



GUIDE TO OPEN DATA:

Overcoming Data Challenges in the Mining Industry

SourceOne[®] 
Enterprise Knowledge Performance System

 Eclipse Mining Technologies[®]

Table of Contents

Current State of Data in Mining 4

 The Current Landscape 4

 Exploring Data Formats 5

Do Database Servers Offer a Solution?..... 6

 Data Warehouses 6

 Data Lakes 7

 Data Lakehouses 7

A Fitting Solution..... 8

 SourceOne® Basics 9

 See the SourceOne System in Action 10

Current State of Data in Mining

Current mining solutions offer varying scales of data openness and accessibility.

They maintain data in a mix of locations, formats, attribution types, licensing restrictions, data definition and access methods. The aggregate of which is a system which on the whole, is found wanting in every metric of **open data**.

Regardless of whether the data is planning, operational, geotechnical, plant or other, the process of moving data between systems, or trying to generate intelligence spanning multiple data stores is difficult, if not impossible, even among solutions from the same provider. This is not true open data.

New breakthroughs in data analysis are appearing on the market at a constant pace. Artificial Intelligence and Machine Learning are making their way into the mining sphere. To take advantage of these tools in the commercial space requires truly open data in a unifying enterprise system.

The Current Landscape

Current mining solutions generally use not one, but many or all the following locations, formats and attribution methodologies:

Location of data storage is a mix of local directories (usually multiple copies on the same machine in different states), local database instances, shared server drives and directories, shared database instances, SharePoint, ERPs...etc.

Attribution is accomplished with filenames, file paths, internally to the file format or a set of database tables. There may be some basic tagging system in place, but it is usually up to the user to maintain.

Licensing restrictions range from text files and databases (which are generally unlicensed) to proprietary binary files accessible through an API or 'data dump' executable (generally licensed).

Data definition is interesting in that even if you have a text file with the data you want, it does not mean you have a description of what the data means. A good example is the use of an integer value to correspond to a specific characteristic (e.g. the number 1 = horizontally rotated model, you can read the 1, but do you know what that 1 means?).

Access methods vary widely. Most solutions touting their data as 'open' still require a specific tool for access. Data text files are

The process of moving data between systems, or trying to generate intelligence spanning multiple data stores is difficult, if not impossible, even among solutions from the same provider.

open by their nature (i.e. I can open a text file up in Notepad) but proprietary binary files often need an API, dumping utility (which results in a text version) or you need to open the tool (which can be licensed), find the data you want, then export it into the desired format.

Formats are a mix of proprietary binary files, a variety of text (txt, ini, csv, xml, json, etc...), databases, Excel spreadsheets, etc. Varying formats make it difficult to integrate data from multiple sources.

Exploring Data Formats

Varying data formats is a big part of the problem, and large solution providers are not only aware of these data challenges, they bottleneck their own use of data. This explains why so many systems are not well integrated even within themselves, much less with other solutions or vendors.

Proprietary binary files offer many benefits including access speed, file size, and a familiar metaphor for organizing data (files and folders). Many of the established solutions have a significant amount of data stored in this manner. However, it is difficult for other solutions to read this data, and rarely, if ever, does the user have an ability to write back.

Common proprietary files such as AutoCAD .dwg or ArcGIS shapefiles allow for native access to data by a multitude of other solutions and is a huge step forward in terms

of openness. However, it retains almost all the other negatives of the proprietary binary formats and adds a couple of its own.

Text or ASCII files are common for configuration data as well as some types of generalized metadata. Most users interact with this sort of data in large scale ASCII data dumps which are usually used for the interchange of data into different systems. As with the earlier two solutions, these are files, and have the same challenges all files have.

Database variants are commonly offered as an off the shelf database solution from many vendors. If implemented correctly, database variants provide multiple benefits. Depending upon how the database schema was made and how the data is exposed, there is an inherent ability to query the data. Most data providers discourage clients from trying to access the data directly and instead provide them with an API for access, negating interoperability capabilities—a primary goal.

The bigger solutions use more than one of the preceding methods, making contextual data access incredibly difficult.

Do Database Servers Offer a Solution?

Most database server solutions serve neither the vendor nor the client well.

If a technology vendor attempts to normalize everything to the point where the native SQL query functions are useful, the performance suffers. If they serialize or blob the data extensively then performance increases, but it is unusable to an external query. So generally, a path of programmatic extensions and views, or an API is given, combining the closed nature of proprietary files with the poor performance of many database solutions into one, behind a wall of compatibility.

However, as industries strive for efficiency by gaining access to and using their data, data server solutions have become quite popular in the last few decades, but more significantly in the last 8 to 10 years. These include data warehouses, data lakes, and more recently, data lakehouses.

Data Warehouses

Although their ultimate purpose (centralize, consolidate and make use of the data) is overall similar, there are significant differences among these three. A data warehouse requires structured data and is designed and built for a specific purpose. The disadvantage of course, is that it does not store raw data and its use is constricted by the purpose for which it was originally built.

Data Lakes

On the other hand, a data lake is capable of storing massive amounts of raw data and each data point is tagged with a unique set of metadata. However, given the amount and state of the data, these usually require a team of data experts to extract analytics and maintain the data, as they can easily decay into a data swamp.

Data Lakehouses

Lastly, a data lakehouse combines elements of the two previous ones: it can store structured, semi-structured, and unstructured data, with the usage flexibility of a data lake, as they allow you to make data warehouses out of your data lake as views, without having to copy any data. Although a data lakehouse can keep up with your evolving business needs, it still centralizes the data around a team of data experts.

But even a powerful data lakehouse has its limitations when it comes to being implemented in an industry such as the mining industry. With its complex data types such as drillhole data, topography data, block model data, and FMS data, just to name a few, this industry has proven to be a challenge for existing data management solution providers.

Then, if it's not any of the above, what is the right data solution for the mining industry?

A Fitting Solution...

Organizations in virtually every industry sector are struggling with how to manage and leverage the enormous volumes of data they are gathering from a multitude of sources. The mining industry is certainly no exception.

For companies in this sector, the explosion of information presents not just challenges, but tremendous opportunities as well. The question is how can companies gain value from their data in ways that can help them meet their unique goals?

A ground-breaking solution such as the SourceOne® Enterprise Knowledge Performance System (EKPS) from Eclipse Mining Technologies provides the key to realize the value. SourceOne was designed to be able to accept and handle complex data types and processes, such as those found in the mining industry.

The enterprise system offers four major capabilities: Data Lifecycle Management, Data Integration, Digital Workflow Automation, and Data Analytics.

The following section describes the main features and capabilities of the SourceOne system, providing CIOs, CTOs, mine managers and other senior-level decision makers in the mining sector with comprehensive information about how their companies can benefit from the System in a number of ways.



Keywords: Data Lake, Data Warehouse, Business Process Management, Data Integration, Data Lifecycle Management, Data Analytics, Digital Workflow Automation, Mining

SourceOne® Basics

SourceOne is unique in that it fuses the benefits of different modern technologies into one powerful solution, a centralized server designed to handle the data, workflows and analytics typically required by a mining company. In other words, an Enterprise Knowledge Performance System.

Having been designed for the mining industry, it can ingest and handle massive amounts of raw data, in a similar fashion to a data lake.

Resembling a data warehouse, it can also store and organize data. Unlike typical data warehouses, the SourceOne system works with any type of data found in the mining industry, and as a result it is not constricted to the purpose that it was originally built for, nor does it need to be redesigned for new uses.

Influenced by Blockchain immutability properties, the System stores and creates permanent, immutable and retrievable data points. The data is distributed to users, allowing any number of users to interact with the data at the same time and then synchronize their work.

Through the use of an integrative open data format and powerful APIs, the SourceOne system retrieves, harmonizes and integrates disparate, unconnected data, facilitating the flow of data across the value chain.

Much like Business Process Management (BPM) systems, the SourceOne Enterprise Knowledge Performance System can digitalize, automate workflows and create tasks. It conveniently incorporates communication tools, so the data and its related observations are accessible from one single place.

Supported by a data abstraction layer, SourceOne data analytic tools allow users to compare disparate sources of data in the same charts or reports, enabling the creation of predictive models.

In addition, the System can also perform complex, mining-specific activities such as digital mine plan creation, modeling, plan compliance, material tracking and others.

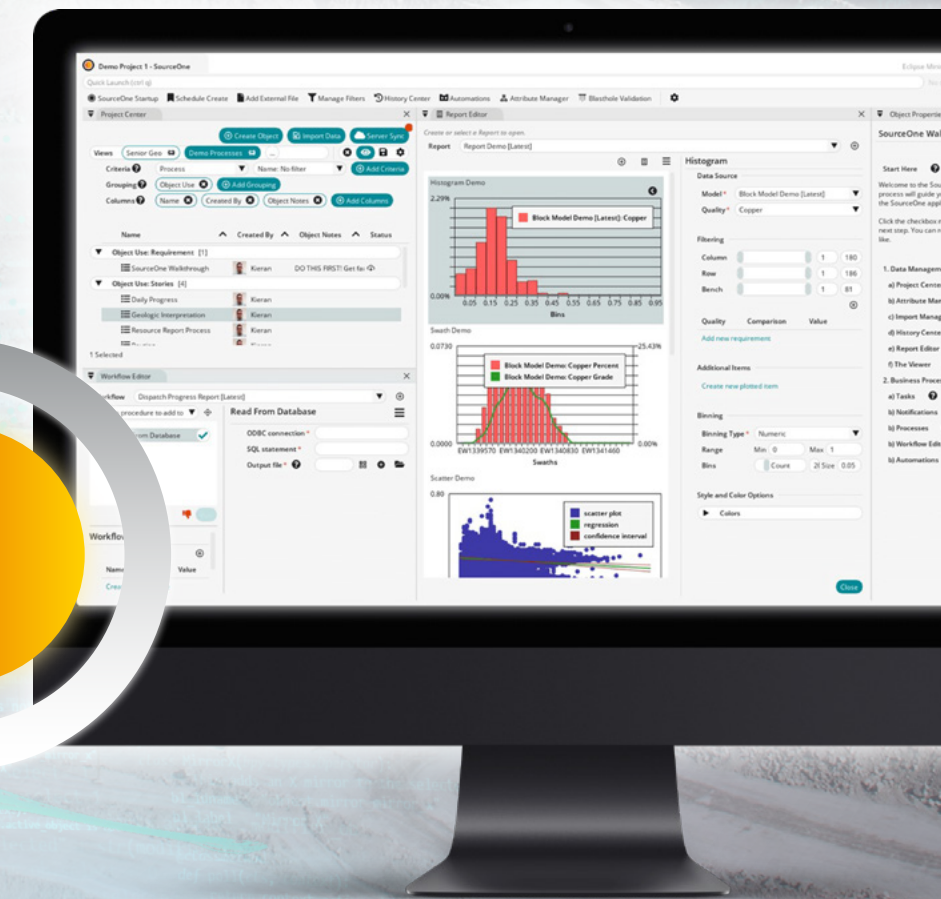
SourceOne architectural flexibility is remarkable. It has a hybrid approach that combines the best of the cloud and local qualities: although it is a cloud,

SourceOne EKPS can also perform complex, mining-specific activities such as digital mine plan creation, modeling, plan compliance, material tracking and others.

collaborative solution, it has the speed and responsiveness of desktop-only applications, as it allows for a copy of the data to be stored locally in the user's machine, enabling the users to work offline when practical, and save money on cloud computing time.

The server can be housed on premises in a customer's own data center or be accessible as a service via a public cloud offering. Either way, the key is that the customer owns the data, has control over who has access to the data and from which devices, determines how it's updated, etc.

Availability via the cloud means companies can deploy the SourceOne Enterprise Knowledge Performance System as quickly as they need to. Registered users of the System can log into a Desktop or Web-based application via a username and password. Each login has a set of roles and permissions that determine what users can access and what actions they can perform within the System.





SourceOne® is an **Enterprise Knowledge Performance System** with intelligent integration that augments the data with **context**, **history**, and **digital workflow automation** to provide **breakthrough insights** and enable strategic decision-making in near real-time driving value generation at all organizational levels.



eclipsemining.com

Copyright © 2021-2022 Eclipse Mining Technologies, LLC. ALL RIGHTS RESERVED.
Eclipse Mining Technologies is a registered trademark of Eclipse Mining Technologies, LLC.
SourceOne is a trademark of Eclipse Mining Technologies, LLC.
This document makes no offers, representations or warranties [express or implied].
Information and data contained in this document is for general reference only and may change at any time.