The Optimum Framework for Managing E&P GIS Data
The Optimum Framework for Managing E&P GIS Data

Author: Gene Rhodes, Principal Consultant, Information Management Services
Gene.Rhodes@Halliburton.com

Introduction
Geographical information systems (GIS) data have become an increasingly important data type to petroleum companies and to business in general. IT leaders recognize that an effective GIS data management strategy and governance framework must address both the short-term needs of current projects and the longer-term strategic business requirements of the enterprise. Timely access to accurate geographic information enables petroleum companies to plan asset and data acquisition, to safely design and construct facilities in some of the most environmentally sensitive regions, and to optimally conduct business operations. To create value throughout the petroleum lifecycle, geographic information data must be well managed, with a set of consistent processes and standards, and of known quality from capture to utilization. In short, like any other E&P data type, a data management framework with well-defined roles should be developed for geographic entities.

While developing such a framework may seem obvious to some, geographic data can present some unique challenges. Geographic data are typically the spatial representations of an organization’s corporate or “golden” data, which may be stored in several disparate data stores. Spatialization is the process of gathering geographical data from various sources and consolidating them into consistent, map-ready formats.¹ In many cases, the business owner of various geographical data are in different departments, but the actual management of these data may be the remit of a “data management” team who leverages a complex IT infrastructure managed by the organization’s IT department.

Moreover, the larger the E&P enterprise, the more likely that sheer data volume will pose a serious challenge. For mature basins, petroleum companies typically will manage tens of thousands of wells and corresponding wellbore data; in some extreme cases, millions of geographic data attributes must be spatialized. In these cases, it may take days or longer to spatialize all the necessary geographic information, requiring an optimized set of processes that utilize complex computing infrastructure to achieve acceptable performance.

Finally, large E&P organizations may have dedicated staff whose sole responsibility is to determine the appropriate tools and techniques for land surveying, remote sensing, cartography, geographic positioning systems (GPS) and other forms of earth-related mapping. This discipline, sometimes referred to as geomatics, geodesy or geodetic management, represents another stakeholder in the geographic data value chain. With so many stakeholders, a formal and accepted data management framework for geographic information is required.
Data Management

Firstly, what is a data management framework and how can it apply to geographic data? The most widely accepted description of data management is provided by DAMA International.

“Data management is the business function of planning for, controlling and delivering data and information assets. This function includes the disciplines of development, execution, and supervision of plans, policies, programs, projects, processes, practices and procedures that control, protect, deliver, and enhance the value of data and information assets.”

The overall data management function encompasses 10 major component functions.

1. **Data Governance** - Planning, supervision and control over data management and use
2. **Data Architecture Management** - Defining the blueprint for managing data assets
3. **Data Development** - Analysis, design, implementation, testing, deployment and maintenance
4. **Database Operations Management** - Providing support from data acquisition to purging
5. **Data Security Management** - Insuring privacy, confidentiality and appropriate access
6. **Data Quality Management** - Defining, monitoring and improving data quality
7. **Reference and Master Data Management** - Managing golden versions and replicas
8. **Data Warehousing and Business Intelligence Management** - Enabling reporting and analysis
9. **Document and Content Management** - Managing data found outside of databases
10. **Meta-data Management** - Integrating, controlling and providing meta-data

All of these component functions should be considered when developing a data management framework for geographic data, but several functional areas are of particular importance. These include data governance, data quality, data security, reference and master data management and, finally, database operations management.
GIS Data Governance

GIS data governance can be thought of as the policies, standards and rules that govern how geographical data are managed throughout the entire data lifecycle. Strong governance entails building standard, repeatable processes that protect the needs of data stakeholders. From initial capture/validation through use and access, consistent application of data governance "can enable better decision-making, reduce organizational friction, reduce costs and increase effectiveness through coordination of efforts." To be most effective, representatives from multiple disciplines spanning key constituencies who have a stake in the use and management of geographical data should work in collaboration to build and maintain data governance.

Successful governance should also define the ownership of different geographical data types. Data owners are usually business leaders or recognized subject matter experts (SME). The owner is ultimately responsible for the validity and completeness of data and defining what this means for geographical data. In larger organizations, data stewards may be appointed to verify data validity and completeness on behalf of the data owner.

Finally, governance should consider the business objectives, organizational structure and cultural norms of the enterprise, and this implies that good data governance will look different in different organizations. Even organizations that strive for consistent application of governance throughout the enterprise may need to consider the unique circumstances of smaller remote affiliates to avoid “one size fits all” syndrome.

Key Best Practice: Assign a business owner to key spatial data types. The owner should be a SME who defines the QC rules that determine if data are fit for use. Because spatial data is a representation of golden source data, the corporate data owner should generally own the spatial copy and work very closely with map specialists, or geodesists, to ensure the proper map projections are applied during the spatialization process.

Geographical Data Quality

Every organization desires data of high quality but what does that really mean? Data quality management is a set of activities that measure data conformance to specifics of use, as defined by the requirements of the data owner. In terms of geographical data, there are several aspects...
to remember. Spatialized data may not be the “golden” version of that data. In many cases, it represents a visual replica of the golden version that can be viewed on a GIS map. The most rudimentary quality control technique for geographical data is viewing the data on a GIS map. While certainly not exhaustive, gross errors with spatial data quality are usually quickly identified by data owners when data are viewed on a GIS map. Visual inspection is clearly insufficient for large volumes of geographical data or for detecting location data which are slightly inaccurate.

Errors with location data raise the question of the source of the error because spatialized data represents a copy of the golden data. Is the error in the underlying data or in the spatialization process? Spatializing golden source data requires quality control techniques to verify that the spatialization process produces a consistent version of source golden data. Furthermore, GIS data and systems also offer the ability to create quality control techniques not easily implemented in non-GIS systems. Geoprocessing techniques allow business rules to be written to verify objects fall within an expected area or volume. When data error associated with the spatialization process has been corrected or eliminated then the quality of geographical data will become a function of the quality of the source data.

✔ Key Best Practice: Develop QC techniques to verify that the visual replica of spatial data is consistent with the source golden data. Furthermore, verify that the spatialization process and QC rules successfully complete before making the data available for access and consumption.

The concepts of accuracy and precision are important when considering geographical data. Accuracy is how close a measured value matches its true or accepted value, while precision is the level of measurement and exactness for the value. Sometimes these terms are used interchangeably, which can lead to confusion. For example, data may be quite precise and yet be inaccurate, as anyone who collects location information may make mistakes or enter that information incorrectly into a database. Inaccuracies may also occur due to distortions inherent in projection systems as well as accuracy standards based upon map scales. The latter tends to be magnified with modern GIS systems that permit unlimited zooming which may create the false impression of improved accuracy.4

Finally, making GIS part of the mainstream interpretation and engineering workflows certainly has many benefits. Providing GIS functionality to end users of interpretation systems allows these users to leverage new techniques and a plethora of existing data in GIS formats. At the same time, the organization must be careful to enact procedures and process around the creation of GIS files in a project environment. Most end users have free reign over their project databases. Such flexibility is desired to enable end users to investigate and model different scenarios. GIS files such as shapefiles in a project environment should be created with attributes that clearly indicate use of the project data. Safeguards should be enacted so that project data in GIS data files never are mistaken for a replica of golden source data.
Key Best Practice: GIS data files created from project data should include attributes that clearly indicate the origin project, cartographic reference system and basis geography of the location data.

Data Security and Access Considerations
Data security management is the planning, development and execution of security policies and procedures to provide proper authentication, authorization, access and auditing of data and information. GIS systems provide another tool to access an organization’s data and in general should reflect the organization’s data and security and access policies. That is, users, departments or organizations that are entitled to specific data should in most cases have the same level of access to the data when using a GIS system to access the same data.

Most modern GIS systems make it very easy to create new layers of GIS data. Layers can be created based on attribute queries or spatial intersections. Care should be taken for any new GIS layers to inherit the access rights of the source layers to protect against potential unauthorized access. This can be particularly challenging when users are able to create and share GIS data files. Data in a file is no longer under database access control, and the only security is file-level permissions and end users’ discretion. This is not typically a robust security model; each organization needs to balance the obligation to secure its data assets and end user agility to collaborate with new and innovative methods.

Traditionally, most GIS data and associated attributes were copies of source data, and the challenge was to verify that source and copy remained consistent. One advantage of copies is that end users cannot change the underlying source data in a GIS system. The cost of verifying consistency came with an added level of security: unauthorized changes to golden source data were virtually impossible to carry out in a GIS system. Today, as more master or corporate data systems become spatially enabled, the potential for accidental or unauthorized access now must be accounted for. The latest GIS toolkits enable software developers to easily extend out-of-the-box GIS map viewers that in some cases permit easy edit and update of golden attribute data in spatially-enabled corporate databases.

Key Best Practice: Updates/edits to golden corporate data should not be allowed in off-the-shelf GIS products. Updates to corporate data should only be made using the tools specifically designed for the particular corporate data base technology. Furthermore, individual users should never be granted access permission to database tables and objects. Access should only be granted to roles, and users should always be assigned a role.

Reference and Master Data Management
Master data management is the control over master data values to enable consistent, shareable, contextual use across systems of the most accurate, timely and relevant version of truth about essential business entities. It follows that reference and master data management is the ongoing
reconciliation and maintenance of reference and master data. GIS data is a replicated spatial representation of an organization’s master data or corporate data.

E&P data owners and stewards are typically responsible for developing QC processes and techniques to verify corporate data’s conformance to standards as well as verifying the consistency of spatial replicas stored in GIS systems. They work closely with geodesists to implement the QC processes. The geodesists also focus on specifying the best map projections, datum and transformation algorithms to be used in the specific areas of operation.

The Optimum Framework for Managing GIS Data

A framework for managing GIS data should provide a mechanism to publish high quality spatial data from trusted corporate data stores and access to that data using tools preferred by the user community. The framework should be rigorous enough to support managing hundreds of thousands (if not millions) of spatial entities in large corporate and/or regional affiliates. It also should be able to scale down and meet the needs of smaller affiliates who may not have the full breadth of data, access to GIS and IT specialists, or as complex requirements as larger business units.

Roles within the Framework

To meet the framework objectives and provide the flexibility necessary to upgrade (or even swap out) major components, the framework should have several layers. This design will allow the ideal division of skills between GIS managers, map specialists, E&P data managers and IT systems and infrastructure administrators. While organizations vary between petroleum companies, in general, the division of skills is typically as follows.

1. GIS managers and geodesists typically are the maintainers of the mathematical definitions of map projections and preferred datum for a given geographical location. This specialized set of resources also develops geo-processing models and tools leading to new approaches in projection and QC techniques.
2. E&P data managers are typically the maintainer of E&P corporate data stores and work closely with the GIS specialist to develop techniques to verify that the replicated spatial data versions are consistent with E&P corporate data. This group also may represent the business owner of corporate data and help define and implement the data standards, rules and processes that comprise a data governance model for corporate data.

3. IT and systems infrastructure experts work closely with GIS specialists and E&P data managers to understand the performance requirements of spatializing vast quantities of corporate data and the storage requirements for both GIS and corporate data. Many companies also maintain their database administration expertise within the IT organization. The IT organization also maintains the proper versions of major system components and works with system stakeholders to plan systems upgrade, migration and release management.

**Framework Layers**
The base layer of the framework contains the corporate data, including the golden location data and its original map projection data. To minimize round-off errors, it is critical that spatialization always use the original measure location data in its original projection.

The next layer is the spatialization layer. This layer will contain the tools to spatialize corporate location data and store an interim version of the data. QC techniques are utilized in the interim version to verify that the GIS data-ready version is consistent with the corporate version and that the spatialization process successfully completed without error.

![GIS Data Management Framework](image-url)
The next layer in the GIS data architecture is commonly referred to as the geo-database. This is a high performance database containing the spatial replica of corporate data. This database must be carefully tuned to provide access to potentially thousands of GIS data consumers and must owner provide the proper level of access to entitled users.

The final layer of the architecture is the GIS data consumption layer. In this layer, entitled users access GIS data in a variety of tools ranging from web-based maps (perhaps from within a portal), to desktop GIS systems and even high-end 3D interpretation systems.

The current trend is to combine corporate databases and geo-databases into a single database. Essentially, corporate databases add a “spatial” attribute for each geographical data type that contains the shape attribute. The spatialization process in effect takes place when the data are loaded or when the location attribute is updated in the corporate database. This approach can simplify the framework and reduce data duplication. As mentioned earlier, this approach must manage user access very carefully to prevent unintended and unauthorized changes to corporate data.

**Conclusion**

Once considered a specialty tool used by experts, geographical information systems and geographical data have become ubiquitous throughout the petroleum lifecycle. A well-defined, role-oriented framework enables a single copy of geographical data of known quality to be easily shared in mainstream interpretation systems, in GIS desktop tools and over the intranet as web map services. By allowing each team to concentrate on what they do best:

1. E&P data managers can build and maintain a high quality corporate data store which protects the enterprise’s investment in data acquisition, interpretation and engineering workflows.
2. Cartographers and geodesists can focus their efforts on GIS tools, projection techniques and areal model building.
3. IT professionals can enable the enterprise to work efficiently and securely with large volumes of corporate and GIS data.

This approach allows E&P knowledge workers with varying degrees of GIS expertise to have seamless access to geographic data, which improves the E&P enterprise’s ability to make actionable business decisions.
References


