cartographic techniques

Using Valid Value Tables in Geodatabase Design to Define Feature Types

Aileen Buckley
Environmental Systems Research Institute, Inc.
380 New York Street
Redlands, CA 92373
e-mail: abuckley@esri.com

Introduction

There are many ways to encapsulate semantic models in GIS and cartographic data. A semantic model is the set of terms used to describe features in the database or on a map. For instance, a semantic model defines whether a low-lying saturated areaperpetually on the landscape is called a swamp, marsh or bog. In order to make maps with GIS data, some part of the GIS data model must contain the data's semantic model so a mapmaker can symbolize the data for the map.

Valid Value Tables (VVTs) are a set of tables that may be used to store a semantic model in a geodatabase by defining the valid combinations of coded values that describe the kinds of features in the database. Coded values are numbers (requiring relatively small amounts of storage space in a database and low impact on digital networks) that represent larger, more descriptive, but inefficient text strings. Drawing data on maps in a GIS is faster when coded values are used to determine which symbols are used to draw features.

The **DI**gital **G**eographic Information **E**xchange **ST**andard (DIGEST) was the first semantic model that ESRI implemented using VVTs. DIGEST uses coded values to define the major kinds of geographic features, their attributes, and the values for the attributes [Digital Geographic Information Exchange Standard, 1999]. Although DIGEST does not identify features as points, lines or polygons, it does provide terms for the types of geographic features that might be included in a geographic database, as well as the types of attributes that those features might have. An example of the five-digit coded geographic DIGEST feature and its three-digit coded attribute is:

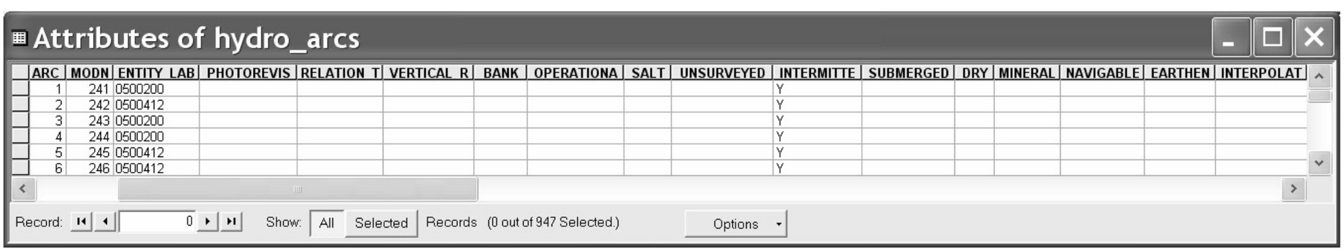
A heliport is encoded with a feature code of GB006 (Airfield) where G indicates that the feature is in the category containing Aeronautical Information and B indicates that it is some type of Aerodrome, and it is associated with the attribute APT (Airfield type) containing a coded value of 009 (Heliport).

In this example, letters are used to differentiate categories of information; however, these could be substituted with numeric values to make the database more efficient.

Coded values are also used by the U.S. Geological Survey (USGS) to define the features in Digital Line Graphic (DLG) files. In the DLG case, the coded values are stored as major and minor codes in pair combinations, and each geographic feature may have multiple major and minor "code pairs". Each major code consists of three digits indicating a major class of features, such as 050 which is hydrography. Each minor code consists of four digits identifying characteristics of a feature. The minor code may describe the basic feature type or it may indicate additional characteristics beyond the basic feature type. For example, the hydrography minor code 0412 indicates that the feature is a stream, and the minor code 0601 indicates that the feature is underground. If a feature has both of these codes it is an underground stream. The complete codes are 050.0412 and 050.0601.

Once extracted from the DLG format, the major and minor DLG codes are all displayed as numerals. For SDTS format data, entity codes, which are text strings of numerals, indicate the major code and the feature type code (e.g., 050.0412 in the example above), and the attributes are stored in fields specific to the feature type.

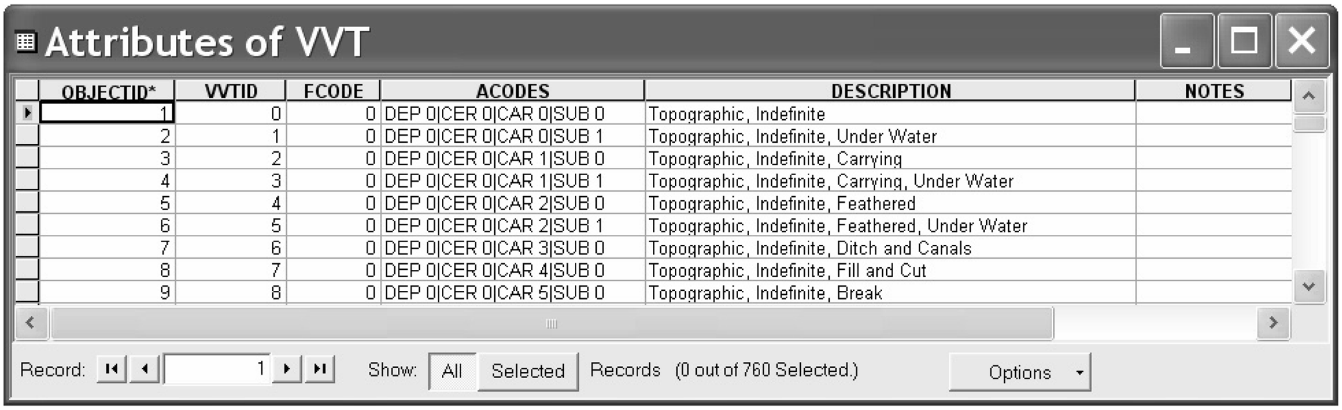
In both of these examples, both the major and the minor codes, or in the case of the DIGEST model, the feature and attribute codes can be considered an attribute of a geographic feature. Although the database storage of this attribute


Attributes of hydro_arcs

ARC	MODN	ENTITY LAB	PHOTOREVIS	RELATION T	VERTICAL R	BANK	OPERATIONA	SALT	UNSURVEYED	INTERMITTE	SUBMERGED	DRY	MINERAL	NAVIGABLE	EARTHEN	INTERPOLAT
1	241	0500200								Y						
2	242	0500412								Y						
3	243	0500200								Y						
4	244	0500200								Y						
5	245	0500412								Y						
6	246	0500412								Y						

Record: 0 Show: All Selected Records (0 out of 947 Selected.) Options

Figure 1. Portion of an attribute table for hydrographic arcs that were imported from the SDTS format. The Entity Label (entity_lab) field contains the combination major.minor code, and the remaining fields are attributes associated with DLG linear hydrographic features.



OBJECTID*	VVTID	FCODE	ACODES	DESCRIPTION	NOTES
1	0	0	DEP 0|CER 0|CAR 0|SUB 0	Topographic, Indefinite	
2	1	0	DEP 0|CER 0|CAR 0|SUB 1	Topographic, Indefinite, Under Water	
3	2	0	DEP 0|CER 0|CAR 1|SUB 0	Topographic, Indefinite, Carrying	
4	3	0	DEP 0|CER 0|CAR 1|SUB 1	Topographic, Indefinite, Carrying, Under Water	
5	4	0	DEP 0|CER 0|CAR 2|SUB 0	Topographic, Indefinite, Feathered	
6	5	0	DEP 0|CER 0|CAR 2|SUB 1	Topographic, Indefinite, Feathered, Under Water	
7	6	0	DEP 0|CER 0|CAR 3|SUB 0	Topographic, Indefinite, Ditch and Canals	
8	7	0	DEP 0|CER 0|CAR 4|SUB 0	Topographic, Indefinite, Fill and Cut	
9	8	0	DEP 0|CER 0|CAR 5|SUB 0	Topographic, Indefinite, Break	

Figure 2. A valid value table (VVT) for hydrographic lines. The VVTID is the foreign key that links to the primary key in the feature class attribute table. The ACODES are the valid descriptors and values for a given FCODE, the DESCRIPTION field carries a user-friendly statement of what the coded value combinations mean.

is complex because it consists of a major code and one or more minor codes, it still is logically a single property or descriptor of the feature.

Using the VVT approach, a VVT table contains the semantic equivalents of major and minor code combinations. There is a row in the VVT table that contains the full description for each type of feature. Each geographic feature attribute table has a VVTID attribute which links to a row in the VVT table. Entries in the VVT table include the feature types as well as any valid combinations of descriptors that apply. In this design, the codes themselves are treated like an attribute with a special kind of attribute domain.

These tables can be used as coded value domains in a geodatabase. Coded value domains can be used to specify a valid set of values for any type of attribute—text, numeric, date, etc. Coded value domains contain a simple value and a description of what that value actually means. The description makes using the coded value domain easier for the user.

VVTs can be used to extend the concept of a coded value domain because they contain a systematic structure for the valid attribute combinations. VVTs hold only those real combinations of feature codes and attributes, rather than all possible combinations. These combinations are easily exposed in common GIS activities, such as feature selection, drawing specified features using definition queries, editing and creating new features [MacDonald, 2002]. Despite these added capabilities, VVTs were mainly designed for cartography because not only the feature type often determines the symbology and labeling, but also by the characteristics of features [PLTS, 2003].

Anatomy of a VVT and Related Tables

The information in the valid value table is at the heart of this approach to modeling a semantic model. The primary contents of the VVT are the codes identifying the feature types and all their valid descriptor or attribute combinations. While other information, such as the feature class that the feature types are located within, can be also stored in the VVT, the most important information carried are the feature types and their valid attribute combinations.

In Figure 3, the main feature type is stored in an integer field called FCODE. There is also a DESCRIPTION field to help the user understand what that FCODE stands for.

The descriptors or attributes that completely describe a particular type of feature are defined by the ACODE range in the FCODE tableACODE range in the FCODE table defines the descriptors or attributes that completely describe a particular type of feature. As with the feature types, these values define discrete properties, not a numeric measurement. The ACODE, for the sake of efficiency, may be used for more than one feature class. Therefore, it is necessary to define all possible values for all the feature classes that will use a given ACODE. For example, aqueducts and pipelines can be at or near the surface, elevated, underground or unspecified; streams can be submerged, and contours can be underwater (for example, beneath the surface of a reservoir). Creating a single ACODE for all of these ensures that the semantic model is complete, and it simplifies the database design. The same ACODE can also be used for other features, such as control markers, telephone lines, and elevated railroad lines. Other examples of multiuse ACODEs that are useful for multiple feature types include positional accuracy (approximate, unknown, etc.) and operational status (under construction, aban-

Attributes of FCODE_VALUES

OBJECTID*	FCODE	ACODERANGE	DESCRIPTION	NOTES
1	0	DEP [0,1,2,3,4]|CER [0,1,2]|CAR [0,1,2,3,4,5]|	Topographic	
2	1	DEP [0,1,2,3,4]|CER [0,1,2]|CAR [0,1,2]	Bathymetric	
3	2	DEP [0,1,2,3,4]|CER [0]|CAR [0,1,2]	Glacial or Ice Surface	
4	3	CWC [0]|RTS [0,1]	Spot Elevation	
5	4	CWC [0,1]|PRM [0,1]	Horizontal Control	
6	5	TAB [0,1]	Vertical Control	
7	6	CWC [0,1]	Benchmark	
8	7	TAB [0,1]	Boundary Monument	
9	8	CWC [0]	Reference Monument	
10	9	CWC [0]	River Mileage Marker	

Record: 1 Show: All Selected Records (0 out of 193 Selected.) Options

Figure 3. The values for geographic feature types are stored in the FCODE table, which includes an FCODE field as well as a user-friendly description of the feature codes. Also included in this table are the ranges of ACODEs that are valid values for those feature types.

doned, etc.) A single ACODE that contains all possible attribute values that can be used for multiple features reduces replication and helps to simplify and clarify the semantic model as well as the physical data model.

In Figure 4, the three-letter attribute type or category is stored in a field called ACODE. The attribute types are listed, each with a unique ACODE value (VALUE). This table provides a unique list of coded attribute values for the entire database. As with the FCODE table, a user-friendly description is included. Notes are also included to indicate any special considerations relating to a particular attribute value.

The VVT is constructed using the information in the ACODE and FCODE tables. The VVT contains the valid attribute combinations, as well as a unique description that makes the VVT more user-friendly (Figure 2). In most cases, it logically easier to first create a VVT such that contains all possible combinations of FCODEs and the ACODEs that can apply to the features. Then it is necessary to delete the coded value combinations that are not sensible. Because each record of a VVT represents a unique combination of attributes, the description must also be unique.

The tables containing the FCODES, ACODES, ACODE descriptions and valid values can be created using any software that allows you to set up tables with rows and columns. Once the VVT is created, the Table to Domain tool in ArcGIS 9.0 can be used to create a geodatabase domain. When the VVT is used to produce domains, quality in the database is enhanced and enforced because only valid feature-attribute combinations are allowed.

To migrate GIS data into a database designed using the VVT approach, it is necessary to create a crosswalk table to link the original feature and attribute codes to the corresponding FCODE-ACODE combinations in the VVT. For example, if one is using DLG data imported from SDTS format, then the entity label and the various attributes are used to determine

Attributes of ACODE_VALUES

OBJECTID*	ACODE	VALUE	DESCRIPTION	NOTES
1	DEP	0		Rise
2	DEP	1	Depression	
3	CAR	0		Not Carrying
4	CAR	1	Carrying	
5	CER	0	Indefinite	Up to 100K
6	CER	1	Unspecified	
7	CER	2		Definite
8	DEP	2	Adjacent Depression 1	Up to 250K
9	DEP	3	Adjacent Depression 2	Up to 100K
10	DEP	4	Adjacent Depression 3	Up to 100K

Record: 1 Show: All Selected Records (0 out of 421 Selected.) Options

Figure 4. The ACODE table contains all the coded value attributes that could be associated with geographic features.

which VVTID a particular feature relates to. Alternatively, the FCODEs and ACODEs can be defined in the native data format and the resulting VVTID can reflect at least part of the native major-minor code description, if they were originally in numeric format. However, if multiple attributes are assigned to a feature, then unique VVTIDs must be assigned to each feature-attribute code combination.

The VVT can be created for the entire geodatabase, or separate VVTs can be created for each feature class. If a single VVT is used, then queries and such can be applied across the entire domain once the VVT is converted to a domain for the geodatabase. If multiple VVT tables are used, it might be useful to include a look up table to define which feature classes use which VVT tables. Each VVT will then relate to the combination of one particular type of geographic data and its geometric representation (point, line, polygon). Using a single VVT for the entire database allows feature representation to vary, for example, between scales. As a general guideline, it might be useful to consider using one VVT if features will change geometry between scales. If working at only one scale, then multiple VVTs, one for each feature class, could be used.

Implications of a VVT Approach

A VVT approach helps to ensure semantic integrity in geodatabase design, and it offers additional advantages for database use and for multi-scale and multi-purpose database specifications. There are a number of feature coding standards in the GIS industry that describe features and their characteristics. Any of these feature-coding standards can be stored in VVTs, making it easier and more efficient for users to work with the data. Many of the current database implementations for these systems are inefficient and difficult to learn, and they do not inherently contain easy to understand descriptions.

The VVT approach starts first with identification of the themes to be included in the database (e.g., hydrography, hypsography, cultural features, etc.) Then the feature classes are roughly identified and become further refined as all possible feature types are determined. Finally the attributes of the features are defined and the invalid combinations of features and attributes are eliminated. In the process of first organizing major kinds of features into feature classes, then determining which FCODEs apply to each feature class, and then which ACODEs apply to each FCODE, some basic rules of thumb can be applied to ensure the integrity of the valid values.

1. The number of possible combinations of valid values should not be excessively large; when it is, it is usually a sign that there are not enough FCODEs. Meaning that one of the ACODEs is really more than just an attribute. Occasionally, one FCODE may need to become its own feature class.
2. In many cases ACODEs have default or implied values, rather than explicit values. Because the descriptions from the ACODE values are appended to the FCODE descriptions to form the VVTID descriptions, it is often logical and sensible to leave the implied value blank.
3. The sequence of ACODEs and ACODE values listed in the FCODE Tables ACODERANGE field should be intuitive, so when users use the coded value domains in the geodatabase, they see a logical, sensible listing of feature types in ArcGIS's applications.

Implementing a semantic model or feature-coding standard using a VVT approach effectively ensures the integrity of the model or standard. This approach ensures that two different feature classes will not contain the same feature types (at least for a single scale model), and it assures that there are no ambiguous feature type descriptions.

Database Use. The VVTID or its easy to use descriptions can be used for selection, specifying definition queries, rendering features, editing and creating new features, and data extraction [McDonald, 2002].

Multi-scale Use. Using VVTIDs does not require the data to be in a particular format in the GIS (point, line, polygon). This is especially useful for multi-scale databases in which features may change geometry through scales (e.g., a building is an area at one scale, becomes a point at a smaller scale and is aggregated to an area at an even smaller scale, then disappears altogether at a still smaller scale). The VVT approach is not representation-dependent and can be applied regardless of the format of the data.

Multi-purpose Use. VVTs could be used for multiple products as well as multiple scales. For example, the VVT could be used to define features that are shown on different types of maps and to help define different symbology for different products. It should be possible to include in the VVT, or a table related to the VVT through the VVTID, the scales and/or products for which each VVTID is appropriate.

Conclusion

While there are potentially hundreds of thousands of possible combinations of possible feature types, in reality there is a much smaller subset of valid code combinations, and even fewer of these that may actually exist in the database. The VVT approach is a good database design because it reflects the logic of the coding standard; it compresses the data representation to only deal with valid combinations; it supports quality assurance (QA), editing, and queries; and it supports multi-scale multi-purpose GIS use.

References

Arctur, D. and Zeiler, M., 2004. Chapter 9: Cartography and the Base Map, in *Designing Geodatabases: Case Studies in GIS Data Modeling*, available: http://support.esri.com/index.cfm?fa=downloads.dataModels.filteredGateway&dmid=3, accessed 6 May 2004.

Frye, C., 2003. The 24:000-Scale Topographic Base Map Data Model. *Cartography and Geographic Information Science* 30(2):163-168.

MacDonald, A., 2003. Feature Coding Standards and Geodatabase Design, An ESRI Technical Paper, available: http://downloads.esri.com/support/whitepapers/ao_/Feature Coding Standards and Geodatabase Design.pdf, accessed 6 May 2004.

Environmental Systems Research Institute (ESRI), 2004. Introduction to Database and Map Production with PLTS, training manual for class offered January 2004.

Digital Geographic Information Exchange Standard, 1999. DIGEST Overview, available: http://www.digest.org/Overview2.htm, accessed 6 May 2004.

U.S. Geological Survey, 1995. National Mapping Program Technical Instructions: Part 3: Attribute Coding: Standards for Digital Line Graphs, available: http://www.geodata.gis.state.oh.us/dlg/standards/3DLG0797.pdf, accessed 6 May 2004.
available:
http://support.esri.com/index.cfm?fa=downloads.dataModels.filteredGateway&dmid=3., accessed 6 May 2004.

reviews

The Salton Sea Atlas

By The Redlands Institute, Redlands, CA: ESRI Press, 2003. Hardbound (ISBN 1-58948-043-0), $79.95. 127 pp., index, bibliography.

Reviewed by Judith A. Tyner, Ph.D.
Department of Geography
California State University,
Long Beach

To those outside of California the Salton Sea is not a familiar place, so why one would want to spend nearly $80 on an atlas about a shallow lake in the middle of the desert?

The Salton Sea is the largest lake in California, covering 376 square miles with a surface elevation of minus 227 feet and a maximum depth of 51 feet. It was formed when the Colorado River flooded in 1905 and 1906 and breached an irrigation diversion dam. While not as familiar as Lake Tahoe, it is a significant recreational area of great ecological importance rich in flora and fauna. These are the basic facts. Over the years, there have been arguments stating that the Sea is an endangered ecosystem or an artificial body of water destined to dry up and, therefore, not worth saving. This atlas provides probably the most complete information about the Salton Sea.

The Salton Sea Atlas was a monumental undertaking, 4 years in preparation, with a team of dozens of geographers, biologists, limnologists, GIS specialists, illustrators, and cartographers. It is divided into two main sections with five subsections plus an index and bibliography. The main sections are the descriptive text and the maps. The subsections include introductory materials that explain the project, use of GIS and the processes involved in creating the atlas. "Physical Geography" describes landforms, hydrology climates (both modern and paleo), and biomes. "Cultural History" treats the human occupance of the area. "Limnology/The Sea Today" focuses specifically on the Salton Sea; "Ecology" deals with life in the sea divided into birds, animals, and fish; and "Future of the Salton Sea" briefly notes the problems. The final section consists of 39 pages of maps totaling 98 individual maps.

Text is not set solid in the usual way, but is often in the form of blocks or boxes interspersed with striking graphics. The pages of this section are a blend of high tech GIS, satellite imagery, and artwork. There are numerous paintings of plants, animals, birds, fish, and reptiles. Using paintings rather than photographs of flora and fauna eliminates the sterile look that one finds with some computer-generated works, and certainly is a major factor in the overall attractiveness of the work.

The maps cover every mappable aspect of the area. Although many focus on the Sea itself, there are some, such as earthquake epicenters that deal with Southern California, while others, such as climate and political districts, show