

C A L E N D A R

- **SME Arizona Conference**
December 4–5, 2005
Double Tree Hotel at Reid Park
Tucson, Arizona
e-mail: meetings@smenet.org

- **NWMA Annual Convention:
Exploring the Modern Minerals
Renaissance**
December 5–9, 2005
Red Lion Hotel at the Park
Spokane, Washington
e-mail: pheyywood@nwma.org

- **Mineral Exploration Roundup
2006**
January 23–26, 2006
Westin Bayshore Resort & Marina
Vancouver, B.C., Canada
e-mail: lelliott@chamberofmines.bc.ca

- **National Western Mining
Conference and Exhibition**
February 7–9, 2006
Grand Hyatt
Denver, Colorado
e-mail: colomine@coloradomining.org

- **PDAC 2006**
March 5–8, 2006
Metro Toronto Convention Centre,
North Building
Toronto, Canada
e-mail: lmcDonald@pdac.ca
Visit us at Booth 312

- **2006 SME Annual Meeting &
Exhibit**
March 27–29, 2006
America's Center
St. Louis, Missouri
e-mail: sme@smenet.org
Visit us at Booth 413

Data Management in the Minerals Industry (Part 1)

This month's article for Pincock Perspectives is the first in a two part series concerning data management in the mining industry and its implications with respect to mining operations and compliance with U.S. Sarbanes-Oxley legislation and similar international initiatives.

Overview of Data Management in the Mining Industry

The management of mining and mineral processing operations involves the collection, organization, analysis and reporting of data from multiple sources. These data sources can be as simple as truck counts in a small quarry operation to as complex as real-time process data used to monitor and control a billion dollar mineral processing facility. Their importance to their respective operations lies in how the data and subsequent information created from that data is used to manage their respective businesses.

As operations become more complex the importance of managing an ever increasing amount of data becomes more important and more significant of a task due to the cost implications of the information created from that data. In mining operations, even the ones that are relatively small, good data management is a requirement; first, to optimize the results of their operations, and second, in response to changing legislative requirements such as U.S. Sarbanes-Oxley legislation that requires company officers to be accountable for the integrity of their financial statements which are ultimately derived from operating data.

Types of Data

The sources and types of data produced in a typical mining operation come in three forms: 1) Structured data, 2) Unstructured Data, and 3) Metadata. All forms are found in most operations and are used to varying degrees in the production of information and the

management of the operation. Some of these sources of data can be either digital or exist only as paper documents.

Structured data is organized in a known or easily decipherable form such as a table of cost accounts in a typical accounting database. This type of data is the most easily utilized. Examples of this type of data are found in most database systems.

Unstructured data which represents the majority of data in any operation or organization is typically found as notes, e-mails, personal knowledge, etc. This type of data is not as easily utilized, but represents the next frontier of information management and knowledge sharing. Desktop file indexing and searches such as Google Desktop, X-1, and others have begun to tap into the wealth of information that is locked up in these sources. Most organizations have barely scratched the surface in utilizing this type of data.

Metadata is the most important type of data and is often the most neglected, but can have a significant impact on data quality, understanding, and general usefulness, because it enables data to be more thoroughly understood and utilized. Metadata is data that describes the properties or characteristics of other data and puts the data into a context that makes it understandable. Examples of metadata are cost accounts used in an accounting system that relate an account number to a cost area such as labor or geologic codes in a database that relate geological intercepts to know rock types. In themselves, a table of cost codes is meaningless, but with the metadata which relates them to a specific cost area in an operation the codes become useful in analyzing cost numbers.

It is the idea of having data that describes other data that is behind the transition to using XML as the preferred file format for transferring data between

disparate systems especially over the internet. An XML file is a format that allows data to be seamlessly transferred between systems because it not only contains the data, but also contains information about the purpose and format of that information. Most of the latest versions of Microsoft Office suite of applications will read and write in extensible markup language or XML format. An example XML file for a two row data table is shown below:

```
<?xml version=" 1.0" encoding=" UTF-8" ?>
<dataroot xmlns:od=" urn:schemas-microsoft-com:officedata"
xmlns:xsi=" http://www.w3.org/2001/
XMLSchema-instance"
xsi:noNamespaceSchemaLocation=" DRILL.xsd"
generated=" 2005-11-30T07:31:17" >
<DRILL>
<DRILL_ID>DRL-001</DRILL_ID>
<DRILLTYPE>DTH</DRILLTYPE>
<DRILLSIZE>NQ</DRILLSIZE>
<DRILLHRS>2570</DRILLHRS>
<DRILLOPER>J. Thompson</DRILLOPER>
</DRILL>
<DRILL>
<DRILL_ID>DRL-003</DRILL_ID>
<DRILLTYPE>ROTARY</DRILLTYPE>
<DRILLSIZE>HQ</DRILLSIZE>
<DRILLHRS>8790</DRILLHRS>
<DRILLOPER>K. Roberts</DRILLOPER>
</DRILL>
</dataroot>
```

Data vs. Information

When most people think of the term data they think of collections of numbers or text, but data is significantly more than that, it is a collection of facts, text, graphics, images, sound, or video that have some meaning in the user's environment (Hoffer, et al, 2002). Data in itself has little meaning and is generally useless until it is turned into information that can be used to manage a process in the business environment in which it exists.

A drillhole database without metadata that relates the individual data items, such as drillhole identification and assay results to qualitative data, such as a drillhole location map and assay items is just data. The metadata puts the drillhole data into context and allows a user to use that drillhole data in the evaluation of a mineral resource and makes it useable information applicable in the context of a resource evaluation.

The management of information in the mineral industry has itself become a science and is at the core of many software and consulting businesses.

Sources of Data and Information

Mining and processing operations and companies collect many types of data in many different systems in various parts of their operations. Some of these sources of data and systems are listed below:

- ◆ Resource Evaluation
- ◆ Geographical Information (GIS) and Remote Sensing
- ◆ Survey and Global Positioning
- ◆ Laboratory Information Management
- ◆ Mine Planning and Scheduling
- ◆ Production Management
- ◆ Process Control
- ◆ Maintenance Planning and Management
- ◆ Accounting and Materials Management
- ◆ Environmental Management
- ◆ Payroll and Human Resource Management
- ◆ Unstructured Data (e-mail, files, Web sites, internal reports, etc.)

This is not a comprehensive list, but represents some of the more common systems, sources of data, and information that may be found in an organization. The extent of this list is an indication in itself of how large a part of mining operations the management of information has become.

Types of Information Management Systems

There are primarily two types of information management systems in use in the minerals industry, as well as in most other industries; flat file systems and relational database management systems (RDBMS). Nearly all software related data management and analysis occur in one of these two types of systems and are used to store, manipulate and turn data into information useable to the organization.

A third type which will also be discussed is office productivity software such as Microsoft Excel, which can technically be considered a flat file system. Because of its nearly ubiquitous usage in the business world, it is commonly used as a simple data management and information system in itself.

In the early days of computing most data storage and retrieval systems were based on proprietary flat file structures unique to a specific piece of software. These systems typically stored different sets of information in structured files in either a binary or ASCII format. As these systems became more complex the limitations of these types of systems became apparent due to the increasing complexity of data storage, data processing, and analysis requirements. Surprisingly, some of these legacy systems persist today even though most enterprise and even desktop software systems have migrated to relational database type systems as the primary storage mechanism within data intensive software applications.

Modern extraction transformation and loading (ETL) tools have made the migration, of information out of legacy systems more easily accomplished and interfacing these legacy systems with a modern RDBMS is not an onerous or expensive task. An ETL tool is used for mapping and transferring information from one system to another in a controlled and repeatable manner.

The limitations of these proprietary flat file systems (Hoffer, et al, 2002) include the following:

Program-Data Dependence – The data stored or contained in a program was not easily accessed by other programs that could potentially use this data.

Duplication of Data – Because data in these systems is not relational data duplication is significant because it often had to exist in more than one location. Duplication of data often results in data integrity issues due to an update in one file and not another.

Limited Data Sharing – Since each application used a unique set of files there was little opportunity to share information across departmental lines that used the same information.

Excessive Program Maintenance and Development Times – Changes in software requirements often necessitated changes in the data structure and therefore applications had to be modified and re-compiled to implement these changes. This resulted in a significant amount of time being spent in updating and maintaining of these types of systems.

The answer to the limitations of flat file processing systems was the RDBMS which is the most popular data storage model in use today and was originally conceived by Dr. E.F. Codd in his 1970 paper titled "A Relational Model of Data for Large Shared Data Banks." (Stephens, et al, 2002) This model was later made main stream by companies such IBM and Oracle and is at the core of most relational database systems used in business today.

The primary advantages of the relational database model over the flat file processing system are numerous (Hoffer, et al, 2002) and include the following:

Program Data Independence – The database engine or relational database management system RDBMS is independent of the program utilizing the data and therefore a single data structure can be utilized by many different programs without affecting the underlying data model or data integrity.

Minimal Data Redundancy – Because data is grouped in logical sets or tables that are related to other tables by a unique primary key field in the table, redundancy in the data is significantly reduced, and in its purest form is completely eliminated. This is a significant issue in extremely large databases, because minimal data redundancy is a necessity for storage efficiency and query performance.

Improved Data Consistency – The existence of data in only one location and the reduction of redundancy in the data lead to more consistent information. An employee's address in the payroll system will always match the employee's address in the accounts payable system if they are using a unified relational database model.

Improved Data Sharing – A RDBMS in a Client/Server environment facilitates sharing of data among many people while still controlling access to the data to only those authorized to see, use, or edit/update/delete the data.

Increased Productivity of Application Development

– Application development time is reduced because the data model is standardized.

Enforcement of Standards – Data structure standards are enforced through input masks that control how data is stored. An example is a typical U.S. Postal Code which could be stored in

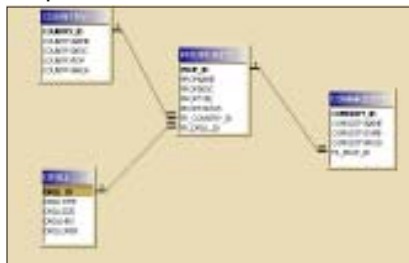
one of several formats or data types. The Zip code 80228 can either be stored as a number or as a text or character string of five characters and may or may not include the Zip plus four code 80228-2950. Enforcement of Standards on how data is stored reduces problems with data integrity later on when that data is queried or utilized in calculations.

Improved Data Quality – Enforcement of standards leads to improved data quality and reliability when data is retrieved and used in reporting information from a database.

Improved Data Accessibility and Responsiveness RDBMS are easily accessed and information made available to multiple users across many areas of an organization. Even novice database users using SQL, can access data contained in RDBMS with very little training. SQL is the acronym for Structured Query Language the common programming language used by most RDBMS to query data.

Figure 1 shows a simple relational database built in Microsoft Access, to manage exploration information including countries, mineral commodities, properties, and drills.

Figure 1
Sample MS Access Relational Database



Although the Relational Database Model and corresponding systems have many advantages they do also have some disadvantages over the traditional flat file system including the following:

Specialized Personnel - are typically required to develop, install, and maintain a RDBMS; although, special skills are typically required in the utilization and maintenance of most advanced software systems.

Installation, Conversion, and Management Cost and Complexity – RDBMS are typically more complex than a stand alone flat file system and therefore are more costly upfront to implement.

Need for Explicit Backup and Recovery – Because an organization typically consolidates its data into a single RDBMS it essentially puts all its "eggs in one basket", This consolidation typically requires structured backup and recovery tools and procedures which are often integrated right into most RDBMS such as Microsoft's SQL Server. It is also why these systems are the most methodically tested and stable software systems in use today. Oracle Corporation even advertises its latest database system as "unbreakable."

Organizational Conflict – The consolidation and standardization of data across an organization often creates organizational conflict because specific groups have specific preferences to how data should be stored, organized, and reported. These conflicts can often be minimized by an implementation team that has an intimate knowledge of the how different groups currently use data or wish to use data in the future and how this fits with the organizations objectives.

Another commonly used data management system used in most organizations today is the spreadsheet; most often Microsoft (MS) Excel. The spreadsheet is a powerful analytical, reporting, and presentation tool and is often used as an ad hoc database.

The spreadsheet is strong in the following areas:

Usability and Familiarity – For most business users the spreadsheet is part of their business day at some point because of its familiarity and ease of use.

Flexibility – Part of the spreadsheets popularity as a data management tool has come from its flexibility to accomplish more than one task in a single environment. It is a calculator, a word processor, and an analytic tool all in a single environment.

Developer Skill Level – The spreadsheet can be quickly used to accomplish basic calculations and data management tasks by users with very little experience and knowledge of only a few program features.

Calculations and Presentation – Even the most complex scientific calculations can be achieved in the spreadsheet and used to create excellent presentations using built in graphing tools.

Areas where MS Excel is weak:

Security – Although MS Excel has some security features that enable it to be password protected and to protect various sheets and ranges in a workbook this security is fairly rudimentary and is actually rarely used by most users.

Maintainability – Maintaining a complex Excel spreadsheet can be a difficult task at best if multiple files have been linked and there are multiple tabs of data and calculations that are interdependent. A small change in one place in the worksheet can cascade through the whole series of calculations and result in errors that are difficult to resolve.

Capacity – MS Excel has a current limit of 65,536 rows of data which is adequate for most users, but is relatively meager in terms of data storage capacity. The next release of MS Excel will expand this limitation. Relational databases such as MS SQL Server can manage terabytes of data without issues. The only real limits being the computing hardware on which it is running.

Multiple Users – Although multiple users can share an MS Excel workbook, MS Excel is not a true multi-user environment where a group of people can be actively working in the same workbook, simultaneously.

Networkability and Connectivity – MS Excel can certainly be accessed across corporate

networks but it does not have built in authentication capabilities such as MS SQL Server which can be integrated with the Windows environment to allow for simplified user access and connectivity.

Data Validation and Data Corruption – Two of the most significant flaws with MS Excel and other spreadsheet applications and the primary reason it is being replaced in the corporate environment by database applications is in the area of data validation and data corruption. MS Excel has no inherent data validation capabilities which lend to its flexibility and ease of use but also require the user to explicitly validate data and calculations to ensure they are correct. Databases validate data on input into the field, if the limits of the data have been pre-defined, before it becomes part of the data set and therefore most input errors are caught.

The second most significant flaw with MS Excel is data corruption. Many mining operations and business users in general rely on MS Excel for day to day operations reporting capabilities, again because of MS Excel's ease of use. Unfortunately, because most of these operating spreadsheets are unprotected, a single inadvertent entry or corruption of a formula in a single cell can result in reporting erroneous numbers. Changes in MS Excel are difficult to effectively track and audit.

It is for this reason alone that many public companies are being forced to move key data

used in financial reporting out of spreadsheets in order to comply with Sarbanes-Oxley.

Sarbanes-Oxley Compliance

The Sarbanes-Oxley Act of 2002 was legislation passed in the U.S. to address reporting scandals and the collapse of several major U.S. corporations in 2001. The passage of this legislation represented a significant change in management's responsibilities in reporting their financial information.

The focus of many companies' efforts since the passage of this legislation has been how to comply with this legislation, especially the requirements contained in Section 404. The control over data and the documentation of internal controls over this data has been a key piece of the compliance effort and has required many companies to adapt their information systems and internal procedures to comply with the requirements of this legislation.

Next month's article will continue the discussion of data and information management in the mining industry and will further explore the emerging area of business processes improvement and business intelligence.

References for the above article provided upon request

Glossary of Terms

ASCII – American Standard Code for Information Interchange is a code for representing English characters as numbers, with each letter assigned a number from 1 to 127. Text files stored in ASCII format are often called ASCII Files (ASCII, 2005)

Client/Server – A network architecture in which each computer or process is either a pc/workstation or server on the network.

ETL - Short for Extract, Transform, Load, three database functions that are combined into one tool to pull data out of one database and place it into another database.

Metadata - Data that describes the properties or characteristics of other data.

Relational Database (RDBMS) – A Relational Database Management System is a system that stores data in the form of related tables.

SQL – Abbreviation for Structured Query Language. A standardized query language that is used to request information from a database.

Terabyte – A measure of data storage equivalent to 2 to the 40th power (1,099,511,627,776) bytes. This is approximately 1 trillion bytes.

XML – Short for Extensible Markup Language, a specification developed by the W3C. XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.



Pincock, Allen & Holt is a consulting and engineering firm serving the international mineral resource industry. Your comments and suggestions are always welcome. Contact Pincock, Allen & Holt • 165 S. Union Blvd., Suite 950, Lakewood, Colorado 80228 • TEL 303.986.6950 • FAX 303.987.8907 • www.pincock.com. Pincock Perspectives is published as a free information service for friends and clients. Information for News Pix is paraphrased from various sources; references available upon request.