OGC
Open GIS Consortium

Data Models and Interoperability

An Open GIS Consortium (OGC) White Paper

1 What is a Data Model?

Data (or content) models are not the most exciting part of designing and implementing a software application - but they are one of the most essential. A data model details how to take real world ideas or objects and make them useful to a computer system. In the geospatial world the focus is on depicting things in the real world as points, lines, or polygons (the geometry ”primitives” we use to assemble locational data about those real world objects) and their attributes (information about those objects). When linked together, a pair (geometry and attributes) representing one or more real world objects, is called a feature.

Software vendors, government organizations, and industry organizations have defined many application-specific data models (those augmenting and built upon internal data models) available for use by specific geospatial disciplines. Application-specific data models are designed to give users “a leg up” in getting their particular application running. Starting with a well thought out model and tweaking it to specific local needs typically shortens the implementation schedule considerably. From the vendor standpoint, these models may make the sales process speedier and more effective.

Industry organizations tackle data models for a variety of reasons that sometimes involve enhancing data sharing among member organizations. The Open GIS Consortium (OGC) is only interested in one type of data models -- one that enhances interoperability.

2 OGC and Data Models

OGC has worked on data models since 1995. One of the Consortium’s first documents was a Spatial Schema, essentially an overarching data model. The Spatial Schema is an abstract information model and not a specific implementation or physical content model. Many OpenGIS® specifications use the Spatial Schema as a key building block. The model, now encapsulated in ISO 19107 (Spatial Schema), forms the basis of the OpenGIS Simple Feature Specification, OpenGIS Geography Markup Language Implementation Specification and the Federal Geographic Data Committee’s Framework Data Layers, among others.

http://www.opengis.org/docs/01-101.pdf

The FGDC Data Content Standards are of particular interest because they are designed for exchange. The objectives of the Cadastral Data Content Standard, for example, include providing “common definitions for cadastral information found in public records, which will facilitate the effective use, understanding, and automation of land records,” and “to standardize attribute values, which will enhance data sharing,” and “to resolve discrepancies … which will minimize duplication within and among those systems.” These standards include both an abstract or logical representation of the each data theme and, as an implementation annex, a physical model for its encoding and exchange.

3 Using Data Models for Interoperability

Where do these FGDC data models come from and how do they enable this level of interoperability? Consider, as an example, the development of a transportation data model. In 2002, the Road Transportation Model Advisory Team (MAT), which included FGDC and OGC participants, began with a survey of how transportation data is stored by local, state and federal government agencies, and specifically, by departments of transportation. Although a number of differences were identified, the team also found many similarities.

For example, when street geometry (the lines that make up road networks) is defined, most organizations use as a building block a simple construct, a line defined by a set of coordinates that include a beginning and ending coordinate. The standard transportation model refers to these simple lines as “segments.” Segments are typically collected into groups to form “paths.” In a path, the segments meet at “nodes.” While different creators and maintainers of GIS road data may call these basic elements different names and store the information in different ways, most have “something like” these basic elements in their existing models.

In addition to geometry, attribute values (descriptive information) are linked to the geometry. Just like the geometry, each organization stores different attributes and calls them by different names. Some have just a few attributes, others have long lists. The Road Transportation MAT discovered that just as with geometry, there are some basic attributes that everyone shares, which are termed “required.” A long list of “optional” attributes is also included in the model if organizations have and want to make more attributes available. Where did that list come from? The team borrowed from ISO/TC 204’s standard on transportation. http://www.iso.ch/iso/en/stdsdevelopment/tc/tclist/TechnicalCommitteeDetailPage.TechnicalCommitteeDetail?COMMID=4559

The FGDC models that detail geometry and attribute structures are written in Unified Modeling Language, UML. UML is a language designed for documenting and writing data models. While UML is very useful for diagramming concepts and relationships to advise implementations, say in a particular vendor’s software package to communicate the model effectively for its implementation of information exchange, a simpler, well-known encoding model is also needed. Geography Markup Language, GML, is open, published and is already supported by many GIS vendors. GML is a set of XML schema packages that can be used to encode geometry and its properties (attributes) independent of any data/content model. Therefore GML is a logical choice for encoding common spatial data exchanges. (To be clear, GML is content model independent but can be used to encode and communicate any spatial content stored in some content model. The actual GML schema developed to define a particular data model is called an application schema).

Once a model is completed and available, say as an application schema of GML, it plays two important roles: one is for those representing their own native data models, the other is for those organizations looking to make their data shareable with neighbors, states, commercial entities in a broader Spatial Data Infrastructure (SDI). Data model implementers can take the common model and use it as the basis of their own data model, developed for their own purposes. If a transportation organization on a remote island in Puget Sound wanted to tweak the FGDC model for its needs – perhaps removing some optional attributes that are not applicable, or including some geometries only used in its system, that is possible. Software vendors can also use the model as the basis for dynamically importing and exporting data into and out of a version in its own internal data format.

The common data model acts as the starting point for these and other data model creators. And, as you might expect, the more similar the features and associations are in a given implementation to a common model, the more easily the information can be transformed on request to match a common model and share with others.

For those who want to make data available for sharing, the FGDC data model is implemented as a “virtual model layer” and their local data model is mapped to the virtual model as part of the implementation. Think of it as a ‘Rosetta Stone’ that is used to translate from one data model to another, except that both models are translated into and out of the FGDC model rather than directly. Direct translation requires a one-to-one map between the two models, which is fine when one has a limited set of models. But when the number of models can run in the thousands the number of one-to-one mappings becomes unmanageable.

There are actually a few ways to create the virtual model layer. One is to use an Extensible Style Language Transformation*,* XSLT, the language used in XSL style sheets to transform XML documents into other XML documents. Essentially the incoming query could be “translated” into a language that the local system understands, then on the way back out, have the data be “translated” into that of the common data model.

A second way to create the interface layer is with custom or perhaps off-the-shelf software to directly transform native data content into a common exchange encoding. An off-the-shelf solution might include a wizard to help the end user “match up” the local data model to the reference one. Solutions will depend in part on the state of the existing data model. Consider for example that some of the transportation data may not even reside in “traditional” GIS programs. Some graphics might be stored in one database, while other attributes are held elsewhere. Further, even the geometry may be distributed between multiple systems - not uncommon with dynamic segmentation or in towns where assessment software holds property information. In those cases, custom software might be required to link the two together and make them “look like” the reference model.

4 The Future

The future vision for sharing spatial data might look like this: Each of the smaller counties or towns hosts its own online GIS. Each uses software and a data model selected to best meet its local needs. It’s easy to imagine that rural counties or towns might have a different data model than urban ones and coastal ones will be different from those inland. But, if each one also makes its data available “as though” it’s in the reference data model, it’d be relatively simple to write an application that knits two or six or an entire state or province, or even a nation’s worth of data together in a single application. And, the data, since it’s accessed “live,” would be as up-to-date as the data held on the local county servers. Also, those interested and who have permission, can work with the data, use it, and understand it, without creating a local copy. This vision underlies many of the ideas of the National Spatial Data Infrastructure (NSDI).

How do we get there and how long will it take? The data models are well underway. Several are in review (http://www.geo-one-stop.gov/participate/status.html) and the rest are due to be completed in 2003. A GML version of the transportation model is also under development, and hopefully others will follow. More and more vendors are including support for GML and Web Feature Service (WFS, which turns out to be a reliable, open way to make feature data available to others) in their software offerings.

The data models and the technology to take advantage of them are coming. It is time for citizens and geospatial software and data users to push for better and more consistent ways to publish and access data of common interest. They need to ensure their local governments require configurable support for GML and WFS specifications from their software providers (vendors), from third parties, and from consultants. Vendors and consultants must demand that models be made available in a single open way, in GML. Finally, local state and federal governments need to work together to create the “human infrastructure” that will make NSDI a reality.