Matt Yacopetti and Stephen Mundell

Improving the quality of geoscientific information

"Garbage In - Garbage Out". Almost everyone has heard this phrase and most would agree with its core message, but in essence it's a cop out! Why? Because it is used to deflect criticism from systems that return useless or poor quality information and answers.

Today, more than any other time in history, data are business assets. All businesses that collect, store, manage, interrogate, abstract, use and deliver data, especially digital data, need to be mindful of its origins, context, relevance and usage. As a consequence businesses are increasingly turning their attention to data quality issues in order to exploit the value of their data & information assets.

There are numerous data assets in the mining industry. Those to be considered here are the original observations and measurements that represent an understanding of the subsurface (i.e. the geoscientific information). These subsurface data types are used as inputs for modelling (both economic and operational) and are interpreted to provide information upon which decisions will be made.

Therefore the value of the original observations and measurements to a mining business extends well beyond the cost of drilling or sample analyses etc. When compared to other asset classes, geoscientific data have properties that are very different from more tangible assets such as truck fleets, dragline buckets or explosive inventories for example. If we accept that data are business assets, then we should also accept that they should be managed as aggressively as any other asset class.

To illustrate this problem, think of a well known physical mining asset and answer the following questions:

1. **How much or how many assets do you have?**
2. **How much is each asset worth to replace?**
3. **How much are the assets worth collectively?**
4. **How many assets are unused?**
5. **How many assets are unserviceable or otherwise unfit for use?**

Now substitute your mine data or your latest drillhole data results for the "asset" and perform the same five-question exercise. Were you able to answer as many questions with similar or greater precision compared to the physical asset class?

As simplistic as this exercise is, it strives to make the point established by Redman (2008) that data are essentially unmanaged assets and that it is easier for businesses to leverage physical capital and human resources than it is to leverage data and information.

It is a simple thing to say that data quality is important, but how can this statement be assessed? If you do have poor quality data, what is its cost to your business and where should you focus your efforts in order to improve the situation? Typically mining businesses have drilling data, assay data, grade control data, and production data etc., all of which are essential for managing the asset. How difficult therefore, would it be to manage that asset if the data was missing, incomplete, inaccessible or hard to understand?

For a business to successfully leverage the full value of its data and information they must be:

- **Accurate (correct)**
- **Accessible (able to be located quickly and easily)**
- **Trustworthy (consistent within context)**
- **Able to be understood (for those it is intended for)**
- **Useable (by appropriate users)**
- **Secure (protected against loss, theft, fraud and degradation)**.

Too often individuals, teams or organisations spend valuable time checking, researching, locating, reconciling and reformatting their own information. In other words they are acting as "data janitors" rather than value adding "knowledge workers".

The reasons for this are complex and involve both the properties of the data itself and the way that humans interact with those properties, including organisational and political. We will explore some of these issues in order to suggest ways in which the data quality of geoscientific information can be improved.

THE UNIQUE PROPERTIES OF DATA AND INFORMATION

Digital data and information can be shared, almost without limitation. As a consequence they are more difficult to secure and protect. Data have properties that can make it behave like a living organism rather than static information (e.g. long lifetimes and the ability to replicate and change). These properties can also mean that data and information can be more difficult to locate and share. Ever since the advent of digital information and the internet, most of us can attest that more data does not automatically equate to better data.

All systems and activities that use data create more of it. Data volumes in their own right are becoming a difficult problem. The amount of time needed to double the data generated and maintained by a business is estimated to be between 12 and 18 months, Whiting (2006). This is due to the fact that data is not consumed with use and can be re-used time and time again. Logically, the existence of poor quality data may have potentially serious consequences, Lacey (2010). The urgency of dealing with this issue is growing, as is the likelihood that poor quality data will be shared with users, both internally and externally.

Who then, in the organisation is responsible for dealing with these issues? Many mining professionals would nominate the IT Department or the Database Administrator. This is despite the fact that these roles are custodial, that is, they manage the data, but they are unlikely to be accountable for data quality at the point of origin. Data management is not purely a technological issue. It is as much an issue of organisational structure and leadership.

PEOPLE PROCESS AND TECHNOLOGY

When an asset is recognised as valuable to a business, the business will invest in its maintenance and management. Try to imagine a viable mining business today that does not invest in its human resources or capital equipment? Data management is no different. Mining businesses should invest in their data management capabilities in order to leverage and secure their valuable geoscientific information assets. A multifaceted approach that focuses on the interchange between people, process and technology will be fruitful.

People

By far the most important is the human dimension. Mining businesses need to recognise that they own the issue of geoscientific information management and to employ the people responsible. Those people need to have the required Information Management skills to do the job. No technology will improve the situation until people in the organisation have clear data quality goals and the processes in place to support those goals.

Corporate and business trends have also played their part. The cyclical nature of exploration and mining activity that is driven by commodity prices has seen periodic downturns characterised by downsizing, mergers and a move to use external service providers. Broome and Cox (2007) have noted that the increased mobility of exploration expertise has resulted in diminished corporate knowledge of data assets.

The ‘people dimension’ can be difficult to change. Despite its importance, the topic of ‘data management’ suffers from a lack of profile. It lacks 'appeal' and frequently fails to capture the attention of management, until there is a problem. Even then, the problem is likely to be associated 'with IT' rather than the lack of an overall data management capability.

Process

Advances in information technology and the need for enterprise-wide computing environments have led to the adoption of more structured data management. For the mining industry the key drivers have been:

- The move away from systems that allow users to manage data as if they own it
- The need to prevent errors at their origin
- The realisation that many errors do not affect the person or group who created them
- The move to systems that support corporate standards and data interoperability.

Processes connect people or groups to tasks. For mining businesses a good process will enable internal and external groups to establish effective data supplier relationships. With geoscientific information in mind, the processes listed below all benefit from an easy connection between people and the task:

- Capture of original observations and measurements to reduce error at the point of origin
- Discovery and access to the captured data
- Differentiation of original data from derivative data
- Identification of metadata (or data about the data)
- Data interoperability — providing the data in variety of standard formats
- Delivery of select data sets to required client systems
- Aggregation and abstraction of data and delivery in desired formats
- Internal data quality alerts and measures.

Redman (2008) estimates that knowledge workers spend on average 30% of their time searching for the data they need, and are unsuccessful at least half of the time. In a mining company where skilled workers are at a premium, the business cannot afford for geoscientists to spend this amount of their time using or creating poorly designed data management systems. This is a major opportunity to improve workplace productivity.

Technology

The challenge for the technology dimension is to connect people to the process. This will encourage the discovery, access and integration of diverse geoscientific information. A well designed Geoscientific Information Management System (or GIMS) will provide an architecture that will enable the geoscientist to capture and access high quality information on demand. To achieve this goal, the business must pay at least as much attention to their work practices and organisational structure as they do to the selection of available solutions.

For mining companies wanting a comprehensive technical solution that supports their individual business requirements, there are a variety of professionally developed, commercial database solutions. These systems may offer all or some of the following enterprise architecture advantages:

- Professionally designed, developed and maintained
- Database hosted by a server based Enterprise Level RDBMS (e.g. Microsoft SQL Server)
- User friendly graphical user interfaces
- Open and persistent data models
- Support for geoscientific data types and interoperability with client systems
- Built-in technology redundancy

MISUSE OF THE HUMBLE SPREADSHEET

One of the most pervasive data management tools available today is the spreadsheet. There is ample evidence from detailed studies to show that error rates in spreadsheets are at levels that would be unacceptable to any organisation, Panko (2005). Despite this, spreadsheets continue to be used in the mining industry for routine data collection and management tasks.

Simply put, spreadsheets are error prone. Very few are created by qualified developers and most organisations lack policies governing their use or development. Many use macros and have links to other spreadsheets or databases and many don't use cell protection. It is common for spreadsheet development to be informal and iterative with extensive revision. Many spreadsheets do not employ modular design principles, and testing with specially designed test data is limited. Many spreadsheets are rushed in development and fewer still are documented.

Why then are they so commonly used? Overconfidence in the accuracy of spreadsheets by the people who create them seems to be one reason why their development appears to be so casual and ubiquitous. Although large levels of overconfidence would seem to be unreasonable, Panko (2005) argues that it is perhaps the most well-established phenomenon in behavioural studies.

Cost and convenience are factors. Selection due to convenience would tend to support informal development controls, whereas cost is a perception problem in as much as businesses may not be convinced of the cost impacts of poor data quality. This is supported by research performed by Eckerson and Sherman (2008) into the main reasons why spreadsheets exist in a range of industries surveyed in North America and Europe. In their ranking, the top five categories were:

1. High degree of local autonomy / control
2. Quick fix to integrating data
3. Inability of the business improvement team to move quickly
4. Low cost option
5. This is the ways it is always done here

Spreadsheets are used due to a perceived lack of viable alternatives.

THE VALUE OF DATA IS MISSING

If mining executives were convinced of the value their own data, then commercial GIMS would be more pervasive than spreadsheets. The fact that this is not the case indicates that many businesses remain unconvinced of the value of their own data or of the costs due to its poor quality. Moseley (2010) reports that in a recent survey by IBM, only 22% of the executive respondents said "data are critical business assets with known value". 27% said "their data are somewhat valuable assets", 16% said "data are an asset with some intrinsic value", 22% said "data are an application resource that always seems broken" and 13% "said data is not even on their radar".

Such analyses indicate that many organisations are content to live in a 'spreadsheet hell' with their inherent disadvantages and are reluctant to adopt a more architectural, enterprise-wide approach. This would seem to indicate therefore that there is a need for more persuasive evidence of direct measures that demonstrate the business value of well managed geoscientific information.

More documented business cases are needed in order to make the connection between people, process and technologies. There is certainly no lack of suitable "war stories" concerning the mismanagement of data in our industry, but very few organisations are prepared air their laundry in public. This then is a challenge for the industry and if embraced; would reflect a growing maturity and understanding of the issues.

THE COST OF DATA QUALITY

The news media is littered with examples of poor quality data causing serious problems. These examples range in scope from lost spacecraft, financial reporting scandals, credit rating errors, military intelligence failures, the unintended release of private information to the public, and the list goes on and on.

These examples demonstrate that data quality is important to all of us and the risks of not getting it right can be higher than we think. Ask yourself the following question. Do poor data management practices occur in my organisation? The logical answer is that it probably does.

Poor quality data can adversely affect businesses on three levels, Operational, Organisational and Strategic, Redman (2008). On an operational level, impacts will be due to the higher cost of production and lower productivity. In a competitive business environment anything that impacts your productivity may erode any competitive advantage you enjoy. On the organisational level, poor data quality can lead to delayed or bad decisions with an attendant increase in risk. It may also affect the trust that other groups place in the data if they need to share it. On the strategic level the affect of poor quality data may be the inability to align people, process and technology within the organisation and make it harder to execute the corporate strategy owing to distracted management attention.

All of these issues have real dollar consequences in both the short, intermediate and longer terms. In addition, there are likely to be significant opportunity costs associated due to inaction or due to the results of poor decisions.

CONCLUSIONS

The original observations and measurements form the basis of all economic and operational decisions made on the future of a mining resource. These data are a business assets and it should be accepted that they be managed like all other significant business asset classes.

For a business to leverage the full value of its data assets, the information must be accurate, accessible, trustworthy, able to be understood, usable and secure. This can be achieved by investing in the people, process and technology components of your business.

People: By far the most important component — technology will never make up for shortcomings in terms or personnel. Data management lacks appeal — raise its profile by understanding and demonstrating its value and importance to the business and investing in the people who collect and manage your data.

Process: Implement rigorous business processes that outline the purpose for the process, have a clear accountability and ownership model, define a standard process and how success of the process can be determined. By having documented and accepted processes, you will minimise the time geoscientists spend using or creating poorly designed data management systems.

Technology: The technology component allows the implementation of the process and makes it accessible and usable by the people — technology does not come first. The technology must be appropriate for the task, support the needs of those who operate it and ensure the accuracy, accessibility, usability and security of the data assets.

These components go hand in hand to ensure the quality of your data. The value of your data to your business goes beyond the cost of drilling, sampling and assaying — think of the opportunity costs of decisions based on poor quality data.

ACKNOWLEDGEMENTS

The authors would like to thank acQuire for the permission to publish this paper as well as Stephen Alpers and Warren Cook for their helpful comments during the review of the draft.

REFERENCES

BROOME, J. & COX, S., 2007: Geoscience Information Management and Access: Evolution of a Key Enabler for Exploration Success. *In:* Milkereit, 2007: *Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration*, 97–108.

EKERSON, W. & SHERMAN, R., 2008: Strategies for Managing Spreadmarts, Migrating to a Managed BI Environment, First Quarter 2008 Best Practices Report, TDWI Research. http://tdwi.org/

LACEY, D., 2010: *Improving Data Quality*. Computer Weekly, March 29. http://www.computerweekly.com/blogs/david_lacey/2010/03/imp_roving_data_quality.html

MOSELEY, M., 2010: The Value of Data, Mastering Data Management, Business Process Management blog, January 21 http://blog.initiate.com/index.php/2010/01/21/the-value-of-data/

PANKO, R., 2005: What We Know About Spreadsheet Errors, University of Hawaii, College of Business Administration, Honolulu, Spreadsheet Research (SSR) Website: http://panko.shidler.hawaii.edu/SSR/index.htm

REDMAN, T. C., 2008: *Data Driven*. Harvard Business Press, Boston.

WHITING, R., 2006: Hamstrung by Defective Data, Information Week, May 8. http://www.informationweek.com/news/global-cio/showArticle.jhtml?articleID=187200771