

DATA MANAGEMENT WITHIN THE PIPELINE INDUSTRY

Robert Brook
Pipeline Industry Manager
Environmental Systems Research Institute (ESRI)
380 New York Street
Redlands, California 92373
rbrook@esri.com

Sheila Wilson
Executive Director
PODS Association, Inc
7921 West 16th
Tulsa, Oklahoma 74127
sheila.wilson@pods.org

Peter Veenstra
Director of Product Development / Technical Committee Chair
Eagle Information Mapping / APDM
14825 St. Mary's Lane, Suite 200
Houston, Texas 77079
pcgv@eaglemap.com

ABSTRACT

Data models are critical to the pipeline industry as they allow users to contain, integrate, and analyze data faster. This need becomes more predominant as pipeline operators implement new technologies that are making more and more data available. This is coupled with the desire for a data system that is stable, allowing a pipeline GIS that can be integrated with other enterprise systems and used to effectively satisfy the increasing requirements of regulatory agencies. Within the industry it is clear that an operator's pipeline data model is the foundation for a new mission critical system; therefore, the model must suit the needs of the operator. When implementing a data storage system, one needs to understand the similarities and differences between data templates, data models, and data standards. While each of these options can provide an operator with different opportunities, the selection of the appropriate type should be based on the organization's data requirements, work flows, processes, and culture. This paper discusses the criteria for selecting a model and the available industry options.

PIPELINE DATA MODELING

The pipeline data model is the foundation of many operator's mission critical business functions, such as pipeline construction, maintenance, and integrity management programs or regulatory compliance. While the model provides the structure for data storage, maintenance, and use, it also provides the environment for solutions, interoperability and enterprise integration. The impacts of a poorly planned solution are significant and far reaching so it is important that the data model suit the needs of the operator. This paper discusses two of the available models, including what an organization needs to know and how the models can meet an organization's needs.

WHAT DO YOU NEED TO KNOW?

The first step in effectively choosing a data model option is to clearly understand the organization's needs, how business is currently performed, and what standard business practices are already in place. The following is important:

- Know where the business is going. This is an excellent opportunity to look over the horizon and make an educated guess regarding the long term direction the business will take. One important step is to identify all the types of pipeline that may eventually reside in the model and into what other systems the model may be integrated.
- Identify the corporate perspective regarding industry standards, particularly when compared to developing a solution designed to meet the organization's specific need.
- What is the overall strategy for information management? Does the business tend to manage information using relational technologies or do individual groups exist with little or no communication between them? Does anyone currently employ spatial management techniques such as found in ESRI's geodatabase or Oracle Spatial?
- How is the organization planning to perform enterprise wide integrations? There are a couple of prominent methods for integration; tight and loose. Tight integration involves the use of foreign keys or related records to form linkages between otherwise independent data stores. In loose integration, reusable web services are the method used to provide access to the desired data.
- What already exists in your organization? Your organization may have gone through other similar selection, e.g. PPDM (Public Petroleum Data Model), from which you can learn.

Next, to differentiate between the available options, the similarities and differences between data standards, data templates and data models must be well understood.

- A data standard defines the data: what should be recorded, how it should be recorded, and how it should be supported by a system in order to retain its full meaning. The standard should enable consistency and predictability in an organization's or an industry's information. A standard is an established format.
- A data template establishes a framework that can be used to collect and store information. It is usually created for a specific purpose but it can be used as a best practice to collect information beyond its original scope. The goal of a template is to establish similarity between like datasets. While a template can be used rigidly to provide consistently, it is meant to provide flexibility.
- A data model is the logical data structure that comes as a result of a database design process. It is the implementation of a data standard or a data template, or even a combination of the standard and a template.

While each of these options can provide an operator with different opportunities, the selection of the appropriate type should be based on the organization's data requirements, work flows, processes and culture.

Lastly, before a data modeling decision is made, it is important to understand what models are available and the similarities and differences between them. With this information you can effectively choose a model that best suits your needs. Below is a description of commonly used industry models.

WHAT ARE THE DATA MODEL OPTIONS?

Four data models are currently available: the Pipeline Open Data Standard (PODS), the ArcGIS Pipeline Data Model (APDM), the Integrated Spatial Analysis Techniques (ISAT) model originally developed by the Gas Research Institute, and the European Commission funded Industry Standard Pipeline Data Management (ISPDM) project. The majority of organizations who are currently choosing a data model select either the Pipeline Open Data Standard (PODS) or the ArcGIS Pipeline Data Model (APDM); therefore, this paper will focus on PODS and APDM.

Before we can effectively discuss the reason to move in one direction or the other, it is important to understand the models. Here is a brief description of PODS and APDM.

The Pipeline Open Data Standard

The Pipeline Open Data Standard (PODS) model was developed and is maintained by the PODS Association and its membership. PODS members include pipeline operators, solution vendors, consulting companies, and regulatory agencies. The PODS model is maintained by members who volunteer their time and is designed as a standard for gas and liquids, for gathering, transmission, and distribution. The standard is for the storage and exchange of all pipeline data required to meet key business drivers such as regulatory compliance, integrity management, facilities and operations management, data collection, land, and many others. The data model is continually expanding to include any and all data associated with the pipeline.

The PODS model began with PODS 2.0, released in 2001, as a modest set of less than 70 tables. PODS 4.0.1 was released in May, 2007, with hundreds of tables. The current model includes tables for pipe location, pipe facilities, geographic features, inline inspections, cathodic protection, close interval surveys, physical inspections, integrity management data, offline events, and more.

The Board and Technical Committee are composed of PODS Association members. At least one-half of both groups are operators. The Board of Directors governs the association and provides direction to the Technical Committee; the Technical Committee makes all changes to the model and oversees the working groups. The Association members work with non-member volunteers in working groups to develop data model and data exchange standards that benefit the entire pipeline industry.

As directed by the Board, the Technical Committee develops standards as well as integrates standards developed by working groups. Standards developed by a working group include tables and fields, data exchange and interchange, and/or best practices. Any standards developed may be submitted to the Technical Committee for consideration and review. The Technical Committee will first determine if the submitted standards are

in agreement with existing logic, submit the standards for member review and comment, then establish the best way to integrate the standards into the data model, as necessary.

The PODS model is relational, allowing vendors and software companies to provide different spatially enabled solutions. One of the primary principles of the PODS Association is that all data models and data exchange standards are “open.” Open refers to several key mandates of the model:

- that the technologies used for model development strive to be open and independent of platform or vendor,
- that membership be open to any company, association, agency or individual,
- that membership directs the activities and priorities of the Association and the model,
- that data model standards developed are non-proprietary,
- that members vote to approve all proposed data standards,
- that funding is raised through membership fee, compliance fees, and sponsorship by members.

The open standard allows operators to use the standard in whole or in part to both store data and more easily exchange data between themselves and vendors or other pipeline companies.

The ArcGIS Pipeline Data Model

The ArcGIS Pipeline Data Model is designed for storing information pertaining to features found in gathering and transmission pipelines, particularly gas and liquid systems. The APDM was expressly designed for implementation as an ESRI geodatabase for use with ESRI's ArcGIS products. A geodatabase is an object-relational construct for storing and managing geographic data as features within an industry-standard relational database management system (RDBMS). The first version of APDM was released in March 2003; in August of 2007 version 4 of the model was released.

The APDM was initially derived from existing published data models and was expanded to meet the needs of gas and liquid transmission pipelines. The APDM was developed by members of the ESRI Pipeline Interest Group steering and technical committees, under the guidance of ESRI. The technical committee includes representatives from pipeline operator and pipeline vendor companies. The model was designed to include a sampling of standard features typically found in 80 percent of pipeline companies, but was tailored to include current key items such as integrity, pipe inspection, high-consequence areas, and risk analysis. The core elements of the APDM were derived from the ISAT, PODS, and ISPDM models.

In keeping with the spirit of other published ESRI models, the APDM is not designed to be a comprehensive or all-encompassing model. Rather, the APDM was created to be a template from which a pipeline operator would start with the core elements of the model, and modify it by adding features or refining existing features. A primary objective of the model was to account for linear referencing of features (stationing). Most transmission pipeline companies refer to the location of features or events that occur along the pipeline

system as events occurring along a route (station series) at a certain distance (measure). Stationing was handled in the model using out-of-the-box technology referred to as routes and measures.

The APDM Association is run by two committees: the APDM Steering Committee and the APDM Technical Committee. Under the guidance of the steering committee, the technical committee is wholly responsible for the development of the model. The committees are composed of elected members from the pipeline industry including operators, solution vendors and consultants. APDM promotes free membership, but membership is reserved for organizations rather than staff.

WHAT ARE THE SIMILARITIES AND DIFFERENCES?

APDM and PODS are not competing models; they simply offer operators options to meet the needs of their organization. Both the APDM and PODS models have been specifically designed for the pipeline industry and each is appropriate for deployment in a transmission or gathering environment. While gathering can be integrated, each model requires that gathering pipelines be stationed. Each model uses linear referencing and absolute coordinates to locate features. The different models both have formal administrative and development committees, established user communities, and have representation from pipeline operators and solution vendors. Both PODS and APDM can be utilized by a GIS system and each organization subscribes to the philosophy that an operator should be able to select from a group of appropriate vendors to provide them with a solution, and that the operator should be able to integrate products from more than one vendor into a solution.

While there are similarities between the models, there are significant differences. PODS is a certified (industry approved, standards body approved) standard that seeks to define in rich detail the features that describe a pipeline. The entire PODS tables structure is meant to be implemented 'as is' and this is what comprises the standard. The operator must have a good understanding of their business needs and the model or, as a standard the model could dictate how the business operates rather than allowing the business to dictate the model.

APDM is a template with a standard core. The standard core is not approved by any industry body but it must be implemented as described in order to promote interoperability between database implementations and in order for business partner software to function. The remainder of the APDM model is a series of best practices which are optional but if selected must be implemented according to a set of rules, formally described as the APDM Abstract Classes. APDM provides an operator with a lot of flexibility but it is open-ended and it can be difficult for operators to clearly understand the requirements, define their specific needs and implement the model.

While the PODS model attempts to describe in detail all the features that define the pipeline, the APDM model seeks to describe the 'response' or 'behavior' of pipeline features as they are edited, or more importantly as the underlying centerline is edited. Both organizations provide support for their model. PODS offers training courses, APDM provides workshops, and both hold open meetings where users can share information and discuss future changes.

APDM is an ESRI Geodatabase, and therefore it explicitly uses ESRI's linear referencing and topology data structures and technology, and has GIS "built into" the model. Its implementations are standard and generally involve the same types of technology. PODS is built to run as a Relational Database Management System and other than storing linear referencing and coordinate information in tables, there is no inherent GIS functionality stipulated for PODS. As an open industry standard it gives pipeline operators the ability to choose the most appropriate GIS application for their organization. PODS allows for a broad spectrum of implementation styles (GIS platforms, levels of integration with GIS). PODS users have implemented ESRI, Intergraph, GE Smallworld, and other platforms.

APDM follows the service oriented architecture (SOA) model for enterprise integration as outlined by ESRI. The SOA provides a modularity of pipeline business logic, which can be presented as service for other enterprise clients. These services are loosely coupled where the interface exists in the enterprise application and remains completely independent of the service layer. APDM does not support tight integration because it can be hazardous to the integrity of the geodatabase. With PODS, the operator can use the SOA through the RDBMS client or PODS can also support tight coupling using foreign keys or related records.

The future may provide another option for operators to review with respect to PODS and APDM. The organizations are currently working to develop a spatial implementation of PODS using ESRI technology. While PODS will continue to work with other spatially enabled software products, the intent of this project is to merge aspects of each model into a single model that calls upon their individual strengths, thereby creating the PODS ESRI Geodatabase. The concept of this model was proposed in early 2007, but it is still in its formative stages.

CONCLUSIONS

Data models are critical to the pipeline industry as they allow users to contain, integrate, and analyze data faster. This need becomes more predominant as pipeline operators implement new technologies that are making more and more data available. Currently there are two main data models available; PODS and APDM. Different business requirements for each operator lead to different solutions. Only by understanding the needs of the organization and the options available can an operator select the model that best suits their organization.

BIOGRAPHICAL INFORMATION

Robert Brook
Pipeline Industry Manager
ESRI

Specific Responsibilities

Joined ESRI in February 2007. Responsible for exposing ESRI to the pipeline industry and lobbying internally for the communities needs. This is coupled with the creation and engagement of ESRI's pipeline marketing campaigns, and conference and public speaking engagements. Rob acts as the global point of contact for existing and prospective ESRI pipeline clients.

Past Experience

Rob has previously held the position of business development manager for New Century Software, served in an oil and gas industry management role with ESRI Canada, managed Dillon Engineering Consultants GIS group, was a partner in EMLAR Environmental, and held several international positions.

Educational Information

B.Sc. - Geography, University of Calgary

Professional Memberships

GITA
Pipeline Open Data Standard
ArcGIS Pipeline Data Model

BIOGRAPHICAL INFORMATION

Sheila Wilson, Ph.D.
Executive Director
PODS Association

Specific Responsibilities

Dr. Wilson joined the PODS Association in 2006. Her primary responsibility is to meet the needs of the members by coordinating between the Board of Directors, the Technical Committee, and the members. She is the Technical Committee liaison to the Board of Directors and addresses technical questions presented by the members. She also organizes the annual PODS User Group Meetings, coordinates the efforts of working groups, and oversees the day to day operations of the association.

Past Experience

Dr. Wilson has been in the pipeline industry for approximately five years with more than ten years experience in GIS. She was a Senior GIS Analyst for the integrity management group at CITGO Petroleum. Her primary responsibilities included pipeline risk assessment data collection and analysis, analysis of inline inspection reports with respect to HCA's, and data collection and maintenance on the pipeline GIS. Previous to CITGO, Dr. Wilson worked at ONEOK, where she converted the legacy pipeline data into a GIS from Autocad and alignment sheets.

Educational Information

B.S. – Mathematics & Education, Northeastern State University, Oklahoma
Ph.D. - Geology, University of Tulsa, Oklahoma

Professional Memberships

GITA
Geological Society of America
Petroleum User Group

BIOGRAPHICAL INFORMATION

Peter Veenstra
Director of Product Development
Eagle Information Mapping

Specific Responsibilities

Peter joined Eagle Information Mapping in October, 2006. He is responsible for leading a team of application programmers in the development of core software for exploration and production and pipeline transmission industry. Peter's role encompasses drafting functional specifications, architecting software solutions, improving software processes and implementing software solutions at client sites. Peter is currently the Chairperson of the APDM Technical Committee.

Past Experience

Peter has 12 years experience performing GIS software and database implementations both domestically and internationally. He previously held the position of Software Architect with GE Pipeline. Peter served as a senior application developer and GIS consultant with M.J. Harden Associates. Before his tenure with M.J. Harden and GE, Peter worked as a GIS specialist with Black and Veatch Engineers.

Educational Information

B.E.S. - Geography, University of Waterloo, Ontario, Canada
Diploma – Geographic Information Systems, Sir Sandford Fleming College, Ontario, Canada
M. Sc. – Geographic Information Systems, University of Edinburgh, Scotland, UK.

Professional Memberships

Pipeline Open Data Standard
ArcGIS Pipeline Data Model