



Open Data in Mining

“ERIK JOHNSON HAS OVER 20 YEARS OF EXPERIENCE LEADING HIGHLY SKILLED TEAMS IN THE CREATION OF HIGH-PERFORMANCE ENGINEERING AND SCIENTIFIC SOFTWARE FOR THE MINING INDUSTRY.”

His industry knowledge and background in mining engineering, geology, geophysics, and geo modeling, provides insight into how we view and use the invaluable asset of DATA in mining.

This is the first in a series of articles that discusses the realities of open data in the mining industry.

This first article outlines the current state of the industry, and highlights the strengths, weaknesses and misconceptions about how we work with data in mining.

The next article expands on the possibilities of how true open data in an agile, enterprise platform can push the envelope of how we work with data in the mining industry.

The final article will tie the two together, and look to the future of mining data applications.

DO YOU HAVE OPEN DATA?

Most people, who think they have open data, really don't.

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, fails 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 platform.

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 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 even impede 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 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. If you have a mine planning solution and want to obtain all the data used in a certain area of the pit (drillholes, blastholes, models, workflows, schedules, etc.), it is a nearly impossible task.

DO DATABASE SERVERS OFFER A SOLUTION?

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

If a 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.

WHERE DO WE GO FROM HERE?

There are various efforts underway to standardize interchange formats, but with little progress towards actual open data. The formats are useful when moving data from one closed system into another closed system, but the scope only addresses simple mining data types such as geometry and block models, and it addresses the data as a single entity. For example, a single block model.

We need more. We need accessible, searchable, contextual, fully described, vendor agnostic, correlate-able open data in a unifying platform.

Think about the current state of your mining data, and ask yourself these questions:

- Where is your data?
- Can you access all versions of all data anytime and anywhere from a single source without licensing restrictions?
- Is your data in a format that is easily understood by third party tools and solutions?
- Can you filter and query metadata (history, context) for analysis?
- **Do you really have open data? And if not, can you imagine the possibilities?**

My next article expands on the possibilities of how true open data in an agile, enterprise platform can push the envelope of how we work with data in the mining industry.

THROUGH PRODUCTS THAT REVOLUTIONIZE DATA CONNECTIVITY AND DATA MANAGEMENT,
THEIR DECADES OF EXPERIENCE IN THE INDUSTRY AND FREEDOM FROM RESTRICTIVE
LEGACY TECHNOLOGY, THE ECLIPSE TEAM IS UNIQUELY EQUIPPED TO BRING A MUCH-
DEMANDED SEA OF CHANGE TO THE INDUSTRY.

A NEW STANDARD FOR MINING SOFTWARE.