BIOGRAPHICAL INFORMATION

Peter Veenstra
Chief Engineer GIS Technologies
MJ Harden /GE Energy
1019 Admiral Blvd.
Kansas City, Missouri, 64106
Telephone: (816) 889-1153
Fax: (816) 889-1300
E-mail: peter.veenstra@ge.com

Specific Responsibilities

Mr. Veenstra joined MJ Harden/GE Energy in 1996. Mr. Veenstra is primarily responsible for the development of software applications and data management solutions for the pipeline industry. His main duties include creating the vision, architecture and design of these products including determining how the current needs of the pipeline industry and GE Energy customers fit into these products. Mr. Veenstra is also responsible for research and development of advanced GIS technologies, data models and processes for improving the current suite of software products and processes. Currently, Mr. Veenstra is serving as the chairperson of the ArcGIS Pipeline Data Model (APDM) technical committee.

Past Experience

GIS Applications Programmer\Developer, MJ Harden, 2000-2002
GIS Consultant, MJ Harden, 1996-2000
GIS Analyst\Project Scientist, Black and Veatch, 1994-1996
CAD Draftsman, Golder Associates Inc, 1994

Educational Information

M.SC. – Geographic Information Systems, University of Edinburgh, Edinburgh, United Kingdom, 1994
GIS.Dip. – Geographic Information Systems Diploma, Sir Sandford Fleming College, LIndsay, Ontario, Canada, 1992
B.E.S. – Environmental Studies (Geography), University of Waterloo, Waterloo, Ontario, Canada, 1985-1991

Professional Memberships

Chairperson Technical Committee, ArcGIS PIpeline Data Model (APDM)

GIS AND DATA MANAGEMENT – WHY IS DATA MANAGEMENT IMPORTANT?

Peter Veenstra
Chief Engineer GIS Technologies
MJ Harden /GE Energy
1019 Admiral Blvd.
Kansas City, Missouri, 64106
Telephone: (816) 889-1153
Fax: (816) 889-1300
E-mail: peter.veenstra@ge.com

Craig Wilder
Project Manager GIS
El Paso Corporation
1001 Louisiana
Houston, Texas, 77002
Telephone: (713) 420-2325
Fax: (713) 420-1997
E-mail: craig.wilder@elpaso.com

ABSTRACT

Although established as a mainstream technology the concept of enterprise GIS remains elusive to most organizations. One reason for this is the need for a clear vision of what effective enterprise data management entails. This paper outlines the concepts of GIS data management as an enabling technology for pipeline operators by answering the following questions: Why is data management important and how does it benefit an organization? What does it mean for an organization to manage properly conceive of and implement a GIS management system? Finally what are the possibilities of leveraging GIS data once this solution is in place? Proper GIS and data management depends on implementing the right technology and capturing and digitizing the organizational and procedural elements of the pipeline business. Once in place these elements form the basis for automating the data capture, data editing, and data visualization components of the system which generates return on investment for the organization as a whole.

INTRODUCTION

Managing accurate descriptions of facilities and operations data for a large transmission pipeline that stretches for hundreds if not thousands of miles (kilometers) presents challenges that can be met by Geographic Information Systems (GIS). Typical transmission pipelines span long thin corridors of land crossing over many jurisdictional and other types of boundaries. Without a GIS, examining such data as a set of tables, analog alignment sheets, disparate CAD drawings, charts and reports does not allow the user adequate options for integrated data visualization, database queries, and map-based network analysis. GIS systems provide a link between the tabular world and the visual mapping or geographic world.

The operation of a pipeline can generate volumes of data. Making timely decisions based on that ever-increasing amount of data can prove to be a challenging task. And, because those decisions are becoming of strategic importance to an organization, that data is ever more paramount to an enterprise's information management strategy. The challenge for transmission pipelines goes beyond just a matter of volume: interpretation and proper analysis of data requires dealing with seemingly unrelated data. The pipeline assets may be located from the latest construction drawings while analysis is performed from individual inspections without absolute locations. How can that and other data be communicated, shared, and acted upon as a cohesive system through the lifespan of a pipeline? Part of that challenge can be met through a uniform and consistent approach to GIS Data Management.

A big challenge for any organization holding large sets of data is how to effectively share that data amongst the various groups that require access to it. Storing data in database tables adds value but does not provide the interpretative power that visualization through a map affords. However, simply digitizing a map and storing it in a flat file (such as a CAD drawing) does not provide flexible access to the data describing the features on the map. GIS Data Management systems provide a link between the tabular world and the visual map or geographic world. It provides an efficient way to electronically store data with an added ability to visualize and perform spatial queries. The underlying technology of a GIS system is the relational database. That means that attribute data for valves – valve type, manufacturer, and installation date – are stored in valve tables and so on but with a spatial component (i.e. where it is located). That spatial component, often a geographic coordinate, allows the visualization of data within a map. It also provides the ability to accurately understand where features are in reference to other features allowing the creation of location map of all the valves installed prior to X date by Y manufacturer as an example. This sort of spatial query capability of GIS provides quick and easy interpretation of lots of data.

WHAT IS GIS DATA MANAGEMENT?

"Pipeline GIS Data Management represents a set of technologies, organization and processes that revolve around creating and managing geographic-style mapping data for the purpose of supporting the business objectives of a pipeline operating company". This definition is offered in replacement to the traditional definition of a GIS which describes the system only and not its role within the enterprise. The idea is to evolve the idea of data management from being a tightly managed, stand-alone system that is domain of a small enclave of highly technical experts to something that supports the requirements and business processes of the entire enterprise. Organizationally many existing GIS systems currently exist that support only one facet of a pipeline business such as engineering, mapping, planning or integrity. Given recent governmental regulations imposed on pipeline operators there are more requests for information from previously uninvolved agencies within a pipeline company prompting the GIS data management system to effectively support the entire enterprise.

The three elements of an effective GIS data management system can be described as a triangle with each element (technology, organization and processes) located at a point on the triangle. Each point in the triangle represents a vital component to constructing a GIS data management system. The technological side of the equation involves the purchase, implementation and maintenance of integrated and compatible computer software, scalable computer hardware and integration with other enabling technologies such as local area networks, the Internet, document and work-order management systems and databases. The organization part of the triangle represents the objectives and organizational structure of a business, which are used to define the requirements of the GIS data management system. Lastly, the process part of the triangle describes in detail how the GIS data management system will be used in the daily operations of the business to satisfy tangible business requirements and daily operations.

The triangle view of a GIS data management system defines the overview of the entire system. The organizational and process components of the triangle drive how the GIS will be structured (from a database design perspective) and maintained (from a workflow perspective). If every business group of a pipeline company (e.g. marketing, planning, design, field services, facility management, engineering, operations and business analysis) are involved in the conceptualization and design of the system then the organizational and procedural elements of the GIS will be effectively captured. The technical aspect of the GIS data management triangle must provide a specific set of baseline out-of-the-box software functionality without customization or additional coding efforts. The following sections of this paper concentrate on some of the technology requirements of a GIS data management system.

Secure and Transactional Multi-user Editing

Accurate and precise data that encapsulates the operational reality of a pipeline company will become that organizations greatest asset. Given the cost of data creation and conversion proper maintenance of such data represents a high priority. It is imperative that these data are created, maintained, preserved and archived in a secure environment. Security not only entails restricting access rights for creating, viewing, updating and deletion of data to specific groups of users but also is the ability to perform transactional edits of data. A transaction is a unit of work without influence from external sources and once completed is recorded and cannot be undone. Transactions involve an encapsulated manipulation of data that could involve complex workflows that may remain in process for lengthy time periods. At the end of the transaction the data is deemed sufficiently accurate for integration with the enterprise data store and is updated to the master database. A transactional environment protects the enterprise data store from accidents, incompleteness, or experimentation while providing a mechanism for daily editing and updates. The system must provide access to the latest valid information in a timely fashion as requested by a myriad of users. Using transactions in a secure environment provides accurate data that is regularly updated to the master database as the transactions are complete. Multiple users can access this information while data is being edited in a real-time environment.

Integrated Geographic Features

Most commercial GIS data management systems represent geographic phenomena (such as pipes, property parcels, valves, environmental areas) as vector-based geometric features. However, organizationally the geometry of the feature and the attributes describing such features are maintained in separate storage mechanisms or databases. The management of the geometry and the attributes of a feature in separate systems create a potential for data integrity issues where the geometry and attributes of a single feature are not congruent or synchronous. The creation, update, and deletion of a feature must be maintained in a single operation and ideally in a single data storage mechanism. A GIS data management system is designed around the graphic depiction of such features on a computer screen or report and thus the maintenance of geographic features (their geometry and attributes) represent the primary responsibility that system. At the heart of the system the user must be able to select the geometry and view the related attributes and vice versa must be able to view or query the attributes of a feature and be provided with a graphical (map) view of the geometry of the feature. The geometry and attributes of a feature are intrinsically inter-related with each other and are maintained in conjunction with each other. This aspect is particularly important to the pipeline community who primarily locate features by some form of Linear Referencing.

Archiving, Auditing and Retrieval

Pipeline operators are increasingly required by regulations to be able to report on and maintain inspections for elements of the pipeline that are no longer under operation or have been historically altered from the current state of operations. A GIS data management system must provide a mechanism for archiving historical data for later retrieval and display. Historical data might comprise abandoned pipeline routes or conceptual plans for new pipelines. This data may also comprise pipelines that have been decommissioned awaiting reactivation. The GIS management system should be able to provide a history of each feature that is currently active in the pipeline system. The history of each feature should similarly be linked to events or activities that occurred to the features of the pipeline within the operations of system (such as inspections, work orders, maintenance operations, repairs etc.). In this manner the lifecycle of each feature can be captured for further evaluation for planning purposes Both archived and auditable data within the pipeline system must be retrievable for reporting purposes. Ideally this information will be tightly integrated with the query interface of the GIS to allow users direct access to archived and historical information.

Access to the Latest Information

The heart and soul of a GiS data management system is to provide the latest, most accurate information to the user community on demand. The GIS data management system must return this data via ad-hoc user interface based queries, batch query, and pre-defined reporting and mapping processes. The system needs to provide data directly to the user through the recombination of data elements using industry standard data access protocols and methodology without the need for custom application development. The primary requirement of the system is to provide easy access to the latest information.

Industry Standard Protocols and System Integration

Successful GIS data management systems utilize industry standard technologies and open technology standards. Closed or proprietary systems do not allow easy integration with other systems. Often these systems require extensive custom application development to facilitate communication with other systems. Over the long term these systems incur substantial maintenance and financial expense. A GIS system that employs established industry standard protocols and technologies has a greater potential for success since these technologies are supported by major technology vendors with established user communities that expect a high level of quality and performance. Utilizing open, standardized and established technologies creates a system that can be extended through the wide spectrum of add-on vendors and service providers with proven implementation track records. Proprietary or customized solutions do not benefit from the latest technologies and are usually the "brain child" of a small group of people representing a significant intellectual property risk to an organization. Open architecture systems are commercially available and have large user communities who can share experiences and technologies to a wider community. The support of open architected systems is easier and cheaper from a technology perspective than custom systems and is scalable to meet the growing needs of the enterprise.

Data Structures and Data Models

A GIS designed to store pipeline features will utilize absolute coordinates and relative positioning (Linear Referencing) along a centerline to locate geographical features. Both positioning methods are accounted for in the current industry standard pipeline data models (PODS, ISAT and APDM). The underlying technology of the GIS management system must manage both methods for locating and storing features.

Pipelines typically span long thin corridors stretching for hundreds if not thousands of kilometers and crossing over many jurisdictional and environmental-related boundaries. Different features of these pipelines can be described and stored in very different manners – as absolute geographic coordinates or as some relative distance from other features on the pipeline. An appropriate reference system for pipelines and similar systems is linear referencing. This approach allows all features to be described as a distance along the length of a pipeline. In addition, features not physically located on the pipeline can also be described with an additional "offset" value representing the perpendicular distance from the pipeline. Regardless of where a feature is located in the geographic world, if it is known relative to the datum or any other linear referenced feature, it can be described in the linear referencing system. In this manner, seemingly disparate data with very different accuracies can be similarly described, stored, and subsequently analyzed. Combining this different data for easy retrieval and analysis is a matter of describing existing and any new features in the same common linear referenced system. To accomplish this, existing pipeline features are first described via an established linear reference system. Any new data is then related to features already known in this system. For example, in-line inspection data contains features with locations described as wheel count or odometer readings. Some of those features, fortunately, refer to valves, taps, tees and so forth that are likely already known in the established reference system. Relating these features and proportionately adjusting those in between (also known as "aligning" data) allows it to then be stored in the same common system. The same can be done for above ground surveys and other inspection data. A GIS system provides the technology, data structures, data model and processes for capturing, managing, maintaining and analyzing these kinds of data in a relatively simple map viewing environment which is readily accessible to many different kinds of users within a pipeline organization.

Once the technology has been established for managing spatial data using both absolute and relative positioning then these data are stored in a data model. Ideally the data model meets two primary criteria. Firstly, the data model is designed to store data in a fashion and manner the meets the operational and procedural requirements of the business (the other two points of the GIS data management triangle). If the data model does not meet the requirements of the business then no return on investment is gained by the organization. Secondly, the original format of the data model must be derived and be in compliance with one of the industry standard pipeline data models. Industry standard data models are designed for the pipeline industry by collaboration between pipeline operators and vendors. Standardized data models provide a mechanism for documenting the content and structure of the data model to provide interoperability between data conversion and application development vendors. Ideally, any vendor writing software solutions for a standard pipeline data model should be able to plug their software products into an operators data management system with a minimum of configuration and customization. Designing GIS systems that supports both forms for spatial positioning and in an industry standard data model stands a better chance of success from a technological stand point.

Enabling Technologies

A well designed GIS data management system utilizing open technology standards and data model design allows incorporation with other enabling technologies and processes. These technologies and processes include, but are not limited to, global positioning system (GPS) devices, hand-held field data collection devices, database replication (across local and wide area networks), remove data access (through Internet, Intranet, remote login), integrity management systems (calculation of risk, integrity management planning, feature assessment), cathodic protection systems, document management systems, engineering design systems, network flow analysis, and real-time system monitoring (SCADA). It is not inferred that a GIS data management system will simply "plug-and-play" with such systems in a commercial-off-the-shelf (COTS) environment. It is suggested that a GIS data management system incorporate the technology and database design that would make integration with such systems simpler by providing the required 'hooks' (such as universal primary key values, standard query language, TCP/IP, relational database management technology, extensible markup language etc.) make such integration easier.

WHY IS DATA MANAGEMENT IMPORTANT?

GIS data management is important to an organization for the following three reasons: the recent exponential increases in the amount of data collected and maintained, increased regulatory pressures, and increased user demands within an organization.

When the original pipeline systems were surveyed in the late 1800's and early 1900's digital technology was not available to pipeline operators. The extent of data management was limited to paper (analog) reports, design schematics and maps. Within the last twenty-five years the advent of digital data storage collection and storage mechanisms has caused an exponential increase in the amount of data that can and is stored. Not only has the quantity of data increased (from thousands of records to billions) but the variety of data collected and required to collected (from facilities, to land records, to documents, to financials, to integrity management, and inline pipeline inspection runs etc.) has increased dramatically. The increase in data quantities not only stem from better collection technologies but also from the corporate landscape of the pipeline industry as pipeline companies acquire operational assets from other companies and are looking for return on investment. The result of which is an integration nightmare as companies struggle to maintain some standardized operational parameters for managing the influx of data from a series of disparate data management systems.

Increased regulatory pressure is forcing pipeline operators to develop digital GIS data management systems. The United States Government Department of Transportation Office of Pipeline Safety and the Federal Energy Regulator Commission has increased the rigor by which pipeline operators within the United States report not only on the location of their facility assets but also on the location of high consequence areas and the level of acceptable risk of operations within those areas. The government is requiring pipeline operators to provide proof of these requirements in digital format and provide an audit trail by which the results of this analysis were obtained. It is not inconceivable that pipeline operators could furnish such reports using manual tabulation and analysis methods. Managing this information in an integrated, open architected GIS data management system allows operators to produce repeatable and verifiable analysis results in a fraction of the time required to do such operations manually even if the data for such analysis are stored in disparate data storage systems. Despite the up-front cost of converting existing data into integrated data management systems the return on investment from being able to generate regulatory reports in an almost turnkey fashion is beyond measure. A well-implemented GIS data management system will also keep this data in an auditable, secured, and archival fashion for later analysis and comparison.

Data management is important because the average user wants easy access to the repository of the stored pipeline information. As technology continues to evolve and become more mainstream access to a GIS management system stops becoming the domain of the 'power user'. Pipeline employees and their employers are demanding that the information stored in the pipeline operators repositories become available to everyone from management on down to field inspectors. What once was the domain of the engineering and drafting departments have important business implications for employees in the following business areas: finance, accounting, customer service, field inspection, field survey, drafting, mapping, design, engineering, land rights management, marketing, planning, executives, environmental, one-call, real-time and network analysts. If all three elements of the GIS data management triangle are accounted for: operations, processes and technology a data management system could be devised, implemented and released (even if incrementally) to support the needs of the entire pipeline business.

THE BENEFITS OF GIS DATA MANAGEMENT

The real power and benefit of a GIS comes in its ability to re-combine data to make useful information in both map and tabular form. Various types of data is divided into different map layers such as facilities, land base, environmental data, hydrology, roads, and geopolitical boundaries. The features on these layers are depicted as geometric elements (such as points, lines and polygons), which are stored as X and Y (and Z) coordinate values. Features that are in proximity to each other, but are stored in different layers, can be overlain on top of each other, intersected with each other, merged together in union, and otherwise grouped to determine relationships. A GIS is able to generate maps based on combinations of these layers allowing for easier interpretation and visualization of data. Spatial data of this nature is often better presented graphically rather than a database report for analysis and query.

CONCLUSION

As GIS can be an integral part of the enterprise information management strategy, its relationship to other systems should be carefully planned. SCADA, financial, work management, and pipeline integrity analysis software can be enhanced with a GIS by utilizing the same data sources rather than maintaining them separately. That leads to what, where and how the data gets stored – all part of a process called data modeling. A solid design of a data model (whether new or an adoption of an existing data model) is necessary to allow proper integration of systems and efficient storage and analysis of data. A needs assessment exercise along with that inventory of related systems with which to integrate is a good first step. This will help you to determine what your GIS needs to be and how it will potentially store and manage pipeline data to be used by other systems. It's important to note that implementing a pipeline GIS solution can be done in stages to alleviate spikes in budgets and to mitigate impact on users. A staged approach can be carried out so that the areas in greater need are addressed first. This approach can also allow those important quick wins to be realized.

Pipeline GIS is being successfully used for pipeline management. It goes beyond a map that must be manually updated. Effective pipeline GIS allows management of all pipeline data through the life of a pipeline with an ability to easily visualize, query, and share information. A pipeline GIS solution can add value and save money by providing more efficient and accurate management of pipeline assets, allowing more informed and timely decisions, and removing duplicate and separately maintained data sources. And, such a solution is attainable particularly since it can be done in a staged approach.

REFERENCES

ArcGIS Pipeline Data Model White Paper (APDM), 2005, APDM Technical Committee

Harmon, J., Anderson, S., 2003, The Design and Implementation of Geographic Information Systems, Wiley

Integrated Spatial Analysis Techniques (ISAT), 2002, ISAT Data Model Book, Gas Research Institute

PODS Association Inc., 2001, Pipeline Open Database Standard - Release 2 Documentation, PODS Association Inc.

Thomas, C., Ospina, M., 2004, Measuring Up: The Business Case for GIS, ESRI Press

Tomlinson, R., 2003, Thinking About GIS: Geographic Information System Planning for Managers, ESRI Press