Enabling a Sustainable Spatial Data Management Strategy

Westheimer Energy Consulting

Spatial attributes are pervasive in energy data types. Energy businesses rely on maps to view their data more now than ever, especially in exploration and production. While maps have always been used for subsurface delineation of oil and gas reservoirs, more and more data that resides on a map has crossed over and is now used as the backbone for visualization of assets, holdings, routes, future expansion, exploration plays and other business decisions with growing financial impact. Spatial data is therefore essential in any overall Master Data Management Strategy (MDM). While the topic of MDM can generate either agreement or angst, sometimes both[1], we will accept for this paper the current definition in Wikipedia, “a set of processes and tools that consistently defines and manages the non-transactional data entities of an organization” and can include “reference data”. The emphasis on non-transactional data is interesting in light of ESRI's ArcGis version 10, and its continuing emphasis on base map layers and operational overlays[2], and some of the best practices developed for display and performance in this environment[3]. In general, the ESRI approach neglects the complexities of collating and concatenating these multiple data types and the challenges of aligning the full available resources to implement such a strategy at this level (Figure 1). A project based spatial data management strategy can be effective without an MDM strategy, but it must be data driven and needs to support business objectives and not simply deliver esthetically pleasing maps.

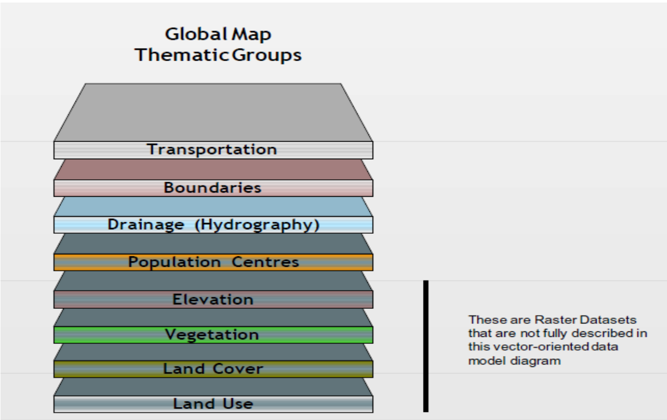

Figure 1: Some of the non-transactional map layers that comprise Spatial Master Data in ESRI's Global Spatial Data Infrastructure (GSDI).

Spatial data can be quite massive in both its volume and complexity, and represents a large part of the corporate data asset that often eludes MDM strategies. Industry standard site assessments that identify Data Management Maturity Model levels for spatial data show that map-based data has above-average numbers of disparate external data sources and working or operational datastores when compared with other oil and gas data types (Figure 2) but is frequently stored only in ESRI-standard original format types. In those same site assessments, independent and national oil companies were found to have higher levels of centralized and standardized spatial data stores than large multi-nationals, putting them higher on proposed GIS Maturity Models (Figure 3). The general oil and gas industry acceptance of ESRI's ArcSDE as the de facto corporate datastore for spatial data also means that more spatial data meets previously identified criteria for Data Management Maturity such as “centralized data management” and can be categorized at Level IV (Dynamic) or above.

True MDM for spatial data must also include knowledge, consistency, governance and sustainable processes in place to publish the data and associated business attributes, in order to achieve more robust data management maturity model levels[4]. In order to meet the observational criteria for Level IV (Dynamic) on the proposed GIS Maturity Model, management must recognize the benefits of spatial data in terms of operational efficiency, in decision making and in the quality development of services deployed to end users. The organization must have an articulated spatial information strategy with centralized management. Actually completing implementation of this strategy and regularly measuring the impact on productivity moves the organization to Level V (Optimized). Non-technical integration issues such as legal, social, policy and institutional constraints must also be considered in moving forward on GIS maturity models[5]. Some best practices evolving from studies of GIS maturity and Spatial Data Integration (SDI) include; use of time stamp meta-data with both cadastral and topographic data models to account for temporal variance, use of feature-level meta-data for efficiency when compared with theme- level meta-data, acceptance of virtual data repositories as a viable alternative to federated databases, visualization for error detection and quality control[6], and the use of spatial mediators and viewing agents[7].

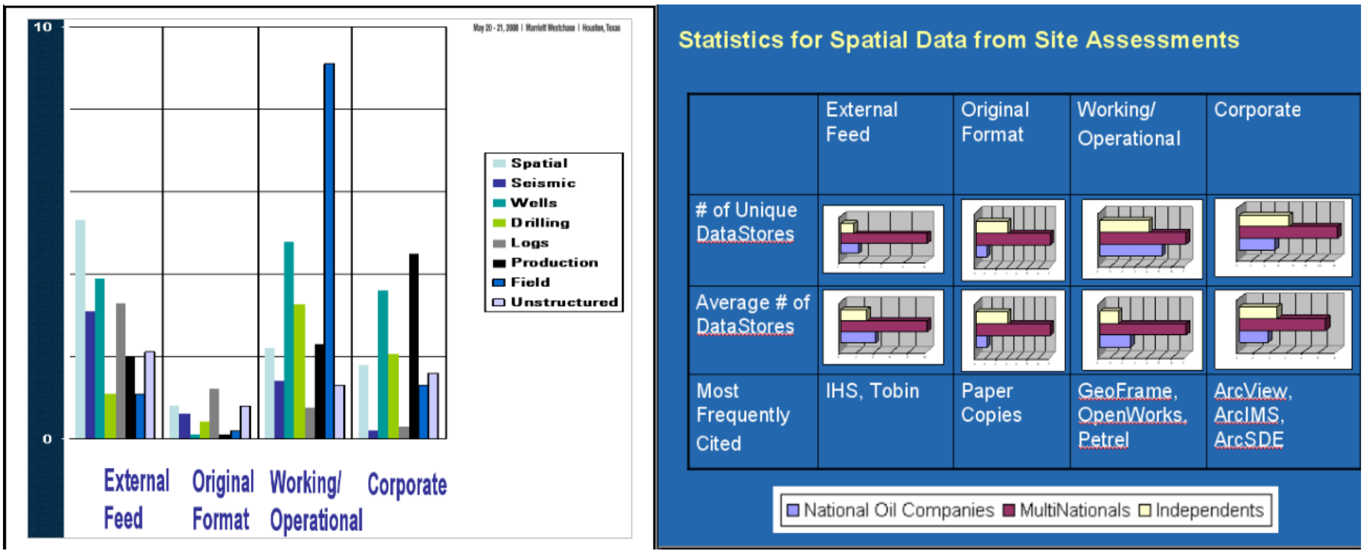

Figure 2: Average number of data stores at the four basic work flow stages for standard oil and gas data streams, and selected statistics from a compiled subset of standardized site assessments, showing the differences in spatial data handling between categories of oil and gas organizations. Note that in many cases the difference between working (transactional) and corporate (master) data stores for spatial data is not fully defined or understood by participants in a site assessment (From SPE Paper #116709, copyright SPE, used by permission of the authors).

Solutions that effectively leverage spatial MDM require a sustainable environment, data management landscape, and process ecosystem to ensure delivery of value to end users. The functional requirements of a sustainable spatial environment parallel those for other petroleum technical data types; enterprise level data accessibility, a standard data model, data access through web services, and business rules for data management, maintenance and quality assurance. A full life-cycle solution for management of spatial attributes includes discovery and collection from field and public domain sources, quality assessment and control, storage in standardized data models, distribution to analysis applications, and capture in knowledge management and audit systems. Effective implementation of the resulting matrix of functional and life-cycle requirements has been shown to have a positive financial impact on oil and gas and other energy and resource extraction organizations, including mining and carbon storage and sequestration.

Level	**Observational Characteristics**	**Resources**	**Technology**	**Process**
Level **V** - Optimized	Strategy is implemented into the organization's operations. Management regularly measures impact of spatial data on productivity and decision making.	Spatial data quality is included in performance KPI's	Data mining and expert systems	Process improvement measured by six-sigma data quality metrics
Level **IV** - Dynamic	Management recognizes benefits of spatial data in operational efficiency and decision making quality. Strategy implementation has started. Spatial data management is centralized.	Spatial data managers are embedded with asset teams	Automated decision support	Defined roles and responsibilities, management support for GIS data governance
Level **III** - Systematic	Spatial data used continuously in diverse business processes. No common data strategy or plan. Spatial data benefits are evaluated based on cost savings.	Best performers are leveraged on key projects	Integrated workflow solutions	Collaboration between business and data owners on functional requirements
Level **II** – Aware	Spatial data usage limited to obvious business processes, separate information systems. Data management is decentralized, managed separately in each application. Benefits not measured or known	Individuals develop their own processes	Unintegrated point solutions	Limited stakeholder involvement, scope creep on projects, business requirements from silos
Level **I** - Base	Coincidental and non-systematic spatial data usage, based on one or few people's own interest.	Capable individuals and heroic efforts	End-User Controlled applications (Excel)	Ad-hoc processes not standardized across locations or between individuals
Level **0** - Obstructive	Counter-productive spatial data processes are imposed by management, failure to allow successful GIS development, disregard for best practices in software engineering	Data management is penalized in job descriptions	Inappropriate tools (CAD)	Deliberate creation of data fiefdoms

Figure 3: A proposed Spatial Architecture Maturity Model (SAMM)

Despite occasional inattention to spatial MDM, most visual representations of Master Data at oil and gas organizations still involve a map view, often integrating Master Data from structured datastores with discovery and management of unstructured spatial data (Figure 4) from shared drives, Microsoft SharePoint portals, internal website and extranet sources[8] With studies showing that 60% to 85% of the data volume in an organization can be unstructured [9,10], but that it comprises 80% of the data used to support strategic decisions, the ability to identify spatial elements from documents and files using automated GeoTaggers will continue to become an important element of spatial MDM. At one point the volume of this unstructured spatial data was doubling every 3 months with the addition of over 7 million web pages a day, many with potential spatial content that can be included in business intelligence searches.

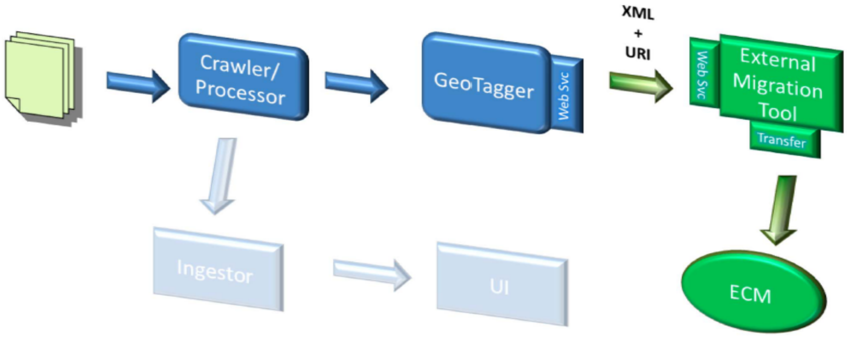

Figure 4: A planned approach to optimizing migration of unstructured data with automated geographic meta-data creation.

Spatial MDM gives end users confidence that spatial data on a map comes from a vetted MDM spatial repository rather than a series of poorly managed shape files, personal geodatabases, Computer Assisted Design (CAD) files, or aerial images without geo-referencing. Current technology such as ESRI Spatial Data Engine (SDE) and Oracle Spatial provide the ideal master published repositories for storing points, lines and polygons as they use intricate geometries that utilize the strength of RDBMS's, especially Oracle. Other data storage repositories which utilize E&P data schema's such as PPDM and Seabed(SLB), can store geometry data in tables, but still lack a methodologies to convert to true geometric objects (such as feature classes in ESRI terminology). E&P has adopted the ESRI SDE technology as it is tightly integrated to the ESRI ArcGIS Server technology, and enables clean indexing, synchronization and versioning. It also can run in smaller enterprise environments such as Windows/SQL Server. Oracle Spatial requires an Enterprise license infrastructure, and is obviously dependant on Oracle RDBMS. ESRI technology does, however, have the capacity to read Oracle Spatial SDO geometries, which is a big plus.

These enterprise level master Spatial repositories can address cartographic meta-data questions about datums and Coordinate Reference Systems, versions and other historical related data that many geotechnical end-users do not consider in their everyday work-flows. And spatial MDM can integrate and bring back non-spatial business data for well locations, seismic surveys, rights of ways or legal and land descriptions. Web based map visualization engines allow for "what if" scenarios and complex query capability to "slice and dice" spatial data and understand the information being used to make crucial business decisions. Ultimately the business and technical risk impact can be measured when a geotechnical end user is able to determine from a technically validated database (Figure 5), for example, if a seismic survey exists that can define a fault generated in an interpretation project[11].

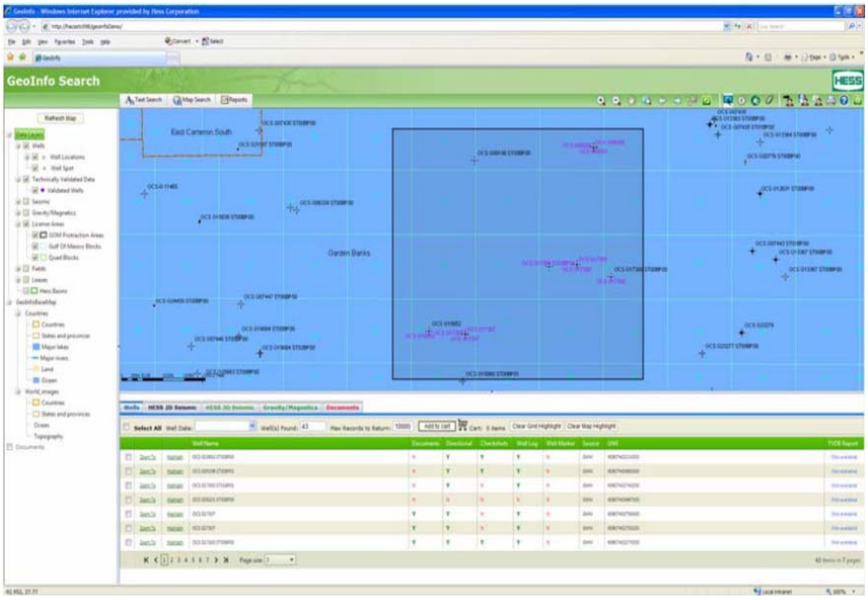

Figure 5: Example of ESRI-based map software for viewing and selecting geo-technical data from a technically validated database.

Spatial data is pervasive across every exploration and production discipline. Understanding the life-cycle, (source, Extract-Transform-Load (ETL) process, non-master repository to master repository work-flows, cyclical updates, interpretation, and derivation of data) is crucial to building confidence in the validity and currency of map data that impacts jobs, careers, and reputations. The impact of spatial MDM is magnified by the increasing volumes and complexity of the geotechnical data itself, and by the high dependency of energy organizations on technology driven work-flows[12]. Energy companies still struggle with the true value proposition for GIS, but sometimes find it in non monetary measures. GIS has value as a standard for inter-agency and public access to information, and public access to oil and gas spatial information is at an all-time high due to the recently raised profile of the industry as drilling continues to move into new onshore shale plays and deepwater.

An example of the implementation of GIS MDM can be found in BP's reports at ESRI conferences in 2007 that GIS protagonists were struggling to promote use beyond specialists and small, unsupported communities[9,10]. BP's deployment began with the adoption of ESRI's ArcInfo in 1989, but was mostly at a small scale with an unsupported, informal network. Without a federated approach, multiple versions of tools were in use and there was a lack of standards and mechanisms for data sharing. According to Keith Everill, formerly the information assurance manager for BP America's Gulf of Mexico Deepwater Exploration, "our approach was inconsistent... we had aligned our operating system tools like Microsoft Windows and Office, but there was no common strategy to govern the use of GIS tools that accommodated a common operating environment". Speaking at ESRI's Senior Executive Seminar (SES), held the day before the 2007 ESRI User's Conference in San Diego, he identified spatial data management issues and a lack of end-user take-up as reasons that BP had been slow to realize the potential value behind a robust exploitation of GIS. Too much time was spent on scripting monthly updates and the multiple versions of the software made it hard to get the value from GIS across discrete business units. A senior manager attending an interactive GIS presentation at a regular operational technology review meeting finally provided the spark and motivation for a high level strategic statement to govern and articulate the vision going forward for GIS, as well as the formation of a group to better understand the return on investment from GIS at BP.

BP's appraisal evaluated the business value proposition of an at-scale adoption of GIS tools, working on a supported set of best practices instead of isolated instances, and thus moving the recognition of value from Level II (Aware) to Level III (Systematic) on the GIS maturity model. A pilot team steering group set out to articulate the value of GIS in business terms across BP. The small team was formed from across all departments to hold workshops, develop an online survey for 150 global operations managers, and conduct interviews. Key findings were that the organization had been slow to realize the value of GIS due to a lack of alignment of people, process, technology and data, leading to disconnected solutions. BP found they had world class GIS access – but to disconnected data. GIS carried a substantial business value and was heavily utilized across the business lifecycle, including isolated applications for exploration spatial search, pipeline and subsurface mapping, environmental analysis, emergency preparedness, compliance, operations and hurricane response. Each of these applications had been built from scratch, usually in isolation along with its own database. These were trapped in electronic offices without the power to integrate or realize the value of an at-scale adoption. While executives up to group vice presidents acknowledged that geographic information was a critical underpinning of the E&P business and played a fundamental role in all decisions, the goal of obtaining the absolute ROI of GIS in a dollar amount became less important to quantify. The hugely powerful assertion that energy is a spatial business and that BP was a spatial company implied that everything they did had a spatial element; from managing assets, through retailing, petrol stations and human resources, and knowing where their offshore or onshore staff were at any one time (an unstructured spatial geo-tagging application was used at one point to track executive travel and spatial data search and filtering played an important role in building an intelligent index to a field study in East Texas). The study advocated for enterprise wide spatial data management based on leveraged GIS applied across the business to improve communication, integrate the business and facilitate decision making.

The team focused initially on the existing impact of GIS, then discussed integrating information systems and the need for a common operating platform. Since BP needed to enable both expert and casual users, recognizing the value of GIS depended on embedding GIS technology into core workflows. BP acknowledged the major impact and applicability of GIS across business assets, and identified the need for an understanding of the full business information lifecycle of the company. This strategy began to crystallize with the deployment of a GIS thin client for users and full GIS for experts, with attention to the full business lifecycle value proposition. Management became convinced of the value through analysis and discovery. Adoption became not about imposing applications, but focusing on individual behavior and demonstrating best practices. Eventually BP began to consider how others could use GIS, including those in marketing and gas processing operations. There were however still questions about confidence in data completeness from the land and lease community. The perennial 'iceberg' that is data management remained a hostile threat to be recognized and navigated.

In order to understand the impact of a compelling business event on the drive for Spatial MDM and shifts along the GIS maturity model, compare and contrast the previous 18-year adoption cycle and time to implement centralization and standardization with the recent deployment of the National Oceanic and Atmospheric Administration's (NOAA) GeoPlatform website in response to the Deepwater Horizon oil spill in the Gulf of Mexico[15]. At an Industry Advisory Council membership meeting on July 15, 2010 in Vienna, Virginia, USA, NOAA's chief information officer Joseph Kilmavicz, who is also director of High Performance Computing and Communications at NOAA, pointed out that the spill made "starkly clear" that there was a need to define an architecture to effectively manage, store and disseminate spatial data to the public and other agencies. The GeoPlatform Web site allows the public to track oil spill and recovery data online using a near-real-time interactive map and access to over 600 data layers from the Incident Command Structure, the U.S. Coast Guard, EPA, the U.S. Fish and Wildlife Service, the U.S. Geological Survey, the Homeland Security Department, NASA and other federal, state and local agencies. The website displays static response base map layers such as spill sensitive areas, lease blocks, sanctuaries, reserves, bathymetry, and parish and interstate boundaries. Dynamic operational overlay data is updated daily when available from sources including oil spill trajectories, assessment results, satellite and field photos, analytical chemistry and wildlife observations, fisheries closures, ship tracking, and environmental conditions such as wind, wave, tides and predicted precipitation.

The public site is currently updated twice daily. While the volume and complexity of data collected for the site is comparable to that of an operating oil and gas company, coming from many types of sensors and sources in the Gulf of Mexico and in different formats, GeoPlatform.gov was launched in ***two weeks*** with the help of geospatial executives from across the government and researchers at the University of New Hampshire's Coastal Response Research Center. The site is hosted on NOAA's Web server farm (Figure 6) and leverages the Environmental Response Management Application (ERMA).

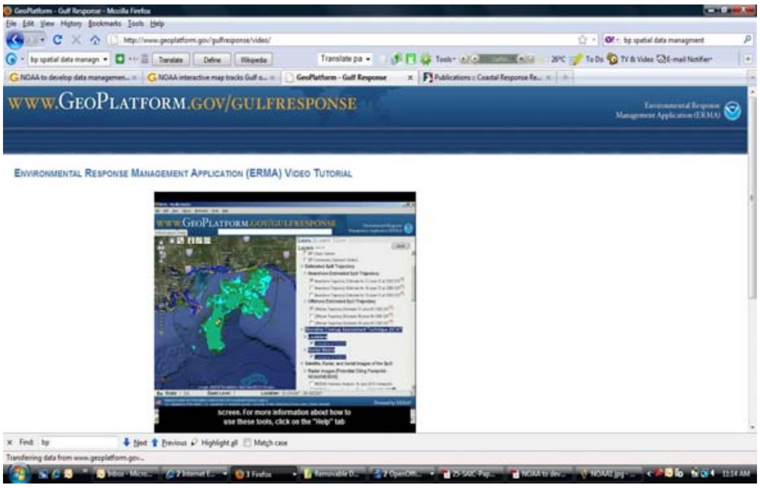

Figure 6. Oil spill response public website

While the near-term purpose of the site is to bring data together in a usable format, NOAA indicates with a degree of inevitability that there will be an evolution as they plan how to best position existing IT systems and data to prepare for the "next crisis". The ERMA solution is part of a broader U.S. Government mandate from the President's Fiscal Budget for 2011, which states that "in 2010 and 2011, federal data managers for geospatial data will move to a portfolio management approach, creating a Geospatial Platform to support Geospatial One-Stop, place-based initiatives, and other potential future programs. This transformation will be facilitated by improving the governance framework to address the requirements of State, local and tribal agencies, Administration policy, and agency mission objectives. Investments will be prioritized based on business needs. The Geospatial Platform will explore opportunities for increased collaboration with Data.gov, with an emphasis on reuse of architectural standards and technology, ultimately increasing access to geospatial data". Data.gov gives the public access to machine-readable datasets generated by the executive branch of the federal government, and according to NOAA, "all data can be geospatially referenced," and put it in a geospatial background for applications such as GeoPlatform. However NOAA also realizes that data management is critical and that bringing data together doesn't solve all problems. Of more and critical importance is how consumers use that information and the work still to be done in that area. The U.S government shares many attributes with oil and gas company management in seeing value in deploying an online mapping tool to facilitate communication and coordination among a variety of users, and providing fast and user-friendly answers with clarity and transparency.

This case study shows that strategies for management of spatial data at any level should depend on the quality of the data and the value of that data to the end user. Industry standard site assessments, a standard methodology for classification of processes and components, and comparisons across capability maturity levels have been used to compare solutions and value statements, resulting in a comprehensive set of best practices and lessons learned for a sustainable spatial data management strategy. One important lesson comes from the lack of end-user take up when IT becomes the focus and the value and use of data to the business becomes secondary. The true value of GIS can then often be obscured by well intentioned, but ultimately IT driven strategies.

Within a MDM context it is essential that the management of spatial data be undertaken with a full knowledge and understanding of the life cycle, the complexities of the multiple data types, their attributes and relationships. A key imperative is that such strategies are directed towards the benefit of the business and that it is the business that takes responsibility for the data it intends to use for interpretation and analysis. It is the business that will ultimately make significant operational and financial exploration and production decisions based on the data (Figure 7). If GIS as a technology is undervalued compared to its full potential, then it is this gap between technology, data and the business drivers that is the key cause. Technology has to be seen to deliver benefit, not just a pretty map. Through the focused, experienced delivery of these approaches, today's energy businesses that have yet to realize full benefits from their master spatial data and master data management will have the ability to advance within industry as competitive differentiators as opposed to falling behind. Technology has to be seen to deliver benefit and value, and the decisions made from maps need the highest accuracy and completeness. One of the resultant benefits is confidence for business decisions and spatial information shared with peers.

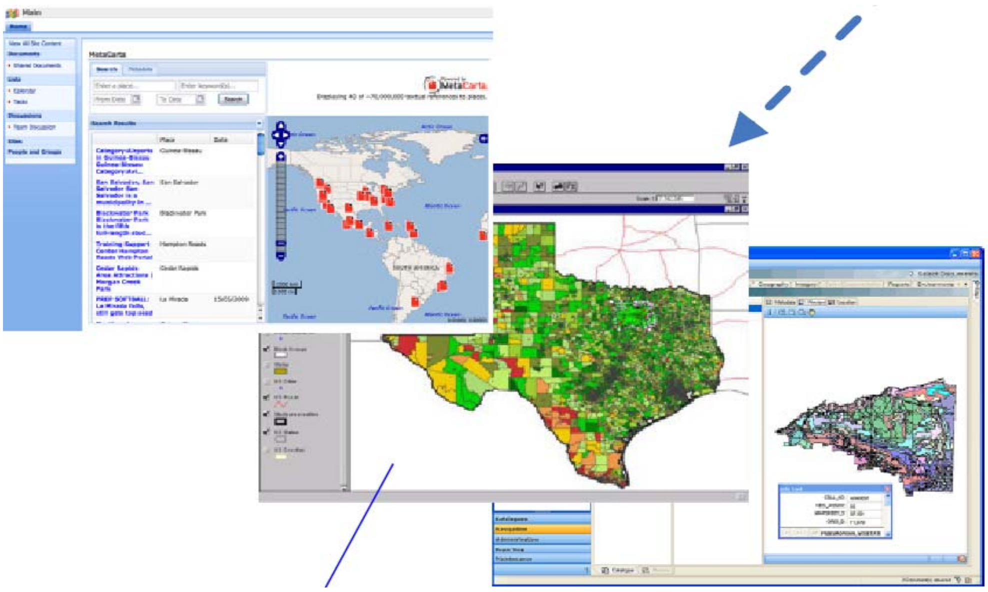

Figure 7: Examples of map views of structured and unstructured spatial MDM from an Enterprise Data Access project for decision support in the mining industry.

References

1) McNaughton, N., 2010. "Putt's Law and the data managers". Oil IT Journal, June 2010, www.oilit.com, accessed 5-Jul-2010

2) Environmental Systems Research Institute, Inc. (ESRI), 2003. "Basemap data model", ESRI ArcGIS Resource center, submitted 05/01/2003, accessed 07-Jul-2010, http://resources.arcgis.com/content/basemap-data-model

3) Jian, 2010. "Code Gallery; Maximum Map Extent", ArcGIS Resource Center, ArcGIS Server»Web Applications»Javascript APIs»ArcGIS, http://resources.esri.com/arcgisserver/apis/javascript/arcgis/index.cfm?fa=codeGalleryDetails&scriptID=16956, Last updated 06-10-2010, accessed 5-Jul-2010

4) Mäkelä, J., Vaniala, R., and Ahonen-Rainio, P.,"The role of user organizations in the implementation of National Spatial Data Infrastructure", Article under review for the International Journal of Spatial Data Infrastructures Research, Special Issue GSDI-11, submitted 2009-03-01

5) Mohammadi, H,, Binns, A., Rajabifard, A., Williamson, I.P., 2005. "Spatial Data Integration", Centre for Spatial Data Infrastructures and Land Administration, Department of Geomatics, University of Melbourne

6) Ulubay, A. and Altan, M.O., 2002. "A Different Approach to Spatial Data Integration", Symposium of Geospatial Theory, Processing and Applications, Ottawa, Canada, http://www.isprs.org/proceedings/XXXIV/part4/pdfpapers/443.pdf, *retrieved 17-Aug-2010.*

7) Miller, L.L. and Nusser, S., 2003. "An Infrastructure for Supporting Spatial Data Integration", National Science Foundation, Bureau of Census, US Department of Agriculture, Bureau of Labor Statistics, US Geological Survey, Iowa State University, http://www.fcsm.gov/03papers/Miller_final.pdf, *retrieved 17-Aug-2010*

8) Byrd, S., and Williams, J., 2010. "Optimizing Migration of Unstructured Data with Automated Geographic MetaData Creation, 14th International Conference on Petroleum Data Integration, Information and Data Management, May 18-20, 2010, Houston, Texas

9) White, C., 2005, "Consolidating, Accessing and Analyzing Unstructured Data". Beye Network, US Edition, Copyright 2004 - 2010. Powell Media, LLC. All rights reserved, retrieved 19-Aug-2010. http://www.b-eye-network.com/view/2098%20;%20http://www.themanager.org/strategy/Unstructured_data.htm.

10) Ponemon, L., 2008, "Survey on the Governance of Unstructured Data", Ponemon Institute Reserarch Report, Ponemon Institute LLC, Traverse City, MI, May 22, 2008, retrieved 17-Aug-2010. http://www.ponemon.org/local/upload/fckjail/generalcontent/18/file/Governance%20Unstructured%20Data%20052208.pdf.

11) Kunzinger, F., Haines, P. and Schneider, S., 2010. "Delivering a Data Governance Strategy that Meets Business Objectives,14th International Conference on Petroleum Data Integration, Information and Data Management, May 18-20, 2010, Houston, Texas.

12 Donovan, T., 2010, "The Synergy of Structured and Unstructured Data in a Geographic Environment". Schlumberger Information Solutions Global Forum 2010, London, U.K.

13) Everill, K., 2007. Defining the Value Proposition of GIS in Business, 17th ESRI Petroleum User Group Meeting, Houston, March 2007, reported in OilITJournal, TW0703_6.

14) Francica, J., A Governor and a Geologist Meet an Environmentalist and a Hospital Administrator. Directions Magazine, June 21, 2007, Directions Media, Glencoe, IL, http://www.directionsmag.com/article.php?article_id=2489 *retrieved 08-Aug-2010.*

15) Yasin, R., Spill prompts NOAA to define its data management architecture. Directions Magazine, July 15, 2010, Directions Media, Glencoe, IL, http://gcn.com/articles/2010/07/15/noaa-cio-kilmavicz.aspx, *retrieved 08-Aug-2010.*

Additional Sources for the Spatial Architecture Maturity Model (SAMM):

D'Angelo, T., Troy, B.: "Data management helps create value" The American Oil and Gas Reporter, July 2000

Finkelstein, Anthony, 1992, A Software Process Immaturity Model, ACM SIGSOFT Software Engineering Notes vol 17 no. 4, October 1992, Pages 22-23.

Mäkelä, J., Vaniala, R., and Ahonen-Rainio, P.,"The role of user organizations in the implementation of National Spatial Data Infrastructure", Article under review for the International Journal of Spatial Data Infrastructures Research, Special Issue GSDI-11, submitted 2009-03-01

*Odho**, M. nd** Ord, R., Understanding and Optimizing Your Firm;s Data Management Capabilities Using Maturity Models. Lakefront Consulting Data Management White Paper, http://www.lakefrontdata.com/style/LakeFront_Consulting_EDM_Maturity_Model_Jan_2009.pdf.*

Contributing Authors:

Michael Dougherty, Jess B. Kozman, Jeffrey Maskell, Thomas Ripley
Westheimer Energy Consultants

Copyright 2010 Westheimer Energy Consultants, a Limited Company Registered in England and Wales - Registration Number 7110859