Storing Expensive and Valuable Environmental Chemistry Data (Newsletter)*

March 28, 2013

Environmental chemistry data are expensive to obtain and valuable and need proper care in storage so they retain their value and return your investment in them. Expenses start with permit application preparation and baseline collections and continue through monitoring programs, analyses, and reporting.

The proper storage of environmental data is in an appropriately designed database, but many organizations use spreadsheets instead because they are readily available and easy for individuals to learn and use. Proper structure and formatting of spreadsheets reduces duplication, facilitates organizing and finding data, and promotes transfer to statistical analysis software.

There are only a few required data attributes (the column headings): collection location, collection date, constituent name, measured quantity, and a quantification flag. Other information can be in additional columns but are not required for regulatory reporting and operational insights analyses. The rows in the spreadsheet contain all the information for a single set of related attributes.

Data format needs to be consistent. Dates, for example, should use the International Standards Organization (ISO) format of YYYY-MM-DD (e.g., 2013-04-01 for April 1, 2013). This ensures data will correctly sort regardless of when or where it was entered. It also quickly identifies duplicate entries. Censored data (those values whose concentrations – the signals – cannot be distinguished from noise in in the analytical process) are reported in various formats by laboratories; most often using the less-than (<) symbol with the laboratory's reporting limit for that sample. This means the quantity column is treated at text by the spreadsheet and those cells cannot be used for computations. This is where the quantification flag mentioned above is used to fix this problem.

The ability to detect very low concentrations of chemicals varies by time (methods and instruments are constantly improving), laboratory, specific instruments, chemist, sample matrix, sample size and dilution, and other factors. There are multiple names and definitions for the censored (unknown value) data but a good general term is 'reporting limit.' When you receive laboratory

---

results that show censored data by having the concentration number preceded by a less-than symbol (<), enter only the number (the reporting limit) in the quantity column and use the quantification flag column to indicate it is a censored value. This flag is a binary, True/False, indicator. For consistency with statistical software analyzing your data enter a value of 0 (zero) when the concentration is quantified (uncensored) and a value of 1 (one) when the concentration is below the reporting limit (censored). To make it easy for everyone to remember these flag values name the column with a reminder; for example, 'BRLeq1' which means 'Below Reporting Limit equals 1'.

As your environmental chemistry data increase you will find different reporting limits for the same constituent. Enter each one with the number provided by the analytical laboratory. There are several ways of calculating and plotting summary statistics and comparing groups with multiple reporting limits. Depending on the concerns to be addressed, the highest reporting limit alone can be used.

These guidelines are a few of the ways to preserve the value of your environmental chemistry data and facilitate the analyses which release the maximum amount of information they contain.