

# A DATA MODEL FOR THE REPRESENTATION OF THE NATURAL LANDSCAPE

Charlie Frye and Aileen Buckley

Cartographic Research and Special Projects  
Environmental Systems Research Institute, Inc.

[cfrye@esri.com](mailto:cfrye@esri.com)      © 1994 ESRI

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 mere graphic text that is not georeferenced to the data used to make the map. In this paper, we demonstrate a GIS data model for physiographic features and by extension named water bodies. We discuss the semantic model (what features to include), the representation (how to define the geometry of the features and their attributes), the symbology (the specifications for both text properties and label placement), and the rules for data capture (how to identify the features from source materials as well as the QC/QA process). We also discuss how the representations, symbology and data capture are sensitive to the software used for mapping.

There are a number of interesting issues relating to such a named features data model. First, it can be inherently multiscale and may be used to create many different maps. This requires that non-point features be captured as polygons and the text be dynamically placed within the extent of the polygon at the scale mapped. There are also interesting questions about the mereological relationships between features – that is, the relationships of parts to the whole. What features are included in a mountain range and when does a gulf become an ocean? Additionally, there is the more pragmatic question of how to capture these features for inclusion in the GIS database. Is it sufficient to digitize the apparent extent from existing maps, can local knowledge be used to define more precisely the extent or can computational methods be used to generate features such as hills, valleys, canyons, and ridges using DEMs and other digital source data?

We review the issues above in hopes that others will be better able to use GIS for making maps that include these features. Without the addition of these names on our maps, the graphics we create will be the poor cousins of the more informationally- and cartographically-rich maps of the past.

**Keywords:** GIS, cartographic representation, cartographic data modeling, physiographic features

## INTRODUCTION

Maps produced by traditional cartographic organizations, both public and private, distinguish themselves from maps produced by organizations using solely GIS methods. The presence of names for natural features is one obvious way traditionally-produced maps can be differentiated from their GIS-based counterparts. These features are often represented by their names alone and not by any distinct boundary that is delineated on the map. On reference maps, this includes names of marine water bodies, such as bays, straights and gulfs, as well as terrestrial physiographic features, such as mountain ranges, deserts, and ridges.

The names for such features that appear on maps are currently preserved in forms not readily applicable to other uses. That is, they are often stored as a text layer for a given finished map product, 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 format. Increasingly, geographic names are stored in a digital database with a point location that sometimes does not correspond to either their actual geographic position or the best cartographic position for text placement. By and large, they are certainly not at this time linked to mainstream spatial information infrastructures.

To further complicate matters, the way the names of natural features without distinct boundaries are drawn on maps is highly tailored to the shape and nature of the feature so as to imply its extent without having to draw a debatable demarcating line on a map. It would be valuable to have a flexible lowest common denominator 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 label for the feature would appear. Digitized correctly, this polygon would be able to encompass the appropriate location for the label at any scale and at the smallest scales could be converted to a point for label 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.

In order to discuss a practical means of doing this, we first need an understanding of the types of natural named features without boundaries that appear on maps (that is, the kinds of features that are included in this category). Then we need 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. This leads to a discussion of the workflow for capturing and mapping these features using GIS.

## 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: all of the names in the database are associated with one or more feature type terms from the 1) ADL Feature Type Thesaurus (Alexandria Digital Library Project 2004) which are drawn from a variety of authoritative sources, including glossaries and government publications, and portions of the names are associated with either the 2) gazetteer type terms from the U.S. Geological Survey or from the 3) U.S. National Geospatial-Intelligence Agency (formerly NIMA). These feature types have specific descriptions and are polyhierarchical (see Appendix A); however, they are simply descriptions with no associated geographic features 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 either points, lines or polygons (see Appendix B). 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 do not even have a name field). It is intended that the names of the feature be derived from the GEOnet Names Server (GNS) for all location except the United States and Antarctica which come from Geographic Names Information System (GNIS) database.

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 (1:1,000,000); this is a slightly more detailed reiteration of the DCW. Level 1 data is equivalent to a medium scale 1:250,000 resolution (horizontal

accuracy: 125-500m). Level 2 data is equivalent to a large scale 1:50,000 resolution (horizontal accuracy: 125-500m). Because of the varying scales, a feature that is represented as a polygon at one scale can become a point at smaller scales; for example, buildings polygons will change to building points and builtup area polygons are represented as town points at smaller scales. The varying feature representations and retaining the names linkages makes using these data for scale-less zoom-able Internet maps problematic.

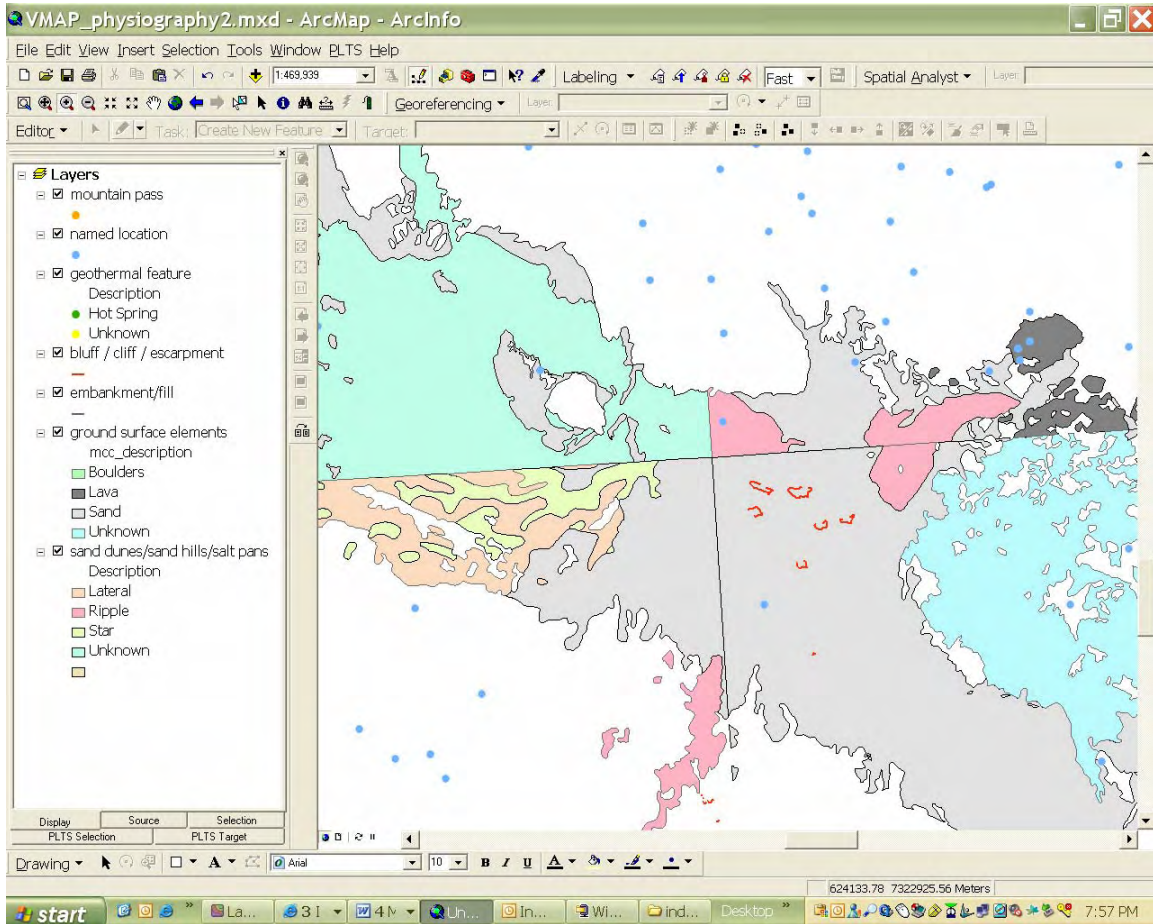


Figure 1. Physiographic features in the VMAP data are designated as points, lines or polygons.

For our research, we determined that we could store features as either points (in a very few cases) or polygons. The features are only used to label the map, and the point and polygon designations we formulated are further categorized so that they can be used to specify particular placement properties for the feature types. These distinctions are clarified further in the “Reclassification of Feature Types” section below.

The left column of Table 1 lists some of the most common types of natural features with without delineated boundaries on maps. The second column 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 and mereological (i.e., the part-whole relation) model for these features in the real world as defined by our languages and cultures. The last column shows whether the information shown on the map is stored as binary (i.e., whether or not the feature exists at a particular location) or predominant (i.e., other kinds of things exist, but they are not sufficient to be mapped.)

<b>TYPES OF FEATURES WITH INDETERMINATE BOUNDARIES</b>	<b>GIS THEME</b>	<b>TOPOLOGY</b>	<b>DATABASE REPRESENTATION</b>
Named Marine Water Bodies	Hydrography	No Overlaps	Text: Binary
Named Physiographic Features	Physiography	May Nest or Have Partial Overlaps	Text: Binary
Islands and Island Chains	Physiography and Hydrography	Nested, no Overlaps	Text Binary
Neighborhoods and Districts / Vernacular Regions	Cultural and Transportation	May Nest or Have Partial Overlaps	Text: Binary
Land Cover, Geology, Soils and Other Surface Overlays	Surface Overlays	No Overlaps	Text and Symbol: Predominant

Table 1. Thematic organization of features with indeterminate boundaries

Many features with proper names exist within these themes, and they often have indeterminate boundaries. For the most part, only 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 labels have 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.

Although it would be possible to extend this discussion to include neighborhoods and districts/vernacular regions as well as surface overlays such as geology and soils, we limit further discussion to natural landscape features on maps, including marine water bodies, physiographic features and islands and island chains. Discussion of these physical features serves to illustrate how our approach may be used for other features with indeterminate boundaries.

### *Named Marine Water Bodies*

The general rule for the topology of named marine water bodies is that they do not overlap. For instance, the Straits of Florida do not overlap the Gulf of Mexico or the Atlantic Ocean. It may also 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). An attribute in a GIS database should be used to define the feature type for marine water bodies. The valid values for that attribute may include historical, relic, and legendary types of water bodies.

Since marine water bodies are typically represented with text, 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 text is displayed so that the largest water bodies have the largest text sizes. Text 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 text will not fit roughly into or just beyond the accepted bounds of the feature, it should not be shown on the map. If the water body is too small to contain its text, 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.

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

- Bay
- Bight
- Channel
- Firth
- Gulf
- Inlet
- Ocean
- Passage
- Sea
- Sound
- Strait
- Infamous Region
- Antiquated Name

An example of an Infamous Region is the Bermuda Triangle, and a feature with an Antiquated Name might include the Sargasso Sea. We developed this set of feature types 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, these factors should be kept in mind and the model can be derived from any of those that currently exist (the ones presented here, ADL, DIGEST, or others).

### *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 and 1946, and Raisz 1957, among others) 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. These regionalizations are used as the basis for further definition of smaller physiographic features. Currently, these features are typically not part of GIS or spatial data repositories. In fact, the digital representation of these features would normally be managed in place names indexes with no associated geometric representation useful for cartography.

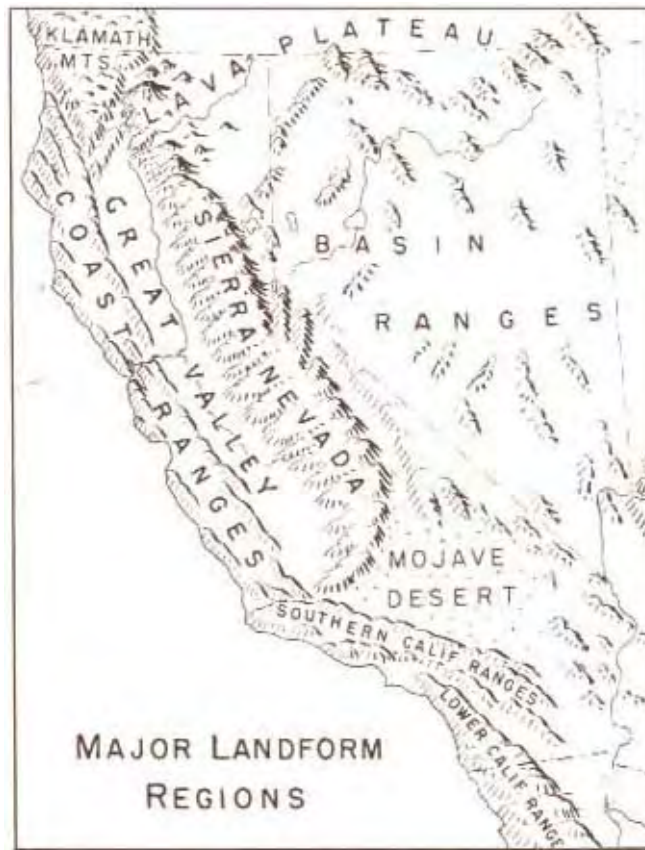


Figure 2. Small-scale representation of named physiographic features and regions (Raisz 1965).



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

Data for larger scale reference maps, however, is even more critically lacking, underscoring the need for a discussion that pertains to features such as peaks, ranges, mountains, valleys, deserts, canyons, flats or playas, passes, etc. For smaller scales, there is a United States Geological Survey map and related dataset (USGS 1992) of named physiographic *Divisions, Provinces, and Sections* that regionalizes the United States based on a 1946 map by Fenneman and Johnson (Figure 4).

The general rule for the topology of named physiographic features is that they may or may not overlap or they may partially overlap. Unlike named marine water bodies, 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 meriological (that is, part to whole) 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 along with peaks. The impact of cultural and linguistic history is not necessarily logical when it comes to the names of these types of features. Another complication is that, unlike named marine water bodies, which we suggest can be well-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 – the name of the feature does not always clarify the distinction.

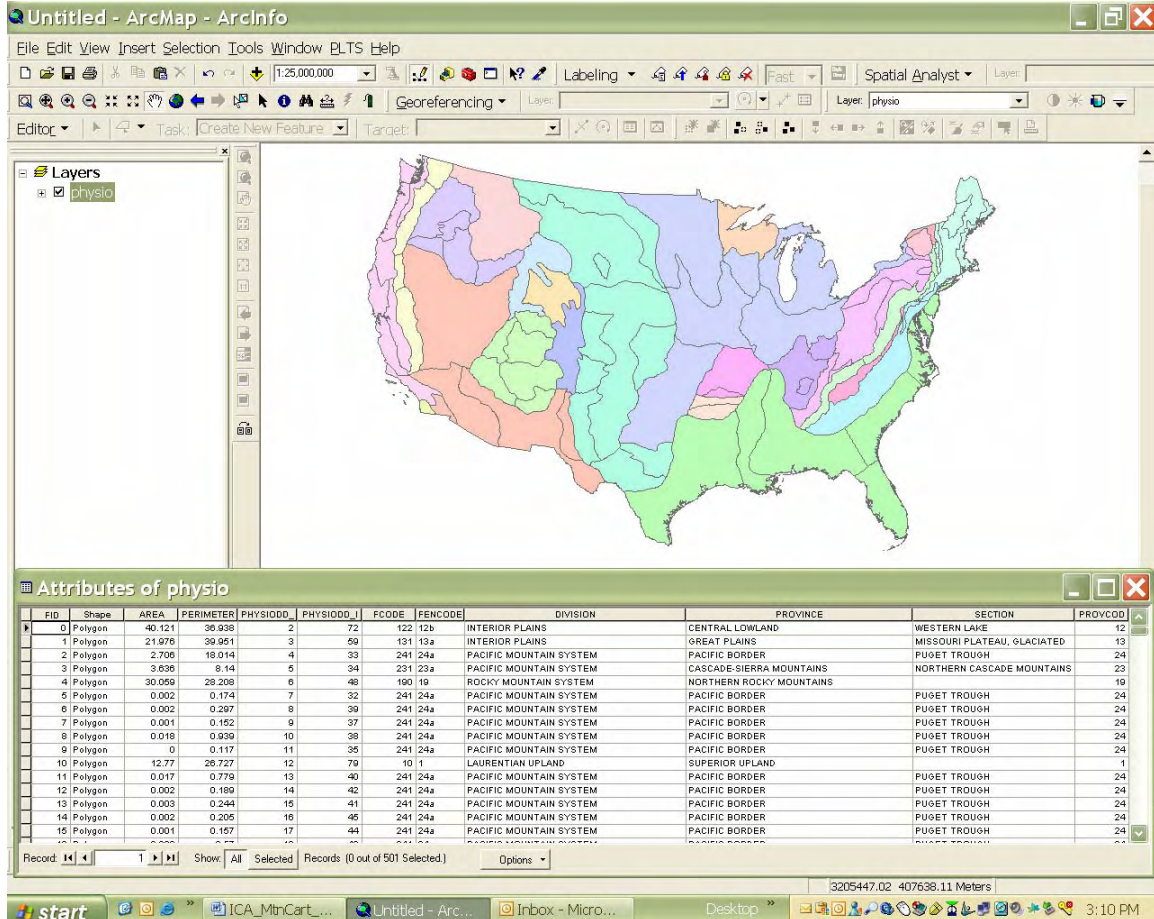


Figure 4. USGS dataset of physiographic regions and provinces for the conterminous United States.

A suggested set of feature types for named physiographic features is:

- 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

There are a few of these feature types that warrant further description. As mentioned above, 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 those exposed named land masses within inland water bodies – islands within named marine water bodies are discussed below.

Named physiographic features, like named marine water bodies, are typically represented on the map with text, unless the feature is too small at a given map scale, in which case it should not appear at all. Unlike named marine water bodies, the names of these features are not aligned to the graticule. Rather, text 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 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 text.

### *Islands and Island Chains*

The mereological relationships for islands and island chains seem pretty straight forward at first. Islands may or may not be part of an island chain. Island chains, however, are represented with text that is by necessity drawn outside the geometric bounds of the individual island features.

A suggested set of feature types for islands and island chains is:

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

Islands and island chains are generally named with these terms as part of their proper name, and there are only a small set of synonyms, like archipelago and chain. This makes the semantic model for and the identification of these features fairly simple and straightforward.

In the cartographic representation of islands, the extent of terra firma is generally neither fuzzy nor indeterminate. The text that labels an island is placed inside the island if it fits

at the given map scale. If an island is too small to contain its text, the text is placed outside the island using methods for positioning text 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 are more complicated, as the text naming the chain is somewhat similar cartographically to the text for a mountain range. Ideally it extends over the extent of geographic space for all islands shown that are officially part of the chain. The text for an island chain may overstrike islands, that is, it does not have to be positioned without exception outside of the geometry of the individual islands. Typically the text for an island chain follows the major trending axis of the chain. If a group of islands is labeled, the style is more like the naming of larger named marine water bodies -- the label is aligned to the graticule and is positioned within the group, large enough to imply virtual ownership of the islands by the group.

## **GIS REPRESENTATIONS FOR CARTOGRAPHIC USE**

In this section, we discuss 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, one 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 representations 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. In the following sections we discuss how specific types of features with fuzzy or indeterminate boundaries should be modeled in the GIS so 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.

### *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]: For the most part, this is informational, but in cases where the semantics of the name do not match the actual type of feature, this attribute stores the actual feature type.
- **SizeClass** [Integer]: This is based on the area of the polygon and is used for two purposes. The first is to determine whether the feature will be represented on a map at a particular scale, and the second is to determine the size of the text 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.

Figure 4 shows the GIS representation of named marine water bodies relative to land masses around Central America. Notice the lines that divide the water into separate and unique named bodies of water. Figure 5 shows the cartographic representation, which does not indicate the feature outlines but does display text that is appropriately sized and positioned within each water body relative to the chosen map scale.



Figure 4. GIS representation of named marine water bodies as non-overlapping polygons with an extent adequate to label the features at multiple scales.

The symbolization and placement of the text is performed by the GIS software with the polygons as the basis for the text style and placement rules. The polygon attributes are used to determine the content of the text string, how to display it, and whether to include it at a given map scale.



Figure 5. 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 the text 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 text for the feature, which only serves a very narrow range of purposes (i.e., maps).

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 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 at 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 name 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]: As above, this is 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.

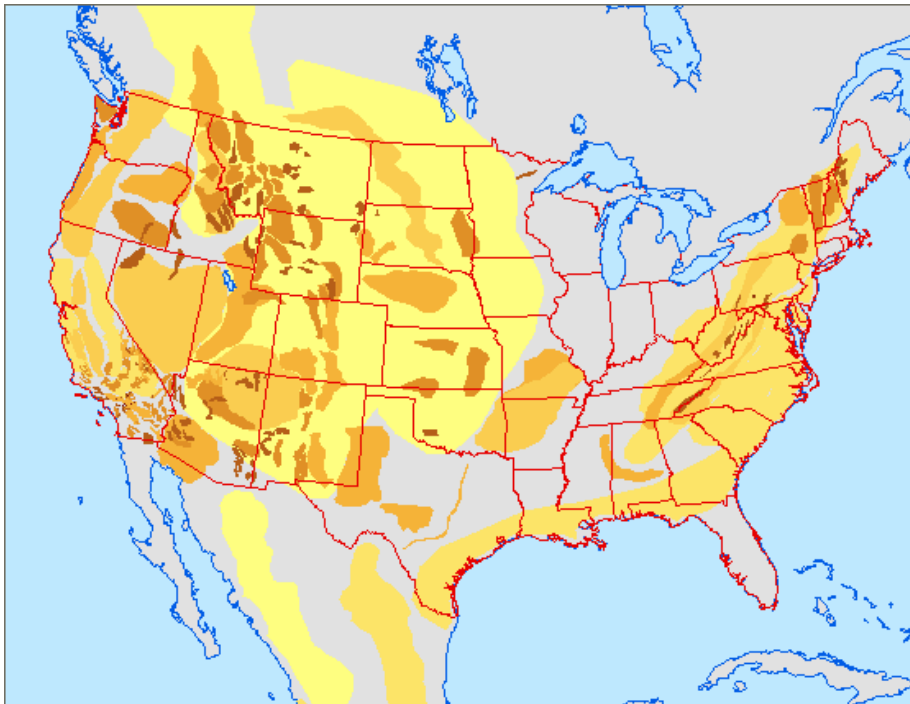


Figure 6: GIS representation of named physiographic features as polygons for a portion of North America. This image is not intended to represent a complete inventory of such features in the area mapped.

Figures 6 and 7 give an indication of how the GIS features for named physiographic features look. In Figure 6, notice that features for all potential map scales are represented; in fact, the canyons in Figure 7 are drawn but at a scale too small to be seen. In Figure 6, the limits of various canyon features stored in the GIS can be seen. The canyon polygons fit within 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.

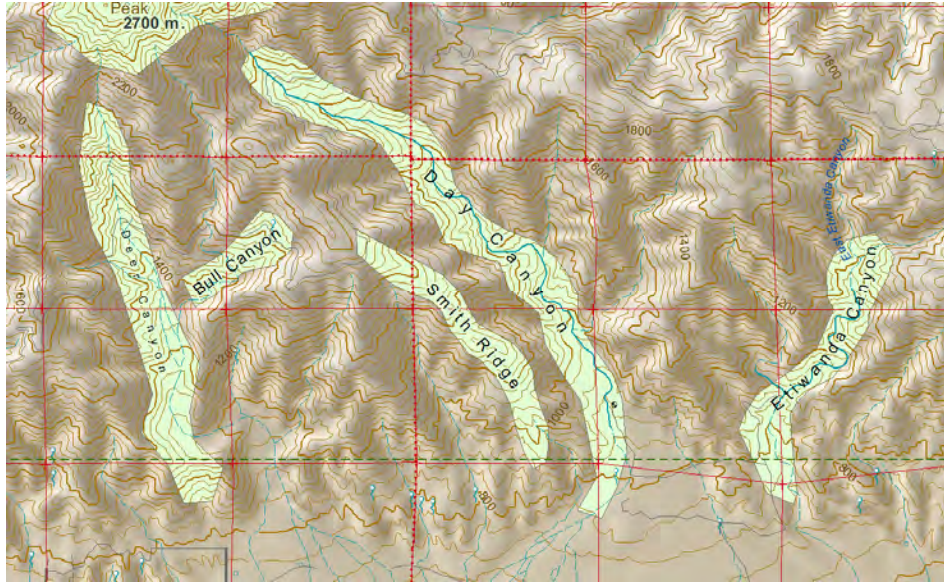


Figure 7: Portion of some of the layers on a 1:100,000 scale topographic map showing the GIS representation (as polygons) and the cartographic representation (as labels) of several canyons in Southern California.

### *Islands and Island Chains*

The GIS representation of both islands and island chains is a polygon. The polygon for island chains should fully encompass all islands in the chain and allow some extra space for the text that should be placed. The island chain name should be placed inside the polygon. Additionally, each island should be labeled within its polygonal outline, but the names may overrun the outline due to lack of space. For small islands that can not contain most of the name, the label should be placed as if the island were a point.

The GIS representation of islands and island chains is polygon GIS dataset 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 that would appear on the map.
- **FeatType** [Integer]: This is used identify which features will be labeled as islands and which features will be labeled as chains.
- **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 7, San Clemente Island is in a group called the “Channel Islands”.
- **GroupID** [Integer]: This is the ID of the group polygon.
- **Island\_ID** [Integer]: This is used as a primary or foreign key in a join or relationship to the named physiographic polygon features.
- **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 8: GIS representation of islands and island region represents the GIS polygon used to represent the Channel Islands chain as a polygon. In the cartographic representation, the islands are labeled with point placement rules, and the island chain is labeled as a polygon.

## RECLASSIFICATION OF FEATURE TYPES

After working with the various feature types from a number of geographic locations (Crater Lake, Oregon; Boise, Idaho; Southern California), we determined that we could reclassify the large number of physiographic feature types into essentially two cartographic feature types: points and polygons. When labeling a map using GIS, it does not matter if the polygon is a canyon or a bluff, if the same text specifications are being used for labeling (that is, the size, color, kerning, etc.) Therefore, the subdivision

of features into categories that 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: any variation in the type specifications and 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. Some examples are shown in Figure 9 along with the placement of the associated feature label. The feature shape types are described below.

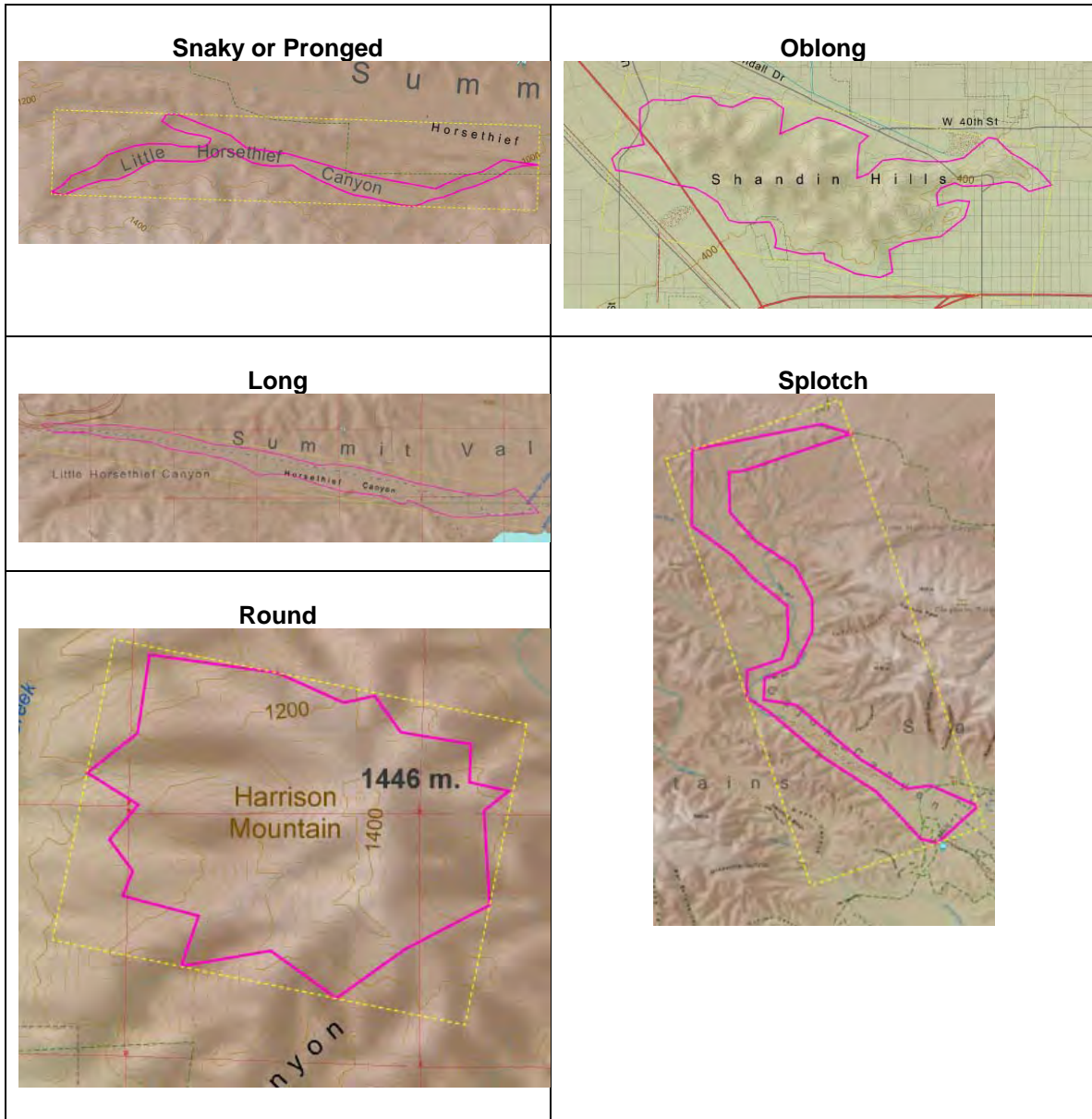


Figure 9. Feature types based on shape of the feature.

Type specifications are a function of the map that is being produced and are handled in the same way that have traditionally been. However, the classification of features by label placement properties is new to this research and bear further elucidation. First, the label placement capabilities of the software will drive the subdivision of features into label classes. If the software is able to provide more refined placement options based on such factors as the size and shape of the feature, then these attributes can be used for categorization. To that end, we developed a set of seven feature types that are used to specify label classes and variations in label placement properties (Figure 9). These are:

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

Each of these has different placement rules and the classification of GIS features into one of these categories is a function of the minimum bounding rectangle (MBR) which is used to calculate the Shape\_Area/MBR\_Area proportion as a as percent and the ratio of MBR length to width (Figure 10).

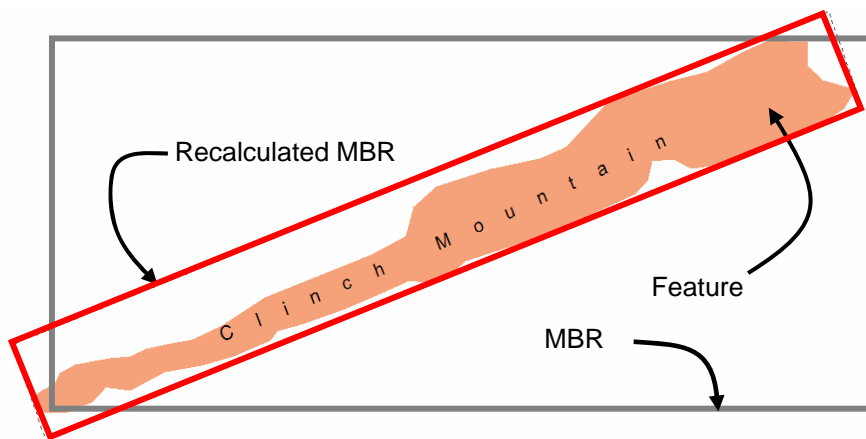


Figure 10. The traditionally calculated minimum bounding rectangle (MBR) versus a recalculated MBR used to classify the features into label classes.

The logic for using these values is then applied in the following calculate statement which assigns each feature to one of the seven label classes:

```

If RatioL2W < 4 and MBRArea > 60%
  Label Type = "Roundish"
Elseif RatioL2W < 8 and MBRArea > 25%
  LabelType = "Oblong"
Elseif RatioL2W >= 8 and MBRArea > 10%
  LabelType = "Long"
Elseif RatioL2W >= 8 and MBRArea <= 10%
  LabelType = "Long and Skinny"
Else
  If RatioL2W < 4 and MBRArea >= 20%
    Label type = "Splotch"

```

```

Elseif RatioL2W < 8 and MBRArea > 12%
    Label Type = "Snaky or Pronged"
Elseif RatioL2W < 8 and MBRArea <=12%
    Label Type = "Snakey or Pronged and Skinny"
    
```

The result is that each feature now has a label class attribute that designates the basic shape of the features. 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. An example of the properties for labeling the roundish and oblong features in Figure 9 above might look something like this:

<b>ROUNDISH</b>	<b>OBLONG</b>
Length to width ratio = 2 MBR = 60% ArcGIS Maplex labeling placement rules: Placement: Curved Try Horizontal First = true May Stack = true Character Spacing = up to 200%	Length to width ratio = 4 MBR = 43% ArcGIS Maplex labeling placement rules: Placement: Curved May Overrun by 36 pts Character Spacing = up to 200%

*Logic for Assigning Size Classes*

In addition, each feature should also have a LabelSize attribute. This attribute is used to set the size classes for the labels and is especially useful if the data set will be used to either produce a maps at number of scales or the create maps that are scale-less such as those used in an Internet application with the ability to zoom. There are two ways that the size classes can be assigned. In the first, an equal interval method is used to divide the range of Shape\_Area values evenly by the number of size classes desired. This works well when there are only two or three size classes. The second is a binary regression method by which each class is divided by half the remaining range area between the smallest polygon and the largest polygon:

$$f(\frac{1}{2}x), \text{ where } x \text{ is the remaining area}$$

Essentially, the classes are determined by initially dividing the range in half and the upper portion becomes the first class (for the largest features). Then the lower portion is divided in half and its upper portion becomes the next class. This is repeated until the desired set of size classes is codified. This method is good for classifying the features into four or more size classes.

Examining the distribution of these shape types in terms of their frequency of occurrence in a variety of out datasets, we found that there many shapes in the roundish category, which require different labeling rules than the oblong and long shapes which are also well represented (Figure 11). Thus, the amount of effort, particularly manual editing of text, is roughly cut in half if this method is used to more appropriately label the features in the first place.

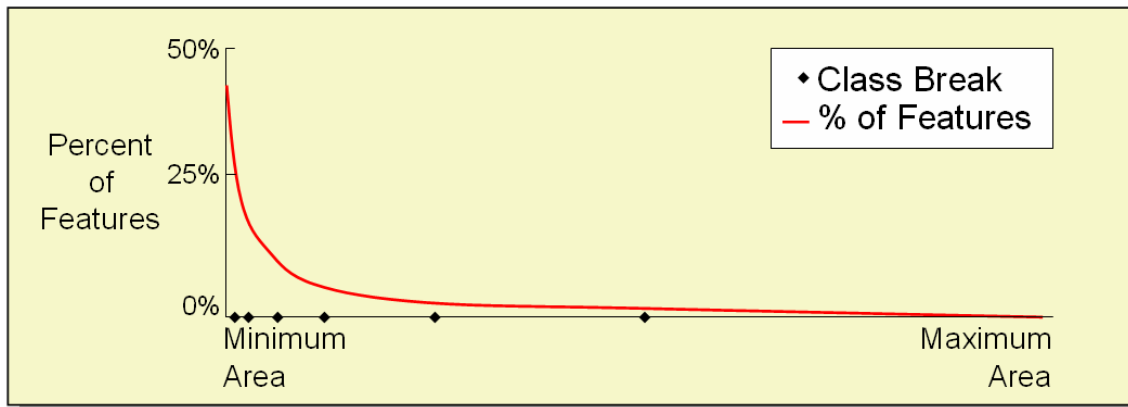


Figure 11. Feature counts for each shape type for five data sets: (1) hydrographic features for the northeastern United States for a 1:1,000,000 scale map, (2) physiographic features of North America, (3) soils for Ada County, Idaho, (4) hydrographic features for Ada County Idaho for 1:5,000 scale maps, and (5) vegetation for Ada County, Idaho for 1:5,000 scale maps.

## DELINEATING NAMED FEATURES OF THE PHYSICAL LANDSCAPE

Because these features with indeterminate boundaries and text representations on maps for the most part do not currently exist in a format that is readily usable by GIS, some thought must be given to how they can best be captured. Essentially, they can either be compiled “manually” by digitizing from existing map documents or they can be “mined” from digital data such as DEMs through automated feature extraction tools in GIS and other software packages. We have explored the feasibility of these techniques relative to the amount of time required, the accuracy of the representation relative to one or more map sources, and the quality of the GIS label placement for the feature.

### *Digitizing*

Digitizing offers the advantages of 1) ensuring the creation of features for all named locations on the map(s) and 2) being able to specify the exact location of the vertices making up the feature. However, digitizing is time consuming and tedious, interpretation of the location of the named places on maps is subjective, and human error can be introduced in the data capture process (*e.g.*, features may be missed). Nonetheless, digitizing ensures that the type and extent of the features on the map are captured, especially once one is trained to determine the location of such features. Additionally, digitizing allows variations in the landscape from place to place (*e.g.*, mountainous western U.S. versus the hilly western U.K.) to be taken into account when

delineating and attributing the named places. One other advantage of digitizing is that this methodology can also be used for other named places such as named water bodies or non-physiographic regionalizations. Automated feature extraction, explained further below, is dependent upon DEMs and cannot be used for hydrographic or cultural region detection.

In the digitizing process, source documents can be used along with ancillary data such as contour lines. The contours often correspond to appropriate edges of the features being captured (Figure 12), although it is often necessary to “jump” from contour to contour or additionally to digitize across contours (Figure 13).

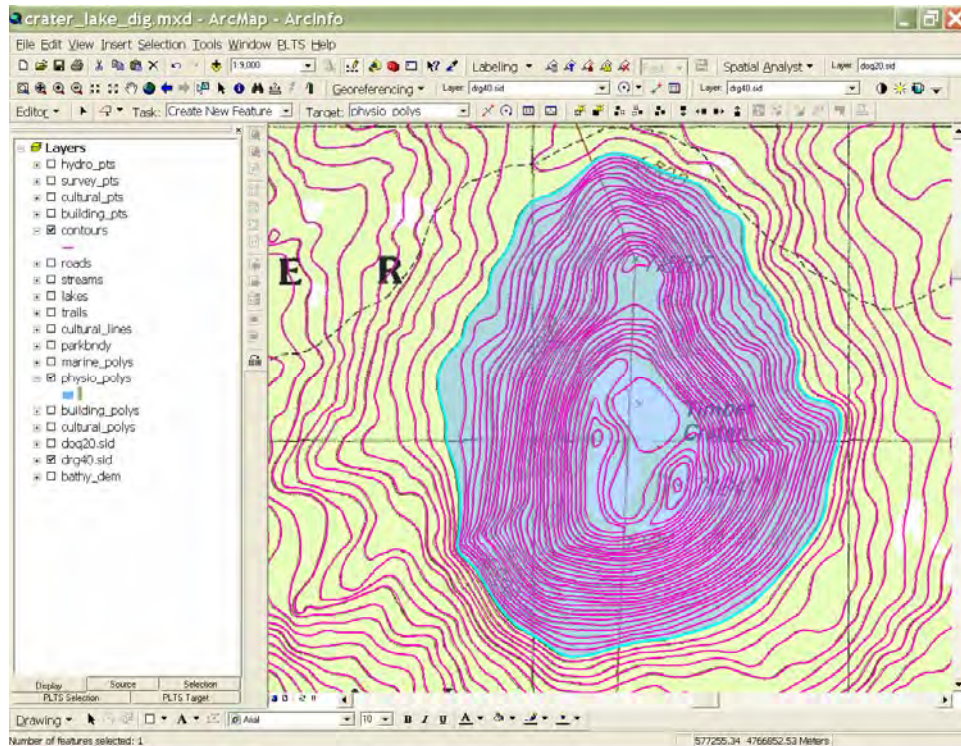


Figure 12. In this case, the extent of the volcanic cone feature coincides with a single contour line which can be traced or copied to create the physiographic feature.

To reiterate, the purpose of the delineation exercise is to capture a polygon within which the text could be shown at varying map scales, so the exact location of vertices is less important than the creation of an appropriate and minimally correct GIS polygonal representation. Using the contour lines can help to speed up the digitizing process, but the coincidence of the feature with the contours should not be construed as a addition of some level of accuracy.

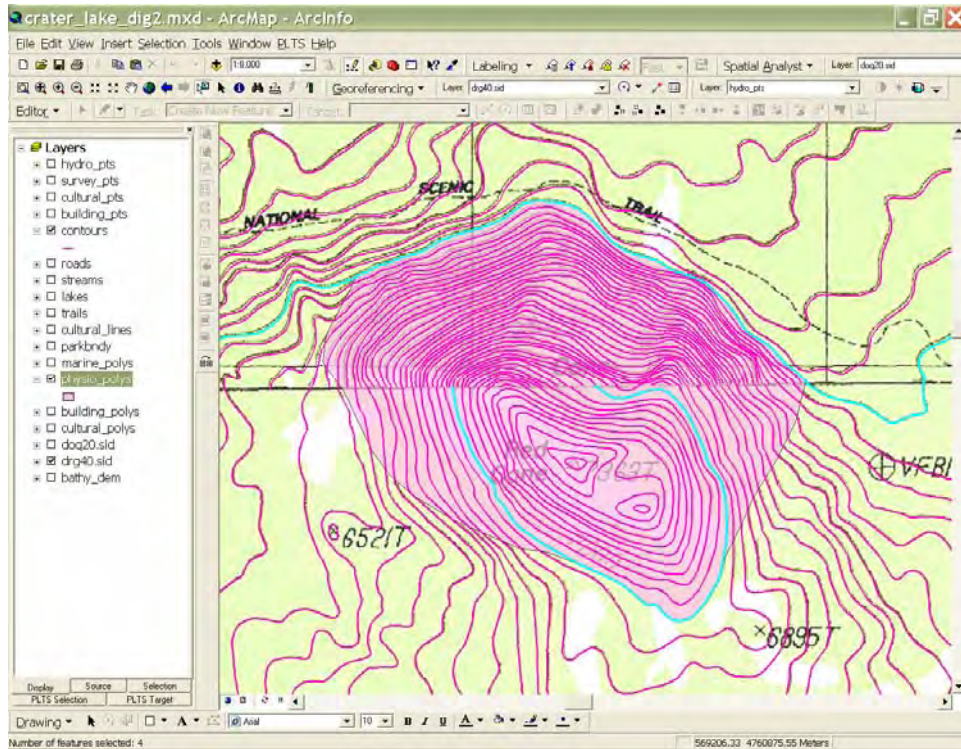


Figure 13. In this case, part of the extent of the volcanic cone feature can be traced from a contour line, but delineation then crosses contour lines, traces a different contour line, crosses contours again and then closes at the first contour that was traced.

### *Automated Feature Extraction*

A variety of software packages include Automated Feature Extraction (AFE) tools. Most of them start with input from a DEM and some have the ability to add ancillary data. In most cases, the DEM is used to create slope and some other value such as slope position or profile curvature which are then further analyzed to produce an estimate of the locations of various landform regions. One approach is to use a series of GIS data processing functions for raster data. An example is the Topographic Position Index extension for ArcView 3.x (Jenness 2005), which is the difference between a cell elevation value and the average elevation of the neighborhood around that cell. This extension classifies the landscape into slope position and landform categories from elevation grids (Figure 14). TPI values are dependent upon scale, or the size of the neighborhood used in the calculation. Using the TPI at different scales, plus slope, the landscape can be classified into both slope position and landform category (Figure 15).

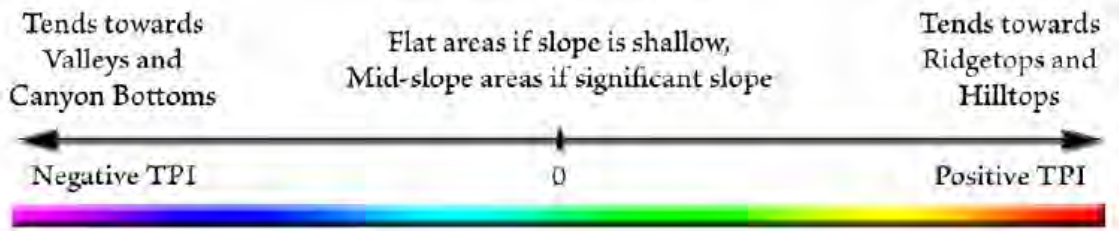


Figure 14. An ArcView extension can be used to calculate the Topographic Position Index. Positive values mean the cell is higher than its neighboring cells while negative values mean it is lower.

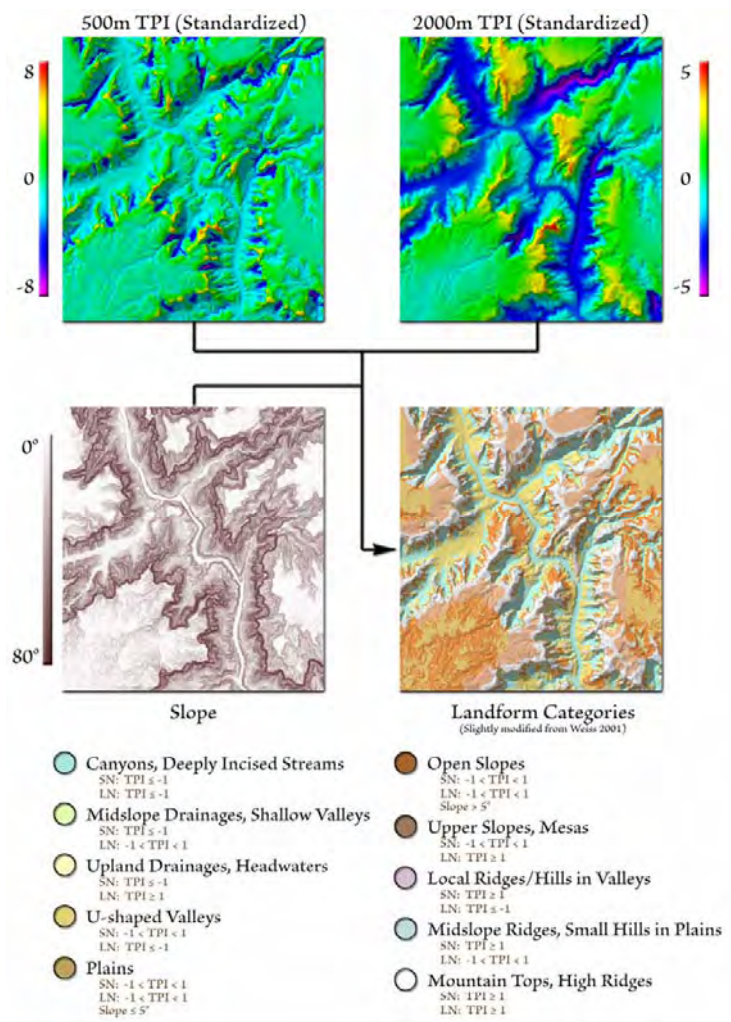


Figure 15. By combining slope and the TPI as calculated from different sized neighborhoods, the landscape can be classified into both slope position (i.e., ridge top, valley bottom, mid-slope, etc.) and landform category (i.e., steep narrow canyons, U-shaped valleys, plains, open slopes, mountain tops, etc.)

Another software package that allows the automated extraction of physiographic features is LandSerf, a free GIS for the visualization and analysis of terrain (Wood 2005). This software classifies a DEM surface into six categories (Wood 2004) - pits, channels, passes, ridges peaks and planar regions (Figure 16). The results of the AFE can be modified by setting two different tolerances: the *slope tolerance* determines how steep the surface can be while still being classified as part of a pit, pass or peak feature (larger values tend to increase the number of point features) and the *curvature tolerance* determines how convex or concave a feature must be before it can be classified into a category (larger values tend to increase area classified as planar regions, leaving only the sharpest features identified).

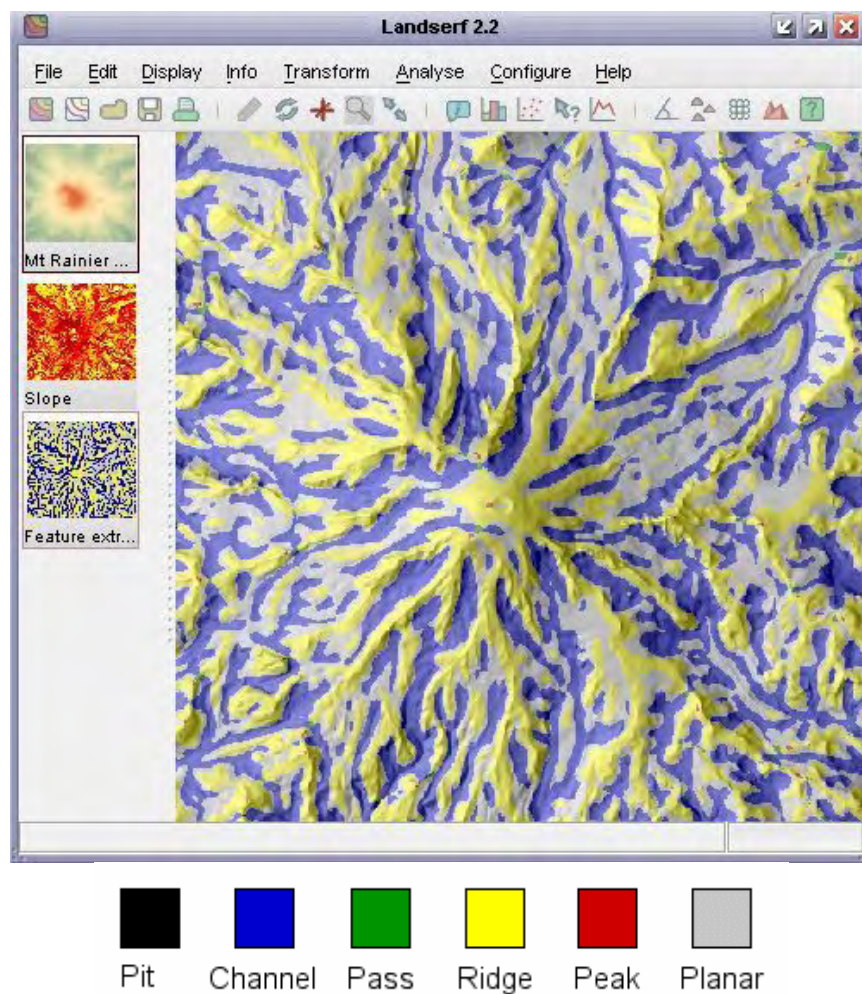


Figure 16. LandSurf uses AFE to determine the locations of pits, channels, passes, ridges, peaks, and planar regions (graphic from Wood 2004).

An added feature in LandSurf is the ability to identify “fuzzy features” since the same feature could be classified into a number of different categories at different scales

(Figure 17). While this is useful for some applications, for the purposes of labeling, it is ultimately necessary to determine a fixed extent of the features that will be stored in the GIS database and used for labeling.

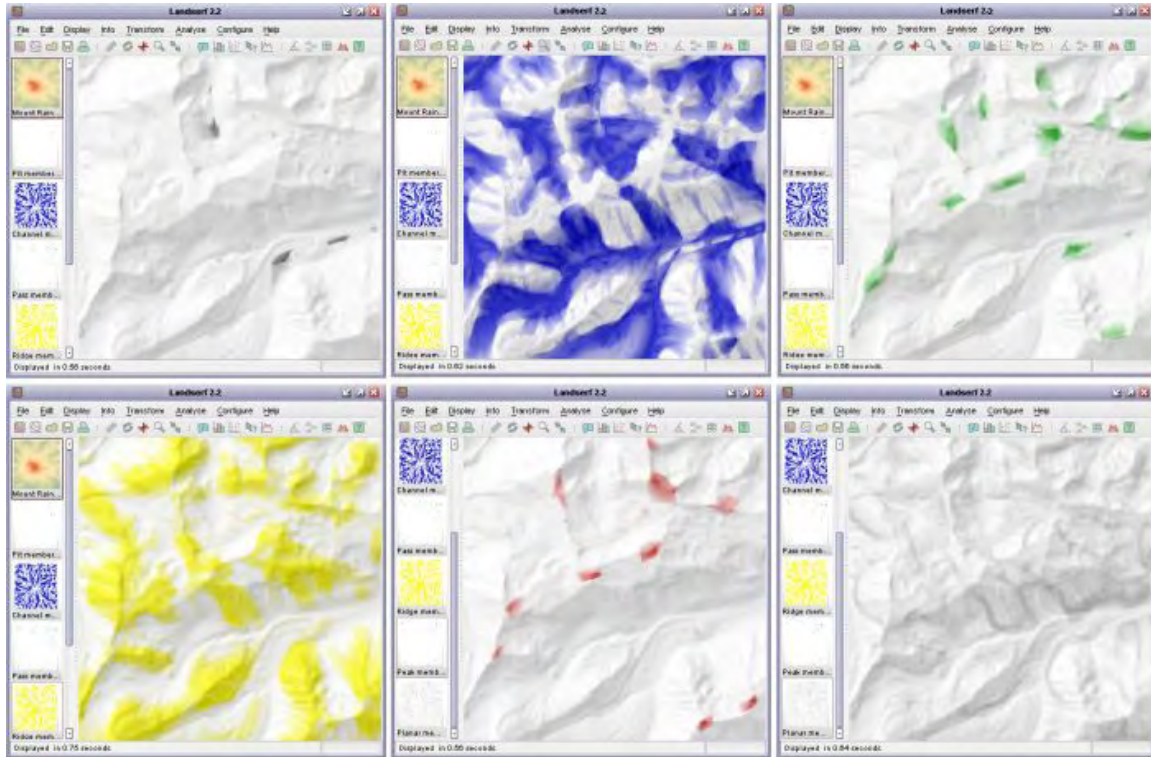


Figure 17. LandSurf's fuzzy classification of features into the six landform categories (graphic from Wood 2004).

AFE can speed up the process of creating feature outlines as well as assure that all features of the same type (e.g., canyons or passes) are created. Additionally, once the features are created, one could intersect the GIS representations (i.e., polygons in most cases) with a database file containing the names of the features, thereby expediting the feature attribution task as well as assuring that all features that need to be named are. However, some names may not end up being associated with their geographic features if the AFE creates a form that does not coincide with the point location of the name. This is especially problematic since the point location of the name in many GIS names databases is associated with the location of the text on the map and not necessarily with the location of the feature in geographic space – exaggeration and displacement may have been used to avoid cartographic symbol/text conflicts. An additional drawback of the AFE approach is that many more features will be identified than have been named. The presence of a particular type of physiographic feature does not ensure the naming of that feature – named features in the landscape are or at least were at one time culturally

significant. Finally, and perhaps most importantly, an AFE approach to delineating the GIS features does not guarantee that the types of features on the map, or their extents, will be the same as the designation of the type or extent of features using the DEM. Resolution and extent aside, the DEM does not inherently contain human interpretation of the landscape features – maps do, since named features in the landscape are a cultural construct, not a physical one, and they are certainly not a consistent or even rational one. Therefore, the extent of a mountain on a map may be greater or smaller than the extent identified through AFE, and the designation of that mountain may be wildly different – on a map, it may be a mountain, a mountain range, a mesa, a butte, a peak or some other derivative. Correlating the AFE results with the cultural subdivision of the landscape is time-consuming and tedious. Keeping in mind that the purpose of delineating named places as GIS features is to label them on a map, the AFE approach may not capture the maximal extent for maps that will be viewed at varying scales (as with Web-based maps that are zoomed in and out). Erring on the side of caution and setting the AFE parameters to overcapture the extent may result in some spatially restricted features to be labeled outside their bounds as apparent on existing maps.

### *Quality Control (QC)*

Once the features have been captured, either through digitizing or through AFE, they should be subjected to a quality control (QC) process. Existing maps would seem to be the best source to compare the captured results against. As suggested, three independent sources should be used to demonstrate the extent of the feature. These sources could also be used to verify the naming of the places although one could also defer to a geographic names authority. In either case, the sources used for verification should be detailed in the GIS database as attributes for the features. Ideally, it would also be useful to have someone with local on-the-ground knowledge confirm the location of the features. Hikers, bikers, campers and others with detailed personal knowledge of the place can provide valuable input.

### **LINKS TO STANDARDIZED GEOGRAPHIC NAMES DATABASES**

As mentioned above, one could 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 is 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) includes names and associated information about places. Places in TGN include administrative political

entities (e.g., cities, nations) and physical features (e.g., mountains, rivers). Current and historical places are included. The position of the place is indicated by geographic coordinates; bounding coordinates and elevation may also be included. As another example, the GEOnet Names Server (GNS) provides access to the National Geospatial-Intelligence Agency's (NGA) 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 [United States Geological Survey](#) (USGS) [Geographic Names Information System](#) (GNIS). The utility of these datasets can be 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 name database does exist, and the challenges in linking to a geographic feature dataset can be overcome, then it might be advantageous.

In any case, the use of a centralized names database assures that the names are maintained in one location and 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 dataset. 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 will 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 apparent extent of the feature.

## CONCLUSIONS

In this paper, we described three different types of natural features with indeterminate boundaries (named marine water bodies, named physiographic features, and islands and island chains) 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 or fuzzy boundaries worked very well to demarcate an adequate extent of the feature name on the map. In our work, we found that GIS text placement algorithms were for the most part insufficient for final text placement, and text had to be hand edited in order to sufficiently and elegantly imply the locations of the features. However, the automated text placement algorithms saved a great deal of time by placing the text so that only minor adjustments were required.

Also important considerations in labeling were the cartographic attributes SizeClass for named marine water bodies and Order for named physiographic feature polygons. This

information allows the GIS to vary the text size relative to the map scale, as only the relative text size variations need to be referenced instead of the specific text sizes (e.g., smaller and larger font versus 10 and 12 point font). Other attributes such as polygons area are not always sufficient to make this determination as features of the same size may vary in cultural and geographic significance (i.e., notoriety). In our work, we automatically assigned this attribute using polygon area for the named marine water bodies, and we manually assigned the attribute for named physiographic features.

In this paper, we assumed that one of the primary purposes 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 to become more involved in the design of GIS databases by contributing their knowledge and expertise.

## REFERENCES

Alexandria Digital Library Project (ADL). 2004. Available at <http://www.alexandria.ucsb.edu/gazetteer/>.

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 *Geographic Objects with Indeterminate Boundaries*, eds. P.A. Burrough and A.U. Frank. London: Taylor & Francis, pp. 257-269.

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

Digital Geographic Information Exchange Standard (DIGEST). 2001. Available at <http://www.digest.org/>.

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.

- Getty Thesaurus of Geographic Names (GTN). Available at:  
[http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/about.html](http://www.getty.edu/research/conducting_research/vocabularies/tgn/about.html).
- Jenness, J. 2005. Topographic Position Index (tpi\_jen.avx) extension for ArcView 3.x. Jenness Enterprises. Available at: <http://www.jennessent.com/arcview/tpi.htm>.
- 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. 1965. *Landforms of California and Nevada*. 1:500,000. Z-Point Graphics, Whistler, British Columbia, Canada. Posted 10-Nov-2004. Computer Database Title. <http://www.panorama-map.com/Raisz/Calif.%20Nevada/Raiszcalif500.html>, accessed 22-Mar-06.
- Raisz, E., and W.W. Atwood. 1957. *Landforms of the United States*, 6th Ed. Cambridge, MA: E. Raisz.
- Smith B. 1995. "Drawing Lines on a Map", in *Spatial Information Theory*. Proceedings of COSIT 1995, Semmering, Austria, New York: Springer Verlag, pp. 475-484.
- Smith B., and Mark, D.M. 2003. "Do Mountains Exist? Towards an Ontology of Landforms", *Environment & Planning b: Planning and Design*, 30(3), pp. 411-427.
- U.S. Geological Survey (USGS). 1992. Physiographic Divisions of the Conterminous U.S. Available at: <http://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml>.
- Waters T., and Evans A.J. 2003. "Tools for Web-based GIS Mapping of a "Fuzzy" Vernacular Geography", in *Proceedings of the 7<sup>th</sup> International Conference on GeoComputation*, University of Southampton, United Kingdom, 8-10 September.
- Wood, J. 2005. LandSerf. Available at: <http://www.soi.city.ac.uk/~jwo/landserf/>.
- Wood, J. 2004 "A New Method for the Identification of Peaks and Summits in Surface Models", in *Proceedings the GIScience 2004 Conference*, University of Maryland University College, Adelphi, Maryland, 20-23 October, pp 227-230.

## Appendix A: ADL Physiographic Features

Used for:	Physiographic Features:	. . ridges
aprons (geological)	. alluvial fans	. . . drumlins
blowholes	. arroyos	. natural rock formations
blowouts	. badlands	. . arches (natural formation)
boulder fields	. banks (hydrographic)	. plains
cones (geological)	. bars (physiographic)	. plateaus
crevasses	. basins	. playas
fissures	. . storage basins	. reefs
furrows	. beaches	. . coral reefs
geological features	. bights	. seafloor features
grades (physiographic)	. capes	. . abyssal features
hammocks	. caves	. . continental margins
inclines	. cirques	. . fracture zones
landslides	. cliffs	. . hydrothermal vents
meander necks	. craters	. . ocean trenches
itches (physiographic)	. deltas	. . seamounts
polders	. dunes	. . submarine canyons
potholes	. flats	. tectonic features
rockfalls	. gaps	. . earthquake features
sinkholes	. isthmuses	. . faults
sinks	. karst areas	. . . fault zones
siphon (physiographic)	. ledges	. . . rift zones
slides (natural)	. massifs	. . folds (geologic)
slopes	. mesas	. . . anticlines
talus slopes	. mineral deposit areas	. . . synclines
terraces (physiographic)	. moraines	. valleys
uplands	. mountains	. . canyons
	. . continental divides	. volcanic features
	. . mountain ranges	. . lava fields
	. . mountain summits	. . volcanoes

## Appendix B: DIGEST Physiographic Features

### D - Physiography

#### DA - Physiography-Exposed Surface

##### Materials

DA005 Asphalt Lake  
 DA006 Alkali Flats  
 DA010 Ground Surface Element  
 DA020 Barren Ground  
 DA030 Land Area  
 DA031 Land Region

#### DB - Physiography-Landforms

DB010 Bluff/Cliff/Escarpment  
 DB030 Cave  
 DB031 Hill  
 DB060 Crevice/Crevasse  
 DB070 Cut  
 DB080 Depression  
 DB090 Embankment/Fill  
 DB100 Esker  
 DB110 Fault

DB115 Geothermal Feature  
 DB145 Miscellaneous Obstacle  
 DB150 Mountain Pass  
 DB160 Rock Strata/Rock Formation  
 DB161 Large Isolated Rock, Boulder, or Rocky Formation  
 DB170 Sand Dune/Sand Hills  
 DB176 Slope Category  
 DB180 Volcano  
 DB190 Volcanic Dike  
 DB200 US-Gully/Gorge UK-Gullies  
 DB210 US-Potential Landslide Area UK-Landslide/Scree  
 DB211 Landslide  
 DB220 Undermined Land  
 DB230 Fan  
 DB500 Bottomline of Cliff  
 DB501 Topline of cliff