

Applying DAMA to Oil Industry Data

Steve Hawtin, Schlumberger Information Solutions

14th Petroleum Data Integration, Information & Data Management Conference

The effective management of Exploration and Production (E&P) data has a major financial impact on a company's performance. Quality data can make the difference between a dry hole and a major discovery. For this reason an initiative to improve data management across a range of industries is bound to be of interest. When that standardisation effort is achieving widespread acceptance it becomes even more important to understand how it relates to the way oil companies manage data.

"Data Management International" (DAMA) is an independent association of information handling professionals. During 2009 DAMA published the first version of their "Data Management Body of Knowledge" (DMBoK). This extensive new standard is being widely adopted in other industries as a framework for performing data management.

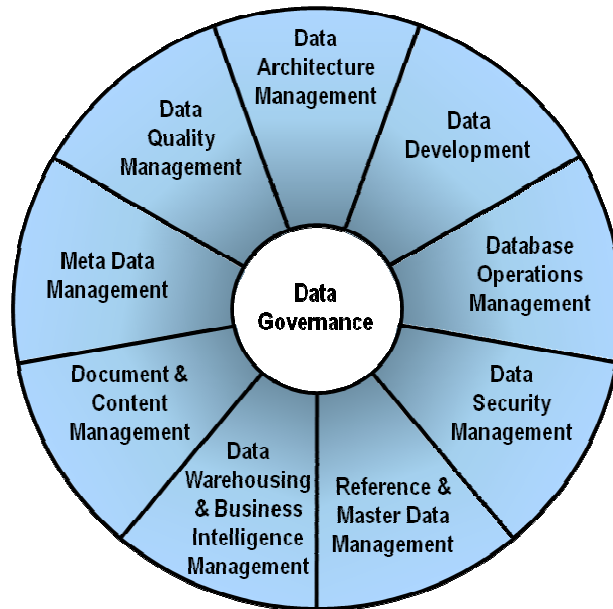


Figure 1: DAMA's 10 Data Management Functions

The DMBoK divides data management into 10 functions as shown above. This paper briefly explores each of these functions in turn and examines how well it maps in practice to the management of E&P data. The DAMA material provides valuable insights into implementing information management, however there are some areas where E&P industry norms differ from the generic approach that is assumed. These key differences are outlined and the underlying reasons for them are explored.

The final section reviews some key aspects of E&P data handling as it is implemented today. It explores the historical reasons for the differences between oil industry standard practice and the picture upon which DAMA is built. While this description will be familiar to those experienced in E&P data management it is hoped that this section will prove enlightening for those who are outside that environment.

DAMA

The “Data Management International”¹ (DAMA) is an independent association of technical and business professionals dedicated to advancing the concepts and practices of information handling throughout the business world. Globally there are more than 50 DAMA chapters. DAMA have created a number of resources for the data management professional including training courses and international conferences.

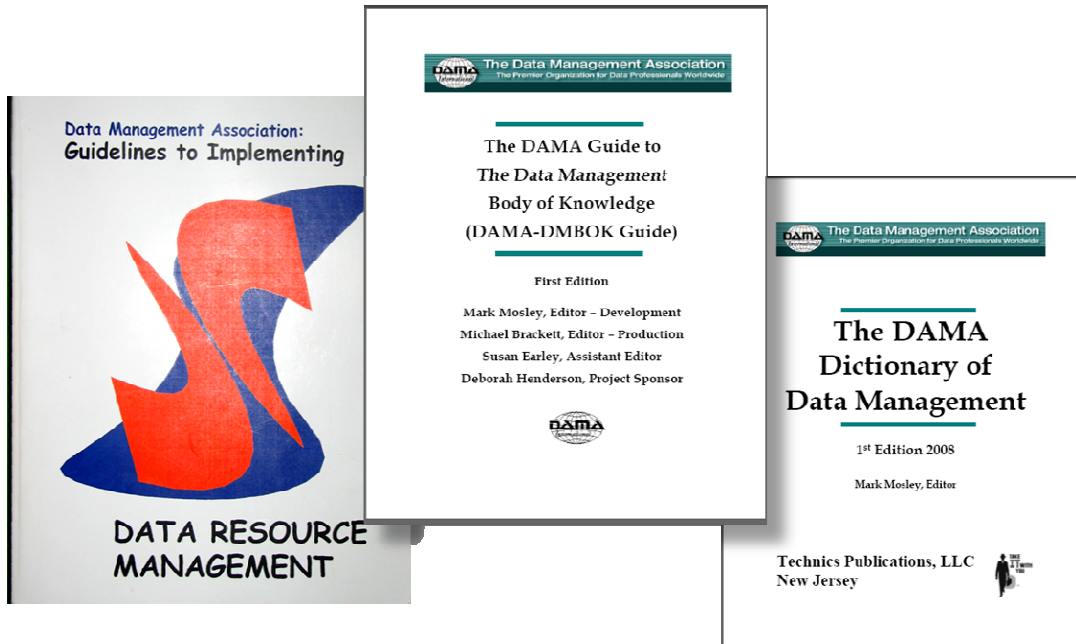


Figure 2: DAMA Publications

DAMA has been responsible for a number of valuable publications ranging from its series of “Data Resource Management” books to the “Dictionary of Data Management”. In 2009 DAMA published the “Data Management Body of Knowledge”² (DMBoK), its definitive guide to the discipline of data management as practiced across a wide range of industries.

The DMBoK is an extensive collection of knowledge arranged within a consistent framework.

¹ The DAMA web site at <http://www.dama.org/> holds an excellent collection of resources for data managers from all industries

² This paper relates to the first edition DMBoK published in 2009 as ISBN 978-0977140084

DMBoK

DAMA has divided the process of data management into 10 groups, each focused on a related set of functions and activities.

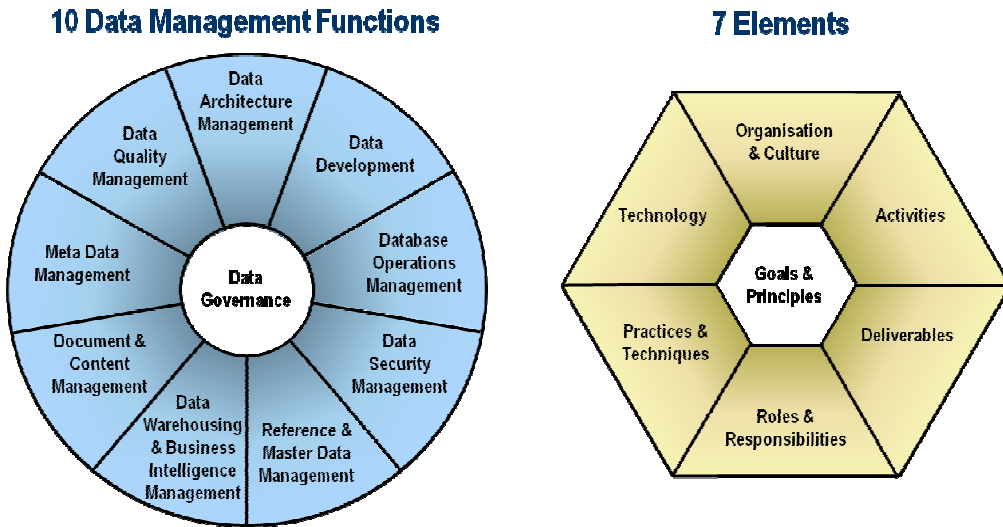


Figure 3: Organisation of the DMBoK

The 10 DMBoK functions are shown above. Each of these is then described by expanding on the 7 elements shown. This framework ensures that this complex topic is explored systematically and in a way that simplifies understanding.

Data Governance

the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets. Data Governance is high-level planning and control over data management.

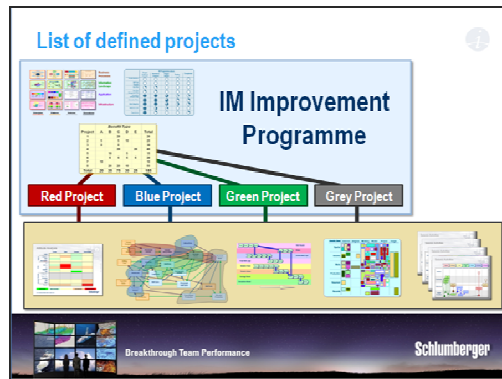


Figure 4: A coordinated programme of IM projects usually delivers greater business value³

Data Governance is the core function of the DMBoK framework, this is why it occupies the central location in the function diagram. DAMA considers that having a clear data management strategy managed by a group which control the resources is the single most important element in the whole process.

³ Slide from “E&P Data Assessments - Are They Worth the Cost?” presented at PNEC13 Houston May 2008

It is difficult to disagree with this view. The E&P companies with the most effective data management practices all have a well defined group or individual that can: coordinate data management; clarify the organisation’s policies; review the existing “Information Landscape”; select the most appropriate projects to implement; allocate resources as appropriate and measure the benefits that they deliver. This “Data Management Board” may be seen as part of the IT group or separate from it, in either case it is important that the topic is seen to have its own voice and that effective data management is perceived as a key element in the organisation’s success.

Data Architecture Management

defining the data needs of the enterprise, and designing the master blueprints to meet those needs. Architecture is an integrated set of closely related views reflecting the issues and perspectives of different stakeholders.



Within most E&P companies the current data architecture was not so much designed but rather emerged from the process of integrating multiple islands of isolated specialised data. So, in many companies this function is focused on documenting the way information flows in practice. Over the last 5 years a growing number of E&P companies have invested time and resources to document their existing information “landscape”. Only once a clear picture of today’s situation is available is it possible to create a complete vision for tomorrow.

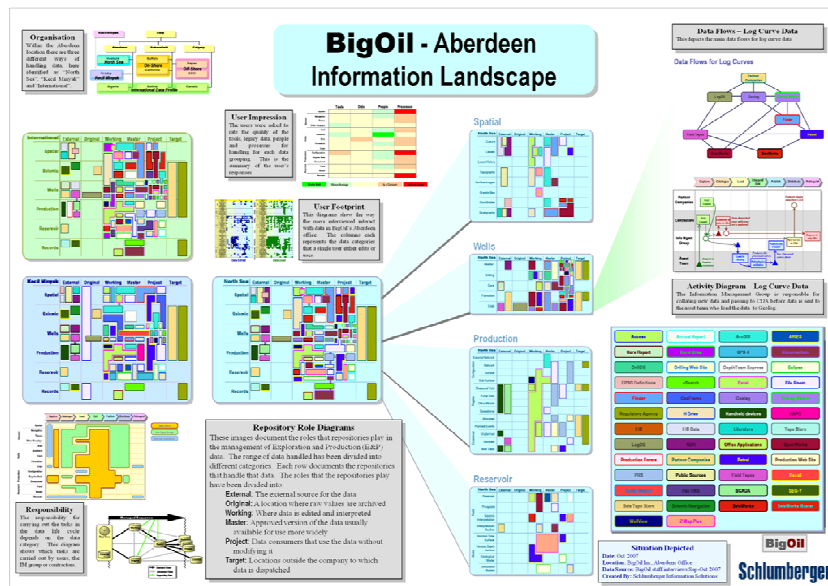


Figure 5: Various tools are available to document an E&P company’s data flows⁴

Tools and techniques for documenting the data architecture of E&P companies have been available for some time. The best of these are able to present the types of complex interactions upon which real organisations rely, and can illustrate the role of data management in the overall enterprise architecture. Once again the insights within the DMBok apply within E&P.

⁴ Picture from “Sharing a common view of the ‘Information Landscape’” presented at ECIM 2007

Data Development

designing, implementing, and maintaining solutions to meet the data needs of the enterprise.

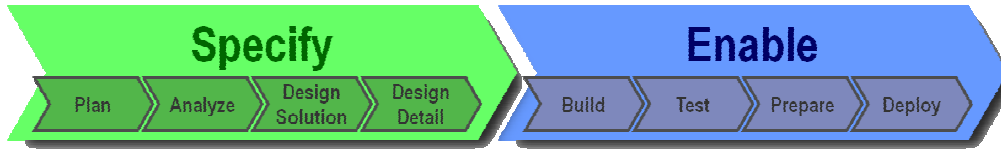


Figure 6: Typical system development lifecycle

DAMA’s description of this function focuses on the system development lifecycle. It emphasises data modeling, data requirements analysis, as well as the design, implementation, and maintenance of related solution components. Most E&P companies moved away from developing customised software some time ago, applications are bought not built. As a result the main concern for Oil Companies is interfacing between pre-existing systems, rather than influencing the way data is handled within a particular application or repository.

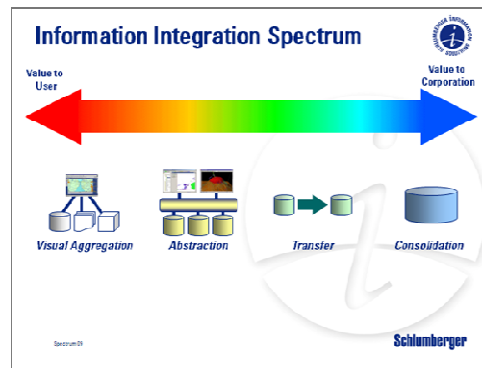


Figure 7: Available integration approaches for E&P data⁵

The E&P industry has long understood the importance, and difficulty, of integrating information. Where DMBok describes Data Development it has some valuable insights. It is less valuable for E&P where it explores data models, database technologies and transfer techniques, appearing similar to the discussions our industry had in the mid 1990s.

Database Operations Management

planning, control, and support for structured data assets across the data lifecycle, from creation and acquisition through archival and purge.



DAMA splits the operations management into two groups of activities, “Database Support” and “Data Technology Management”. These section titles might appear to emphasise the Database Administrator’s (DBA’s) role in the management of data. However the body of knowledge not only describes the more traditional DBA responsibilities, such as tuning relational database performance, but also activities that within the E&P world would usually be performed by other staff, such as understanding the data technology requirements.

This is another area in which DMBok seems to slightly differ from E&P best practice. For example, while DAMA states that “*database support is at the heart of data management*” this

⁵ Slide from “The Data Integration Spectrum” presented at AAPG Cairo 2002

is clearly not true for most oil companies. There is never a single relational database that holds the important data, it is spread across a collection of databases, directories and other repositories. The operations management element must coordinate all these disparate resources to enable the company's business activities.

It is common in oil companies to have an IT group that takes on the more generic parts of this support, for example providing storage and ensuring network connectivity, and to have a separate set of "Petrotechnical Data Managers" that specialise in the handling of E&P data. This approach is more likely to be able to meet the specific needs of the users. This Petrotechnical support is often delivered by a vendor that specialises in handling these categories of data.

This division of the operational elements into generic IT and Petrotechnical groups makes having a clear definition of the roles important for all concerned. Within the best E&P companies, standards such as ITIL⁶ are employed to clearly define these services, however even when less formal approaches are taken it is usual to closely monitor the activities being performed.

Data Security Management

planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information.



The DMBOK clearly lays out the process of implementing data security: understanding the requirements, defining the policies, setting the standards, describing the procedures and managing the entitlements. These concerns are exactly the same within the oil industry and for this topic the material is a valuable guide.

When data is maintained in shared facilities, for example within CDA in the UK or PetroBank in Norway, the main concern is with correctly identifying the entitlements. These issues are commonly encountered in other industries.

Reference and Master Data Management

reference Data Management is control over defined domain values... Master Data Management is control over master data values to enable consistent, shared, contextual use across systems



The divisions between reference data, meta data, business intelligence and document management that are a standard approach in most industries do not hold for E&P data. There would be no consensus between E&P disciplines as to which items fell into each category. In DMBOK reference data describes the identifiers shared across transactions, especially those involving more than one domain. Master data is a "golden record" that contains the definitive values for a particular data category.

This is the one topic where the E&P world is most different from that described within DAMA. The concept of an integrated "Master" set of E&P data has been the subject of heated discussion for at least the last 20 years. Many companies, vendors and industry associations have tried to implement a technical system able to manage the complete range of important E&P data, these efforts have failed to deliver the anticipated benefits. As a result there is a caution about using the phrase "Master Data" without qualifying its exact meaning.

⁶ The "Information Technology Infrastructure Library" (ITIL) is a widely employed set of standards for managing Information Technology published by OGC.

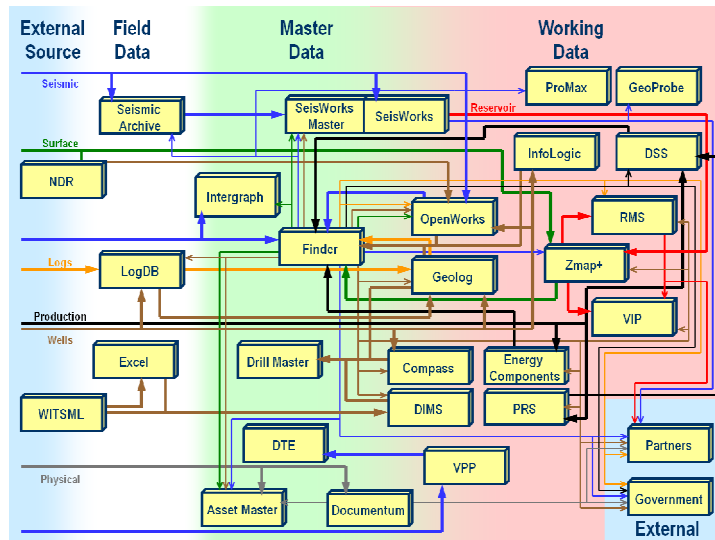


Figure 8: “Master” data is usually held in a range of locations⁷

At the moment, in practice, all E&P companies hold their “approved” data in a range of repositories each of which maintains part of the complete corporate picture. This situation is the inevitable result of integrating a wide range of distinct “islands of data”, each of which has been tailored to deliver the “best in breed” performance for a particular group.

This plethora of different databases, file hierarchies and other repositories makes the need to share reference data even more acute. It also makes the clear definition of these shared elements quite a challenge.

So the DMBok assumption that clear reference data can be easily defined and that all the stakeholders can agree which elements make up the “Master Data” certainly does not apply in E&P.

Data Warehousing and Business Intelligence Management

planning, implementation, and control processes to provide decision support data and support for knowledge workers engaged in reporting, query and analysis.



In DAMA a Data Warehouse is a combination of two primary components. The first is an integrated decision support database. The second is the related software programs used to collect, cleanse, transform, and store data from a variety of operational and external sources.

Most E&P companies that have implemented decision support databases have done so for focused elements of their business, for example maintaining descriptions of potential exploration prospects to allow various business scenarios to be tested when selecting new plays. These initiatives have to precisely define their scope, otherwise they risk being swamped by potential data sources. This means that such initiatives can be best thought of as just another discipline, rather than employing the techniques that most industries would refer to as business intelligence (BI).

E&P companies do, of course, employ BI techniques for enterprise reporting, implementing management dashboards and scorecards, as well as exploiting analytic applications, though these activities are typically not closely tied to the structured E&P data. Often Petrotechnical

⁷ Picture from “Lessons Learned from Site Assessments” presented at PDIM 2006

teams will be called upon to provide specific data elements to enable the BI activity, but this is best seen as part of the data management operations.

Document and Content Management

planning, implementation, and control activities to store, protect, and access data found within electronic files and physical records... Document and Content Management is the control over capture, storage, access, and use of data and information stored outside relational databases



The effective management of documents is an absolutely essential component of E&P data management. In some companies it is common for whole disciplines to do the majority of their analysis using Microsoft Excel and to deliver all their results using Microsoft PowerPoint.

The DMBoK description of this topic has a number of valuable insights, however it could further emphasise the importance of employing domain experts to define the taxonomies, policies and procedures that should be implemented. In our industry a few attempts to implement effective document management have resulted in widely discussed failures. These have often resulted from a generic IT group attempting to impose a system without sufficiently understanding the specific requirements of the E&P user community.

In addition most E&P companies would place the management of physical assets such as magnetic tapes and chemical samples under this same topic. This type of “Asset Indexing” is important within the E&P industry because physical assets tend to be regularly accessed, for example to reprocess a seismic survey or examine a core sample. In the majority of other industries physical items, such as records, are kept mainly for legal reasons and often not re-examined.

Meta-data Management

planning, implementation, and control activities to enable easy access to high quality, integrated meta - data.



There is a confusion about the meaning of the term “meta-data”. DMBoK defines it to be data about the data that is held, for example describing the locations where it is to be found, who is entitled to access it and the processes that have been performed upon it. In most cases E&P companies have not yet achieved enough consensus about their existing data architecture to create widely used meta-data descriptions.

Many companies are in the process of implementing meta-data repositories that document particular activities. For example, the dispatch of data to government and partners is usually related to contractual obligations, so keeping a precise record of the actions taken is crucial. However even with this level importance there are plenty of companies that track the activities using informal procedures.

This is another area where DAMA’s description contains many helpful elements, but some aspects of implementation within E&P are quite different.

Data Quality Management

planning, implementation, and control activities that apply quality management techniques to measure, assess, improve, and ensure the fitness of data for use.



All oil companies apply information to reduce risk, so the quality of information has a direct impact on the organisation's performance. This means that Data Quality Management (DQM) is one area that has, in recent times, received a high level of attention.

DAMA emphasises that a rigorous and systematic data quality programme is necessary to provide a solution within acceptable costs. This involves more than just correcting data, for example ensuring that the lifecycle for data creation, transformation, and transmission is improving data quality so that the resulting information meets the needs of all the data consumers within the organization. Within this framework DQM is seen as an on-going process to maintain quality, not a single event. These concerns all clearly apply to E&P data.

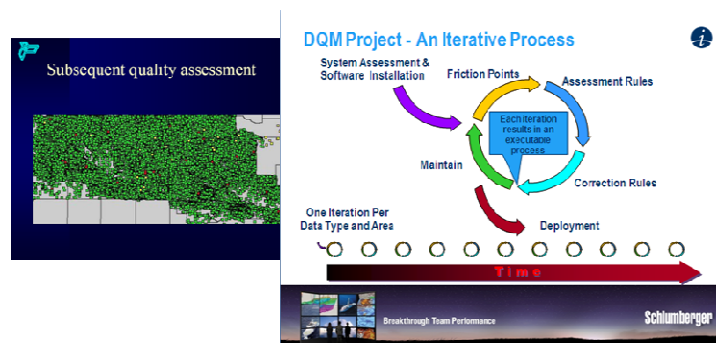


Figure 9: A number of specialised tools simplify the quality checking of E&P data

The complexity of E&P data has meant that some quality efforts employing generic techniques have failed to deliver the anticipated benefits. However it is now widely appreciated that the most effective DQM programmes utilise tools that have been tailored to deal with E&P data. If these tools can also automate the process of identifying and correcting data issues they can be far more efficient than traditional manual techniques. When these types of approaches are employed data quality projects deliver significant business benefits.

Why does E&P data handling differ from DMBok?

The majority of the DMBok provides a valuable guide to managing E&P data, however as we have seen there are some elements where E&P data handling does not fit with the DAMA description. It is valuable to review the underlying reasons for these differences.

There are few industries where information has as large a financial impact as it does in E&P. At the same time the majority of the information required for key business decisions comes from experts applying their best judgement to interpret a situation from ambiguous measurements. This combination of uncertain interpretation, the need to consolidate multiple disciplines and high financial impact means that each group of experts has had sufficient funding to demand customised tools for their own purposes.

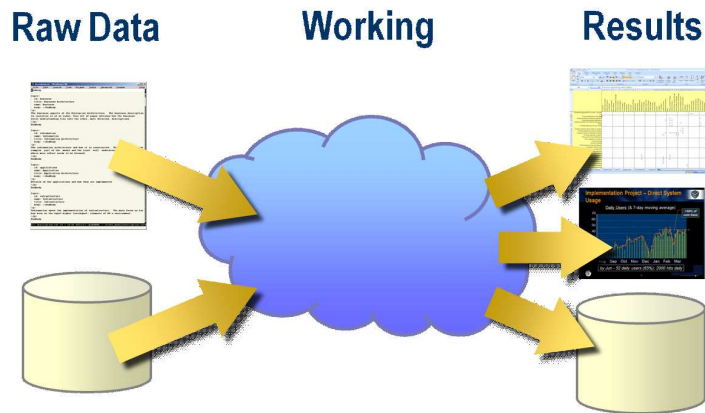


Figure 10: Most disciplines take raw data, work with it and deliver results

From the point of view of any single E&P discipline data can be seen as divided into raw input, working data and results. However one discipline's results often supply the raw data for another group. These relationships mean that the information flows in a complex web of potentially looping interactions.

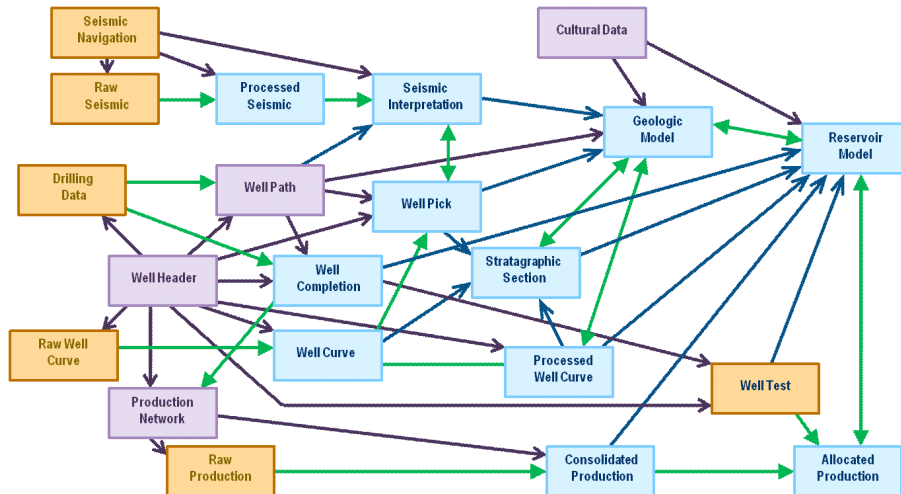


Figure 11: E&P data typically flows in complex networks

No single database

Modern oil companies use a wide range of different databases, application files and directory structures to hold their critical data.

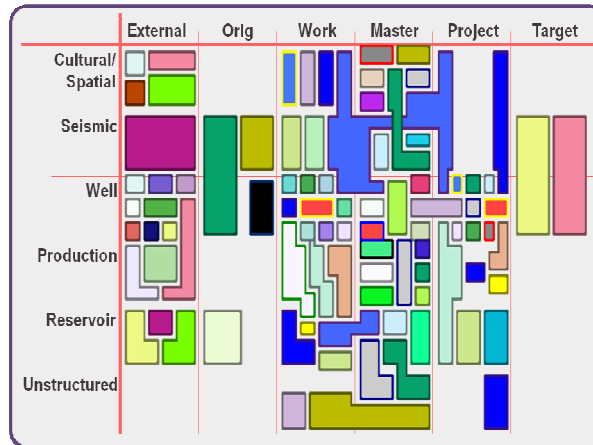


Figure 12: Even a small oil company will typically use many data stores

In part this situation reflects the history of each company, islands of specialised information have been integrated into today's complete environment. However that is not the only reason for this complexity. Each data source enables a different set of activities, and is finely tuned to make the domain it supports as efficient as possible.

It is this level of flexibility that enables all the various domains to contribute towards corporate decisions. In future it may be possible to reduce this complexity, however for the moment the most effective data management approach is to enable the users to function within the world as it is.

No single data structure



Figure 13: A data representation of an oil well is not the same thing as an oil well

Throughout the 1990s there were a number of attempts to create a comprehensive standard for all important E&P data. These efforts did not achieve their goals, not because they failed to attract sufficient funding, or through a shortage of high calibre participants, but rather because it was impossible to achieve consensus on the best "shape" of the data.

In retrospect it is obvious that different specialists describe the same thing in incompatible ways. To take an example, a Petrophysicist, Geophysicist and Reservoir Engineer all focus on

different aspects of a well, it is not surprising that the data models adopted by the tools they use have incompatible definitions for how a well is represented.

Use of innovative technologies

In most industries a large proportion of the structured data is held in some form of relational database. In E&P funding for individual disciplines has been sufficient to allow most of the specialized tools to invest heavily in improving their data storage mechanisms. As a result most of the key data repositories used in E&P are either based on non-database technologies, or wrap the database in so many additional constraints that they cannot be conventionally accessed.

This means that DAMA’s focus on relational databases, for example in “Database Operational Management”, is not appropriate when dealing with the majority of E&P data.

Separation from IT

The delivery of the standard IT elements, the networks, storage and computing platforms, has normally been done by service companies or internal groups that are not E&P specialists. It is common for E&P companies to have a separate Petrotechnical Group that is responsible for "subsurface data management" but has to rely on the "generic IT" group for the underlying services.

This division between standard IT and Petrotechnical Support has allowed the IT specialists to focus on their strengths, that is the ability to efficiently provide the infrastructure, without burdening them with the requirement to comprehend the complex world of E&P data. The Petrotechnical Support group meanwhile has been able to build an expertise within this complex domain. It is common for some or all this function to be contracted to a specialist service provider, in this case the company benefits from the experience of the service vendor in other clients.

Conclusion

The effectiveness of the information management within an Exploration and Production environment has an enormous impact on business success. There are some areas where, for good historical reasons, applying the genetic DAMA approach is not appropriate in an E&P context. However, for the majority of this complex field the DAMA Body of Knowledge provides an invaluable resource. The DMBok framework should be part of every E&P data professional’s toolkit.

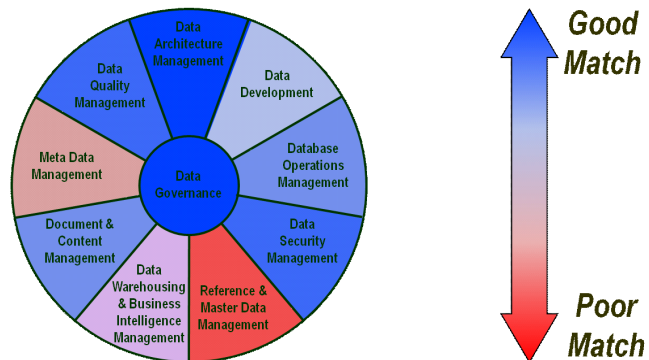


Figure 14: Match to E&P practice of the DAMA Functions