integer or Boolean]: This is used to denote whether a feature is part of a group of features also represented in the GIS

database. For example in Figure 9, San Clemente Island is in an island group called the "Channel Islands".

- **GroupID** [Integer]: This is the ID of the group polygon.
- **Sources** [Integer]: Ideally this is used as a primary or foreign key for a join to a look-up table of sources. Unless an agency has performed the primary research to determine the name of a feature, this field can be used to identify at least three independent sources demonstrating that the feature name exists in the public domain.

Figure 9 shows the GIS representation and the associated type style and placement for an island group that contains a number of islands. Both the islands and the island group are stored as polygons in the GIS dataset, but the polygon for the island group would not be displayed on the map; it is shown here to illustrate the extent of the polygon in the GIS dataset.

RECLASSIFICATION OF FEATURES BASED ON SHAPE TYPE AND AREA

A reclassification of the features based on their geometry (i.e., shape type and size) can be used to further refine the placement of the type on the map. When labeling a map using GIS label placement algorithms, it does not matter if a polygon is a canyon or a bluff if the same text specifications (that is, size, color, kerning, curved or horizontal placement, etc.) are being used for those polygon labels. The subdivision of features into categories, such as canyon or bluff, which are not helpful when labeling is pointless as well as unproductive. Instead, the categorization of features into label classes should be dependent on two things: (1) any variation in the type specifications and (2) any variation in the label placement specifications. If the labels will look different or be placed differently, then they need to be in different label classes so they can be handled differently by the GIS label engine. Often times, type placement variations are dictated by the size

"The subdivision of features into categories, such as canyon or bluff, which are not helpful when labeling is pointless as well as unproductive."

Figure 9. GIS representation of islands in an island group. In the cartographic representation, the islands are labeled using point placement rules, and the island group is labeled as a polygon.

and shape of the feature; therefore, it would be useful to have a method for classifying the features based on these properties. This classification then becomes the basis for the creation of the various label classes that can be used to specify the style and placement of the type. Examples of some features in different shape classes are shown in Figure 10.

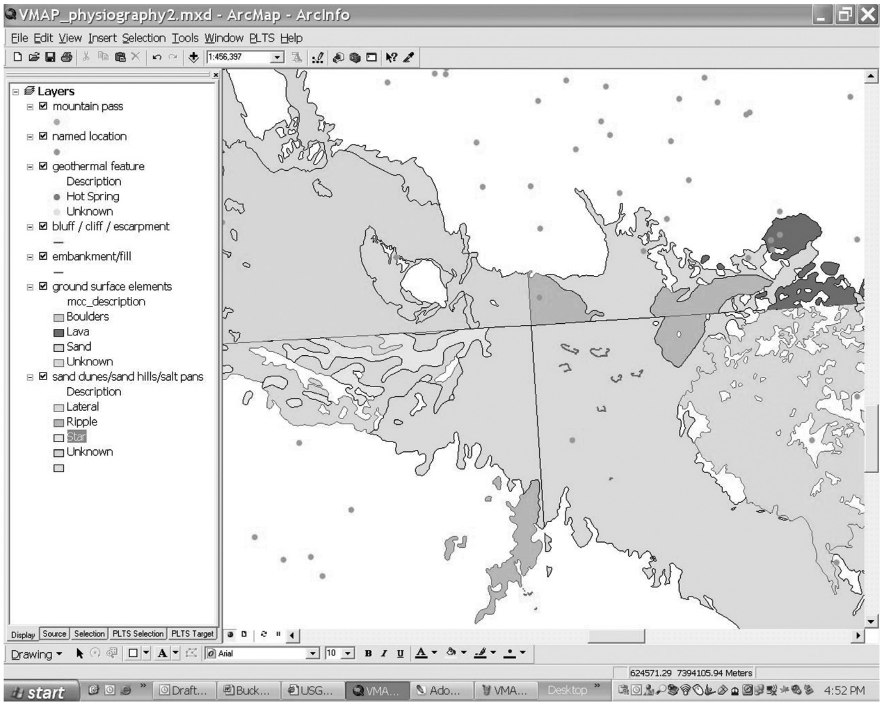


Figure 10. Features can be classified based on shape type which allows label classes to be created that are used to specify how the type is to be placed. (see page 80 for color version)

The classification of features by label placement properties is new to this research and bears further elucidation. First, the label placement capabilities of the software will drive the assignment of features to label classes. If the software is able to provide more refined placement options based on such properties as the size and shape of the feature, then these attributes can be used for label classification. To that end, a set of seven feature types that are used to specify label classes and variations in label placement properties were developed. These are:

1. Long
2. Long and Skinny
3. Oblong
4. Round
5. Snaky or Pronged
6. Splotch
7. Snaky or Pronged and Skinny

"The classification of features by label placement properties is new to this research . . ."

A cartographer might choose to specify different classes based on the requirements for their maps, keeping in mind that the assignment of features to different label classes is a function of the placement rules and type specifications, which are in turn a function of the map being produced. These seven classes, however, seem to work well for a variety of maps at different scales.

It would be extremely tedious to sort these features into classes by interactively selecting them and calculating a ShapeType attribute to store the unique class numbers that are then used to separate the features into label classes. Instead, it would be useful to have an automated method for performing this sort. This can be done using the minimum-bounding rectangle (MBR) which is then used to determine the proportion of the shape area to the MBR area (Figure 11). A better solution is to use the Rotated MBR, or RMBR (Brinkhoff *et al*., 1993), which allows rotation to align the bounding rectangle to the axes of the feature. The RMBR area is then used to calculate: (1) the Shape_Area/RMBR_Area proportion as a percent, and (2) the ratio of RMBR length.

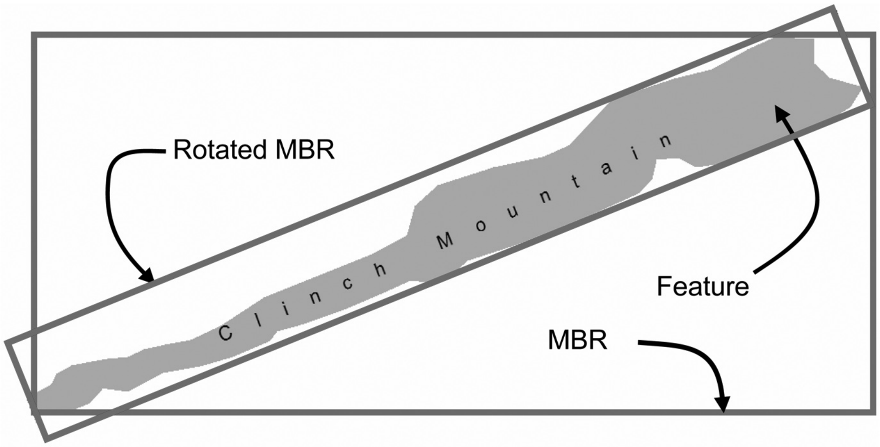


Figure 11. The traditionally calculated MBR versus the rotated MBR which is used to calculate two metrics used in the classification of features into label classes.

The logic for using these values is described in detail in an article by Frye (2006) and is not elucidated further here. The result, however, is that each feature can then be assigned to a label class using an attribute that designates the basic shape of the feature. Combining that with any required variations in the text specifications creates the full set of label classes used by the software to properly label the features. This helps to overcome deficiencies of the software to place text with greater variations for placement and style. As a result, the time required to edit the automated type placement can be greatly reduced.

LINKING TO STANDARDIZED GEOGRAPHIC NAMES DATABASES

An additional step might be to link the geographic features to a standardized names database. The advantage of doing this is to be able to manage the names of places in one location. For databases that will be used to create a number of map products, using a centralized standardized database makes even more sense as any updating of names in the database would be reflected across all maps products.

In some cases, it may be desirable or necessary to defer the decision about which names to use on maps to a geographic names authority that has such a database. For example, the Getty Thesaurus of Geographic Names (TGN, 2004) includes names and associated information about places. Places in TGN include administrative political entities (e.g., cities and nations) and physical features (e.g., mountains and rivers). Both current and historical places are included. The position of a place is indicated by geographic coordinates, and bounding coordinates and elevation may also be included. As another example, the GEOnet Names Server (GNS) provides access to the NGA's and the U.S. Board on Geographic Names' (US BGN) database of foreign geographic feature names or, for names in the U.S. and Antarctica, the GNIS. The utility of these datasets can be

"For databases that will be used to create a number of map products, using a centralized standardized database makes even more sense as any updating of names in the database would be reflected across all maps products."

limited, though, as the point location or bounding box position of the place name may not coincide adequately with the geographic feature in the GIS database. If such a names database does exist, and the challenges in linking to a geographic feature dataset can be overcome, then it might be advantageous to use this approach for name management.

In any case, the use of a centralized names database assures that the names are maintained in one location, and that the contents of that database can be reflected in all the maps that an organization creates. The amount of time to render the labels would be impacted by the size of that database that is related to the features. To speed up processing time, the geographic names fields that will be used for labeling could be permanently joined to the feature attribute table in a product version of the GIS dataset that contains the features geometries. Additionally, the names database could contain much more information about the place names such as historical, vernacular or variant names. The identification of sources (as with the physiographic feature polygons) could be circumvented by using a standardized geographic names database, although it might still be desirable to indicate the sources used to determine the geographic extent of the feature to be used for type placement.

CONCLUSIONS

This paper described three different types of natural features with indeterminate boundaries (named marine water bodies, named physiographic features, and islands and island chains and groups) and how they are modeled in a GIS database to support cartographic representation. Each of these three general types of features differs semantically and topologically. For all but summits, polygons with cartographic attributes were used as the GIS representation of the features. This framework for representing features with indeterminate boundaries worked very well to demarcate an adequate extent for the type on the map.

Type placement algorithms in GIS were insufficient for final type placement in the earlier stages of this research, and type had to be hand edited in order to sufficiently and elegantly imply the locations of the features. These shortcomings of type handling in GIS are well known by cartographers. Nonetheless, the automated type placement algorithms saved a great deal of time by placing the type so that only minor adjustments were required. An enhancement that continues to be explored to address this problem is the automated division of feature types into different label classes based on their shape type and area properties. These properties are then used to more carefully specify the style and placement of the type through the use of label classes. This enhancement reduces the amount of time and effort required for type placement and style changes.

"Modeling GIS data with the appropriate geometry and attributes for cartography will help make the data more versatile for multiple map purposes . . ."

Also important in the consideration of type style were the cartographic attributes SizeClass for named marine water bodies and Order for named physiographic feature polygons. These attributes allowed the GIS to vary the type size relative to the map scale, and only the relative type size variations needed to be referenced instead of the specific type sizes (e.g., smaller and larger font versus 10 and 12 point font). Other attributes such as polygon area are not always sufficient to make this determination, as features of the same size may vary in cultural and geographic significance. Assignment of this attribute using polygon area for the named marine water bodies was done through automation, and the attribute for named physiographic features was manually assigned.

In this paper, one assumption was that a primary purpose of a GIS database is to make maps. Modeling GIS data with the appropriate geometry and attributes for cartography will help make the data more versatile for multiple map purposes, and with this increased utility, cartographic GIS data may become more prevalent. The development of appropriate models for cartographic GIS data requires that cartographers become more involved in the design of GIS databases by contributing their knowledge and expertise. Since cartographers know that the type for such features as physiographic and marine places should be on many maps, it would benefit both the cartographers and others who need this information to have the data in their GIS databases. As a result, the GIS data will be more cartographically rich and the GIS software can be used to aid in the sometimes tedious and time consuming task of placing all that type on the maps.

"As a result, the GIS data will be more cartographically rich . . ."

REFERENCES

ADLP *(Alexandria Digital Library Project)*, "Correlation data relating words/phrases in place names to associated classification terms from three typing schemes", June 4, 2004, <http://www.alexandria.ucsb.edu/gazetteer/>, (June 7, 2004).

Arctur, D. and Zeiler, M., 2004. *Designing Geodatabases: Case Studies in GIS Data Modeling.* Redlands, California: ESRI Press, 393 pp.

Brändli, M., 1996. Hierarchical Models for the Definition and Extraction of Terrain Features. In Burrough, P.A. and Frank, A.U. (Eds.) *Geographic Objects with Indeterminate Boundaries*, 257-269. London: Taylor & Francis.

Brinkhoff, T., Kriegel, H-P. and Schneider, R., 1993. Comparison of Approximations of Complex Objects Used for Approximation-based Query Processing in Spatial Database Systems. *Proceedings of the 9th International Conference on Data Engineering*, Vienna, Austria, 40-49.

Buckley, A. R., Frye, C., Buttenfield, B. and Hultgren, T., 2005. An Information Model for Maps: Towards Cartographic Production from GIS Databases. *Proceedings of AutoCarto 2005*, March 18-23, Las Vegas, Nevada.

Couclelis, H., 1996. Towards an Operational Typology of Geographic Entities with Ill-defined Boundaries. In Burrough, P.A. and Frank, A.U. (Eds.) *Geographic Objects with Indeterminate Boundaries*, 45-55. London: Taylor & Francis.

DIGEST, *Digital Geographic Information Exchange Standard*, 2001, <http://www.digest.org/>, (May 26, 2006).

Fenneman, N.M., 1938. *Physiography of the Eastern United States.* New York: McGraw-Hill Book Company, 691 pp.

Fenneman, N.M. and Johnson, D.W., 1946. *Physical Subdivisions of the United States* (Map): U.S. Geological Survey, 1:7,000,000, 1 sheet.

Frye, C., 2006. Shape Types for Labeling Polygon Features with Maplex. *Cartographic Perspectives* 54:69-73.

TGN (Getty Thesaurus of Geographic Names), "About the TGN", July 13, 2004, http://www.getty.edu/research/conducting_research/vocabularies/tgn/about.html>, (June 9, 2006).

Jenness, J., "Topographic Position Index (TPI)", March 22, 2006, <http://www.jennessent.com/arcview/tpi.htm>, (June 9, 2006).

Lobeck, A.K., 1932. *Atlas of American Geology*. New York: The Geographical Press, Columbia University, New York, 92 pp.

Lobeck, A.K., 1947. *Historical Geology of the United States.* New York: The Geographical Press, Columbia University.

Mark, D.M. and Turk, A.G., 2003. Ethnophysiography Workshop on Spatial and Geographic Ontologies, September 9, prior to COSIT03, Ittingen, Switzerland.

Raisz, E., "Landforms of California and Nevada: 1:500,000", Map published 1965, November 10, 2004, http://www.panorama-map.com/Raisz/Calif.%20Nevada/Raiszcalif500.html, (June 9, 2006).

Raisz, E. and Atwood, W.W., 1957. *Landforms of the United States*, 6th Ed. Map, Cambridge, MA: E. Raisz.

Smith, B., 1994. Fiat Objects. In Guarino, N., Vieu, L. and Pribbenow, S. (Eds.), *Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology, 11th European Conference on Artificial Intelligence, Amsterdam, 8 August 1994*, 15-23. Amsterdam: European Coordinating Committee for Artificial Intelligence.

Smith, B., 1995. On Drawing Lines on a Map. In Frank, A. and Kuhn,W. (Eds.) *Spatial Information Theory –A Theoretical Basis for GIS*, International Conference COSIT '95, Vol. 988, Lecture Notes in Computer Science, 475-484. Berlin: Springer-Verlag.

Smith B. and Mark, D.M., 2003. Do Mountains Exist? Towards an Ontology of Landforms. *Environment & Planning b: Planning and Design,* 30:3:411-427.

Smith, B., 1996. "Mereotopology: A Theory of Parts and Boundaries", *Data and Knowledge Engineering* 20: 287-303, also <http://ontology.buffalo.edu/smith/articles/Mereotopology1.pdf>, (June 9,2006).

USGS (U.S. Geological Survey). 1992. "Physiographic Divisions of the Conterminous U.S.", January 3, 2002, <http://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml>, (June 9, 2006).

USGS (U.S. Geological Survey). *USGS Topographic Maps Illustrating Physiographic Features*, August 26, 2002, <http://rockyweb.cr.usgs.gov/outreach/featureindex.html>, (June 7, 2006).

USGS (U.S. Geological Survey). *National Mapping Program Standards*, December 2, 2005, <http://nationalmap.gov/gio/standards/>, (June 7, 2006).

Varzi, A. C., 1997. Boundaries, Continuity, and Contact, *Nos* 31, 26--58.

Waters T. and Evans, A.J., 2003. Tools for Web-based GIS Mapping of a "Fuzzy" Vernacular Geography. *Proceedings of the 7th International Conference on GeoComputation*, University of Southampton, United Kingdom, 8-10 September.

Wood, J., *Landserf*, June 11, 2005, <http://www.soi.city.ac.uk/~jwo/landserf/>, (June 9, 2006).

Wood, J., 2004. A New Method for the Identification of Peaks and Summits in Surface Models", *Proceedings of the GIScience 2004 Conference*, University of Maryland University College, Adelphia, Maryland, 20-23 October, 227-230.