Management of Mine Water Quality Data from Permitting to Compliance to Closure

Susan Wyman[1], Makinsey Harmon[2]

Whetstone Associates, 104 W. Ruby Avenue, Gunnison, Colorado 81230, USA;
[1]swyman@whetstone-associates.com; [2]mharmon@whetstone-associates.com

Abstract Data management is a critical link between effective mine water monitoring efforts and informed data analysis and decision making. Developing efficient systems for sample tracking, data validation, storage, retrieval, presentation, and analysis are increasingly important in organizations with limited time and resources. Mine permitting in the United States requires ever-increasing standards for verification and validation of laboratory water quality data. Routine compliance reporting, evaluation of Best Management Practices (BMP) effectiveness, monitoring plan modification, and closure decision-making also rely on accurate and accessible water quality data. Today's technical professionals expect that water quality data be understandable, usable, sharable, secure, and transferable among interested parties.

Keywords environmental database, data validation, statistical analysis, mine permitting, compliance monitoring

Introduction

Large quantities of environmental data are generated during mine permitting, operation, and closure. Successful projects require efficient, accurate, and cost-effective management of environmental data. Gone are the days of dozens of spreadsheet files as the central repository for environmental data. Mining companies, regulators, and consultants have made great strides over the last 20 years toward robust water quality and environmental database systems. Centralized storage of environmental data provides an "official" version of the data, which serves as a valuable resource for project consistency and accuracy. Data management is a critical link between effective mine water monitoring efforts and informed data analysis and decision making.

The purpose of this paper is to describe the relational database structure, list some of the software currently available for managing environmental data, provide general requirements for environmental databases at various phases of the mining project, and provide specific tips and tricks unique to databases for mining projects.

Alternatives to Relational Databases

Several alternatives exist for managing environmental data, depending on the size of the project, scope of the investigation, budget, and user capabilities.

Spreadsheets –Spreadsheets such as Excel are convenient ways to chart and analyze data. Environmental data is typically presented in crosstab format, with samples listed across the top and parameters down the side, or vice-versa. However, spreadsheets are limited in size (1,048,576 rows by 16,384 columns in Excel 2010), are easily editable, and standards are typically not enforceable. Users can delete data within a spreadsheet, and most spreadsheets are not set up to require values or enforce restrictions on the data being entered. While useful for data presentation and transfer, spreadsheets are not recommended as a se-

cure, archival repository for large environmental datasets.

Flat files – The laboratory electronic data deliverable (EDD) is typically provided as a flat file, consisting of one analytical result per row. Flat files take up a lot of file space due to duplication (each record contains information for the site, station, date, time, and sample ID). Although a spreadsheet flat file may be sorted and filtered, flat files are generally difficult to read and interpret. The flat file is a useful mechanism for transferring environmental data, but not for using the data.

Laboratory Information Management System (LIMS) – Some laboratories provide online access to their LIMS, which allows for downloading flat-file data or exporting crosstab queries to Excel spreadsheets. Laboratories typically limit the length of time that data is available, and users are limited to software features developed by the lab's IT team. Queries are limited to laboratory results (not lithology, depth, sub-basin, or other spatial qualities) and are therefore less robust than with a fully developed environmental relational database.

Environmental Relational Databases

Relational databases organize the data into tables, and link those tables based on defined relationships. These relationships enable the retrieval and combination of data from one or more tables with a single query. At the most fundamental level, an environmental relational database will consist of Stations, Samples, and Results in a one-to-many ($1 \rightarrow \infty$) relationship (Fig. 1). Lab parameters are listed in a lookup table (LUT), also known as a valid value list (VVL), which forces the database to use consistent parameter names (*e.g.* "Mercury, dissolved" and not "Hg, diss" or "Mercury, diss."). If a lab is reporting parameters using non-standard names, the Parameter_Aliases table is used to generate consistent parameter IDs.

A more complex relational database structure for a mine site is shown in Fig. 2.Although the relationships may seem complicated at first, the fundamental database structure is the same (Stations $1 \rightarrow \infty$ Samples $1 \rightarrow \infty$ Results). Relationships are set up to investigate correlations between borehole geology, lithology, and water chemistry (Fig. 2). Samples can be grouped by station type (monitoring well, domestic well, seep, spring, pond, exploration borehole), sub-basin, site, lithologic unit, or nearly any common criteria. Data validation can be automated by tracking cooler receiving temperatures, sample holding times, blind field duplicate samples, and laboratory QC batches (Fig. 2).

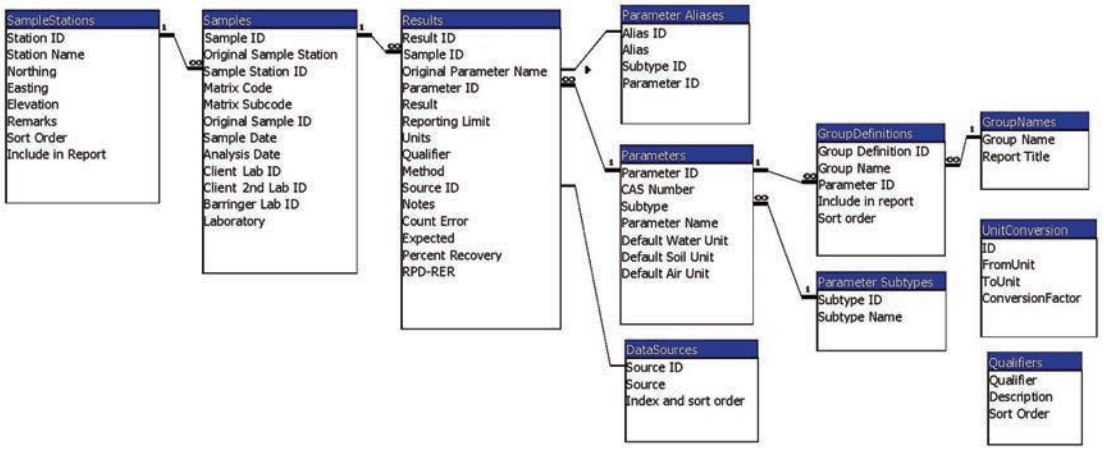

Fig. 1 *A simplified Environmental Relational Database contains tables for Stations, Samples, and Results.*

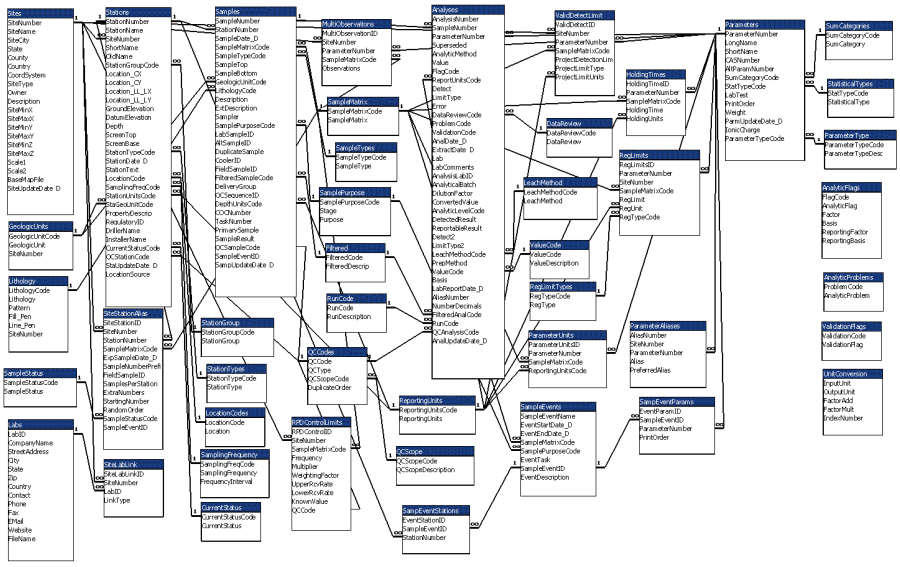

***Fig. 2** Relationships in a standard Environmental Relational Database.*

"Bulk data" refers to data that doesn't easily fit into the samples/analyses data model, such as time sequence data such as from level loggers and multi-parameter probes, downhole sensor data sensor, shaft or well water quality profiles, and two- and three-dimensional survey data. These datasets often contain large numbers of sample points that are best organized differently than typical laboratory results. While such datasets were typically stored outside of an environmental relational database, some databases have the ability to efficiently handle, store, and retrieve bulk data.

Relational Database Software

Numerous environmental database software packages are available commercially, including EIM, ChemPoint/ChemStat, Enviro Data/EnviroSpase, EQWin, Visual Site Manager (VSM), DataSight, HydroGeo Analyst, EQuIS, ESdat, and others. Historically, custom environmental databases were developed for specific projects using Oracle, Access, SQL Server, or other common database platforms. However, many projects are moving toward standardized, supported, off-the-shelf software. Several factors may be considered in selecting the environmental relational database software for a mining project:

- Initial and ongoing costs and number of user licenses
- Ease of use
- Portability, transferability, and shareability between users
- Ability to import and export between applications (*e.g.* Surfer, RockWare, EVS, EnviroInsite. EPA's ProUCL, Air Force ERPIMS)
- Regulatory requirements (some agencies require a specific software package, for example Colorado Division of Public Health and Environment formerly required EQuISand has now adopted EnviroData®. Some agencies simply require that the data be exportable between different databases).
- Support, flexibility

Our firm has standardized most of our mine water quality and geochemistry databases on the Enviro Data® software based on technical support, flexibility, and price (US$1,000 Viewer, US$4,000 Single Use Full, US$5,000 Concurrent Use Full). Enviro Data® operates on the Microsoft Access® platform (for up to about 200,000 records on a network or 500,000 records on a standalone PC) and on Oracle or SQL Server platforms for larger database applications. We also maintain databases

in MS Access and ChemPoint/ChemStat. The concepts presented below are applicable to most environmental relational databases, and are not unique to a specific software package.

Numerous water quality relational databases have been developed for mine sites using a user interface that runs on the Microsoft Access platform because this software is inexpensive and the databases are easily shared. Customization of the default database lookup tables is essential for managing data sets generated by routine mine water quality monitoring and unique geochemical studies related to waste rock leachates and whole rock analyses. Standard tools within the database software are combined with customized procedures to meet today's stringent requirements for data validation and verification. Practical databases accommodate an infinite number of problem flags (reason codes) which map to a finite number of validation flags for data presentation and reporting.

Relational Databases for Mine Permitting

One purpose of a relational database for mine permitting is to manage data from groundwater, surface water, soil, whole rock, geochemical leachates, and less often, from air and tissue samples for use in establishing baseline and estimating potential impacts from the proposed project. The care that goes in to developing study plans, logging boreholes, collecting samples, and analyzing samples must continue with the environmental database. Database considerations for mine permitting include the following:

- The framework for the long-lived project database is typically developed during the mine permitting phase. Effort spent in setting up the database correctly will pay off in the long-term.
- Accurate lists of parameters and parameter aliases are set up early in the process. This is not the time to have the summer intern mix all of the phosphate (PO_4) and phosphorus (P) results.
- Use the database to quickly check that the detection limits are well below the regulatory limits. If regulatory limits differ in specific media (*i.e.* surface water or groundwater), the Sampling and Analysis Plan should set the detection limits below the lower of the two regulatory limits for potential impact analysis.
- The initial project analytical suite typically contains a comprehensive list of parameters to be analyzed. The database should provide an efficient and trusted means for paring down the analytical suite to specific constituents of concern, using statistically accepted methodology.
- The database may be used to determine baseline levels of constituents of concern (COCs) using commonly approved statistical methods. The definitions of "baseline concentrations", "upper protection limits", or "background levels" may vary based on state regulations. A simple calculation of mean plus two standard deviations may be appropriate for establishing background of major ions. Statistical analysis of censored (non-detected) results is more complicated and must typically follow established statistical protocols. Internal calculations can be coded in the database, or most databases can export to *.csv for import to EPA's ProUCL for determination of baseline concentrations.
- Legacy datasets often exist where a historical mine is being permitted in the present day. These datasets may lack the laboratory QC packages, reporting units, or portions of the hard copy data may be illegible. Project-specific flags may be added to any questionable data, while well-documented historical data may be used without flagging.

The baseline data is used for decision-making by project proponents and regulators. As such, the database must provide robust and accurate datasets for defensible evaluation of potential project impacts.

Relational Databases for Mine Operations and Compliance Monitoring

During operations, most mining projects are required to meet general or site-specific water quality standards at designated points of compliance (POCs). Database considerations for compliance monitoring include the following:

- A relational database is useful for rapid comparison of laboratory results against the site regulatory limits.
- The comparison must be made using validated results. Laboratory analytic flags are generated by the lab and relate to laboratory instrument precision, comparability, accuracy, and bias based on laboratory control samples, laboratory control duplicates, calibration verification samples, matrix spike, matrix spike duplicates, laboratory duplicates, method blanks, and equipment blank samples in accordance with the laboratory Quality Assurance Plan. Validation flags are assigned by the independent data validation officer based on receiving temperatures, blind duplicate RPDs, cation-anion balance, TDS ratio, and a number of other factors (Table 1). The data validation officer may identify information for a particular sample that is retained in detail in the database but does not result in a specific validation flag code.
- An exceedence of a specific parameter during one sampling event does not necessarily indicate that the site is out of compliance. The database may provide statistical evidence that the result was an outlier, particularly if no other changes in trace metal or major ion chemistry are seen. If the exceedence is recognized early, the lab can re-run the sample either within or out of holding times. (Depending on the lab, samples may be held for three to six months after analysis.) The mine permit will often specify a matrix for confirmation sampling and accelerated monitoring, leading up to mitigation and monitoring to verify the effectiveness of the mitigation measures.

Problem Code	Analytic Problem Description	Maps to Validation Code[1]
D1	Dilution due to matrix interference	V[2]
DL	Dilution, no validation flag required	v
D+	Dilution introduced high bias	J+
D-	Dilution introduced low bias	J-
DT	Dissolved or total not indicated (legacy data)	Q
EA	Estim., <-MDL due to matrix interference	J
EB	Est. based on equip blank detection	FB
EQ	Extended qualifer, no ValidFlag required	v
FB	Est. based on field blank detection	FB
FD	Field duplicate %RPD criteria not met	J
H	Exceeds holding time	J
H1	Sample rec'd & analyzed past hold time	J
H2	Rec'd in time, analyzed past hold time	J
H3	Orig in hold time, re-analyzed past hold	J
J	Estimated value	J
R	Rejected value	R
Rt	Rejected on basis of TDS ratio	R
T0	Cooler <0 deg C	T[3]
T10	Cooler >10 deg C	T[3]
T2	Cooler <2 deg C	v
T6	Cooler >6 deg C	T[3]
z	unknown	z

Table 1 *Example Problem Codes and Associated Validation Flags.*

[1] *Validation codes U, J, R are derived from the EPA Functional Guidelines for Inorganic Data Review (EPA 2004)*

[2] *Data validator may override validation codes based on results of QC process (e.g. D1 may map to J in validation code if warranted by the QC findings)*

[3] *Only specific analytes are temperature sensitive; metals preserved with HNO_3 do not require flagging on the basis of receiving temperature*

Relational Databases for Mine Closure

Mining projects may be required to meet water quality standards at designated POCs for a fixed time (the post-mining monitoring period). Database considerations for mine closure include the following:

- After mining ceases, bonding is in place for post-mining monitoring and mitigation.
- The relational database often contains decades of monitoring data which may be useful for reducing the list of parameters, to save on analytical costs.
- Automation of data collection, data capture, and reporting are important to save costs in the post-production phase of mining operations.
- The database should facilitate statistical trend analysis (such as Mann-Kendall analysis or quantitative linear trend analysis) to identify whether any trends exist in post-mining water quality.
- The database should have the ability to plot data "on the fly" to visually identify trends and potential outliers.

Mine closure monitoring may be tied to a fixed period of time, or may be linked to demonstrated performance or benchmarks. In either case, efficient evaluation and presentation of monitoring data is critical for mine closure and bond release.

Conclusions

Data management is a critical link between effective mine water monitoring efforts and informed data analysis and decision making. Developing efficient systems for sample tracking, data validation, storage, retrieval, presentation, and analysis are increasingly important in organizations with limited time and resources. Environmental relational databases can be developed using a number of readily available software packages. A well formulated and properly implemented database will serve as a robust tool for mine permitting, compliance monitoring, and mine closure.

Timely data management is required to provide information for compliance monitoring programs that require reporting of water quality results that exceed established permit reporting levels (PRLs). Statistics can be analyzed from within the database, or results can be exported to a variety of formats, including EPA's ProUCL, for rigorous statistical analysis. Given the current agency trend of requiring third party validation of datasets, robust database management is necessary for meeting data quality requirements in-house or for "farming out" the development (input) side of relational databases. Once a database is correctly developed and populated, complete or limited datasets can be extracted as necessary. Mining companies, agencies, and consulting firms can easily utilize these datasets using the UI viewer software.

This approach to managing water quality, whole rock, and geochemical data at mine sites provides a cost effective method for meeting today's increasingly complex requirements for data capture, storage, analysis, and reporting.

References

GeoTech Computer Systems 2013.Enviro Data software. http://www.geotech.com/envirodata.htm

Rich, David 2002. Relational Management and Display of Site Environmental Data, Lewis Publishers, 429 pp.

United States Environmental Protection Agency (2004) USEPA Contract Laboratory Program Functional Guidelines for Inorganic Data Review Final. EPA 540-R-04-0. October 2004.