

A Crucial Part Of GIS For Oil And Gas Database Models

A P&GJ Staff Report

The structure of a pipeline company's data model is important to any company's geographic information system (GIS) implementation. Database model selection is one of the first steps in building a GIS since all of the data for the GIS will reside in this model or be tied to this model. APDM, ISAT and PODS are three database models available. Each year, pipeline companies discuss their use of database models and associated enhancements.

In this roundtable article, panelists from different facets of the database model spectrum discuss their experiences with APDM, ISAT, and PODS. Participants include: Scott Hills, senior staff scientist, Chevron; Debra Rohrer, team leader engineering Services Geographic Information and Mapping Services (GIMS), Columbia Gas Transmission; John Utley, administrator, Data Management and GeoFusion Engineering Systems El Paso Corp., and Ed Weigele, general manager at Willbros Engineers, Inc.

P&GJ: *What is the importance of a data model, standard or otherwise, in the pipeline industry?*



Weigele: When the first standard data model (ISAT) was introduced in the mid 1990s it was a significant event. For the first time the industry had a model that was not totally custom, and the pipeline operators benefited significantly. Today, data standards are important, but in a different way. With ISAT, PODS and APDM as the predominant players, each has a different place and definition.

It is imperative that the model be an enabler and not too restrictive or dominant in the software selection process. If the software platform is too specific to a model, you may limit your flexibility in the future. If your model is too universal, you risk not being able to fully leverage the core GIS platform you have selected.



Utley: A data model is of paramount importance to an organization. It represents a concise representation of any particular organization's deployment of data. I wholeheartedly support the concept of a standard data model for all the obvious reasons:

- Provides organizations an opportunity to not reinvent the wheel or at least provides a starting point.
- Provides an anticipated environment for

vendors to write software against, thus lowering development costs and providing a wider choice of applications to deploy on the data model.

- Allows for interoperability between companies and better supports integration activities.
- Provides proof of concept, implementation strategies, and best practices to the industry.

While there are wonderful reasons for embracing standards to-date, the industry has not necessarily reaped the benefits of a standard data model. Within each data model standard, there are numerous iterations that are actually implemented. Generally the deviation from the standard is driven by vendor applications that support only a particular flavor of the target model or have extended attributes that are not necessarily dictated by the standard model. In fact, most vendors end up supporting several iterations of a single application that run on variations of customer models. In theory this should not happen. A single code base should suffice for all iterations of the model, regardless of client or vendor.

It is worth mentioning that the APDM effort has taken an interesting approach to creating a standard. Instead of creating a detailed comprehensive model that would eventually fracture into a specific organizations, or vendor iteration of the standard, the APDM team chose to only define a core set of standard tables and behaviors that are required. The rest of the model is only modeling the relationships for event types (line, points, and polygons). The content and structure of any particular table is totally up to the implementing party and their particular business requirements.

An operator organizes their potential data sets into behavioral categories (standard behaviors) and as new behaviors are identified, they are implemented into the core abstract classes. The idea is that if the core structure and the basic relationship scheme are adhered to, vendor software should be able to be written that would anticipate the model schema and content. The only problem is that the model is light years ahead of the available software; to date, the only software available is implementation specific, but this concept is worth watching and does hold promise for the industry.



Hills: The importance of a data model in business is independent of the model's particular application: in general, data models are intended to provide a structured design for the storage, maintenance, and

use of information. Consistency in storage and maintenance facilitates use of the data. Business-specific considerations come into play when you consider model design.

A well-designed model specifies storage structures and rules which are appropriate for the type of information being managed. In the business of pipeline operations, structures and rules should model pipeline facilities and the activities associated with their use and maintenance.



Rohrer: A model provides a framework. Just as you wouldn't build a house without a foundation, it's unimaginable to implement a GIS for a pipeline company without a model. A model is an enabler, establishing conventions, sometimes flexible, sometimes rigid, for how the data will be stored, accessed, used, and even how it will behave.

A pipeline GIS must be amazingly flexible while adhering to established procedures and conventions to ensure data integrity. This requirement stems from its need to serve many end-users with very different needs. One community might require polygon processing, for example footages of pipe for tax purposes, while another, linear referencing to align data spatially along the pipeline for analysis. A data model is what allows one to design their GIS accordingly, ensuring that the correct relationships are established to tie the data together and meet a wide spectrum of end-users' needs.

Standardization of models for an industry allows us to do this more effectively as a body — leveraging points of commonality — to get the best value in meeting industry needs. Vendors can program to a model because certain conventions are established. This, in turn, provides companies with flexibility in product choices — reducing timelines and costs normally associated with customization.

Our model, APDM, is not a "standard" in the traditional sense. It is a template that requires specific core classes be implemented in a standardize fashion.

P&GJ: *What data model are you using and how long has it been in production?*

Rohrer: Columbia Gas Co. uses APDM.

Hills: Chevron Pipe Line Co. has been using PODS since 2002 to manage the data needed to maintain alignment sheets, and comply with regulatory reporting requirements. The mileage represented has varied with acquisitions and divestitures, but is currently at a little over 10,000 miles.

Utley: El Paso Corp. is using the ISAT model; we have over 43,000 miles of pipeline in this model and have been using ISAT sense 2001. Previous to ISAT, El Paso's pipelines were staged in a myriad of systems and models — some standard, some proprietary. ISAT has proven to be a good first step in addressing El Paso's data needs.

P&GJ: *How does a data model integrate into an enterprise system at a pipeline company?*

Rohrer: We regard our GIS as a core application within our company. That means, from an enterprise perspective, our GIS model has to be able to support our enterprise vision. Enterprise integration can occur in a variety of ways. It can be extremely rigid where tables are linked very tightly or it can be managed very loosely where data is "consumed." (I expect that while there may be instances where integration is tight, those instances will be the exception as the cost to maintain that level of integration is significantly higher.)

Our model complements our vision because it was intentionally designed to allow for extendibility and interoperability. We can "grow" into our enterprise vision. That growth needs to be thoughtful and deliberate. It represents a lot of money and a lot of impacts. It will also change. The technology will change. The demands on the industry will change. Even if you could "do it all tomorrow," next week, next year, and in five years, there will be new requirements and technologies. The growing of the "enterprise GIS" is ongoing. Models have to be able to handle that continuing evolution.

Hills: Integration generally involves using one of two approaches to link data records in the enterprise and pipeline databases. Data linkage depends upon having a means to identify related records (e.g., using a common unique identifier) in the two databases. The approach then used to link them depends upon the particular enterprise system and the integration requirements.

Some enterprise systems provide an API or Web service which can be used by applications to access the contents of the associated database. Other systems allow access to the actual underlying database tables which enables publication of a system's contents using "views" of its database. If publication of the enterprise database itself is not feasible, either approach may require implementation of ETL (extract/transform/load) utilities to replicate extracts of data from one database to the other.

Weigle: As the industry evolves, we are quickly realizing that working within an enterprise environment is essential to fully realizing the value of the investments that have been made in GIS technologies by pipeline companies. Also, as the evolution of GIS continues, we are seeing applications that extend well beyond the data-rich operations and engineering environments that we are limited to with the models in use today. Integration into

the enterprise will require consideration and modification of the model design.

Utley: There are a couple schools of thought surrounding data model integration. The most widely used strategy is to practice straight foreign key integration between complimentary systems. This is a simple concept that uses unique keys or defined attributes to tie systems together. It is generally thought that the GIS model should carry the location attribute while the integrated business system should carry the bulk of the event attribute data.

In theory, this is a simple concept but in practice can result in a web of cross-platform relationships and keys that can be difficult to maintain or even present an obstacle to progress. Perhaps the better strategy is to implement and deploy something like service-oriented architecture (SOA) within your overall system strategy. The SOA construct allows organizations to create business objects that hand shake between systems at a generic level.

Provided all systems subscribe to the standard SOA paradigm, organizations should be able to construct services in a manner that is maintainable and readily available to integrate with other business systems within the organization and avoid many of the pitfalls associated with hard integration across systems within the organization. I would add that no organization has fully realized the goal of fully integrating their GIS data model into the entire enterprise. There is great opportunity in the coming years to fully realize this potential.

P&GJ: *What data models make sense for transmission, gathering, or distribution?*

Weigle: There are established data models for the transmission and distribution systems and certainly the transmission models can be utilized for most gathering systems. The issue arises when there is not a consistent reference mode for gathering. I have seen cases where there is a combination of linear-referenced data along with lines on a township map which can cause conflicts with system design, model design, and with application sets. Since the gathering systems will require much the same application set as the pipelines, linear models will work best.

Application preference and GIS core system preference will certainly lead to the appropriate model of choice. Distribution systems have traditionally not used a linear-referenced centerline, so the ability to work with stationing is typically limited to the high-pressure trunks in their service areas. There have been implementations in the distribution environment that have used a combination of linear modeling and traditional network models that work quite well together.

Rohrer: Linear referencing is fundamental to transmission. Our models must be spatially enabled. We have to line data up on the pipeline. Those requirements are what led us to APDM.

Hills: As the regulatory environment expands to include gathering lines, I'm not sure it makes much sense to discriminate between models based on their suitability for transmission and gathering systems. In fact, both PODS and APDM originated as models for maintenance of transmission pipeline data, but both are also now being used to manage data for gathering systems.

Utley: The data models I am familiar with (ISAT/PODS/APDM/PODS Spatial) support the stationing concept used in the transmission pipeline industry. Any of these would be a good choice for a transmission pipeline, based on the strength of these models' support of linear referencing.

I have seen and even participated in the deployment of gathering systems in these linear-referenced models, although I am not necessarily convinced that any of the linear models actually do the gathering world justice. In practice, the gathering system tends to have components of both the transmission domain and distribution domain. That said, it would be my assertion that gathering systems are best supported by the spatial Geodatabase models (APDM or PODS Spatial).

In this environment, events that adhere to the linear referenced paradigm could be deployed as any transmission feature and events that do not support linear referenced paradigm could be modeled as non-stationed features. Based on my limited experience with the distribution domain it appears best supported by standard non-stationed models that are available from GIS vendors.

P&GJ: *How do you view the direction of the current models that are in use in the next three years?*

Hills: I think we'll see developments on several fronts, all driven by the desire for integration and additional functionality. First, I expect PODS will continue to deliver new functionality, including that developed by PODS Working Groups — which partner with other industry organizations such as NACE — and modify the existing model so that it fully realizes the Association's vision of "a data storage and exchange standard."

Second, PODS and APDM are working together to identify design changes that will make their models more compatible. At the very least, I'm hopeful the result will enable data synchronization between the two models so that companies can enjoy the benefits of both.

Third, I expect to start seeing other efforts that will improve the ease with which these models can be integrated with other information systems, whether the latter are enterprise systems or work-process specific application databases. The most promising approach for enabling this integration involves the use of Web services. For this to succeed, however, it will require development of another industry standard.

Fourth, and perhaps most interesting, is

the potential to extend the use of Web services originally designed to support integration to enable automated, event-driven processing as Web services. Imagine, for example, automatically updating the assessment of a line's risk of third-party damage by using changes in the monthly rate of one-call tickets associated with that line.

Utley: I believe in the next three years that the APDM and the PODS Spatial efforts will become the GeoDatabase models of choice. Ideally, from an operator's perspective, there would not be much, if any, difference in the two models, and vendor tools will run on either model transparently. There is a decision point that is being debated in the GeoDatabase community regarding the actual deployment of the components that make up these models as features or events. Features store the geometry of the component while events are interpolated along the route based on an attribute location.

Generally, the traditional way to stage a GIS was to deploy feature through linear referencing — which is fundamental to transmission pipeline — is potentially better supported by the event-based scheme. Based on my simple observations, it appears that the event-based scheme supports processing of data better while the feature-based approach supports display performance better. The jury is still out on this issue, but the result will probably be some sort of a hybrid model containing both features and events — or perhaps an event-based system published to a feature-based system for general display consumption.

The RDBMS standard, PODS, will continue to be viable for the next three years although many of these users will migrate to the Geodatabase technologies. Oddly enough, the event-based implementation of the Geodatabase technologies may breathe some measure of life back into the other RDBMS standard, ISAT; an organization can extend its current ISAT configuration into the Geodatabase world by registering ISAT tables with the Geodatabase. This concept has been used in the PODS world for some time but was only recently adopted by the ISAT community. This scheme for exposing the ISAT database to the robust Geodatabase technologies holds great promise for organizations that have a large investment in ISAT integration or applications that can not be easily replaced in the Geodatabase world.

Rohrer: I think the challenge will be ensuring the expandability of models to meet the requirements of larger and more distributed audiences—enterprise GIS. Different companies will approach enterprise GIS in different ways. Some will bring data back into the GIS, others may “push and pull” data between the GIS and other systems, still others will display data as it correlates through linear referencing.

Most likely, a given company's solution will combine all of these, depending upon their requirements and data integration strategy. Our customers, through exposure to GIS technologies, be it Google Earth, our company GIS, or their personal navigational system in their cars, are becoming increasingly savvy in their expectations. The models must be as adaptable as the organizations that use them.

Weigle: The next three years will define the application provider and services industry and what models are used. As the core platform providers define the offerings to the pipeline industry, there will be a consolidation of the data models that are supported. If the current models cannot keep up with the changes and demands of the users, they will begin to consolidate as well.

This consolidation will also bring new players into the market which will help drive the integration of the data models into an enterprise model. As this is done, traditional enterprise players will have a significant influence in the direction and changes to the models to drive integration with other corporate decision tools. **P&GJ**

(GITA's 16th Annual GIS for Oil & Gas Conference takes place Sept. 24-26, 2007 at the Marriott Westchase Hotel in Houston, TX. Visit gita.org/oilgas for the full program and to register online.)