

# Why the Cost of Data Validation May Be Higher Than You Expect



Engineers, environmental consultants and regulatory officials make important and expensive decisions based upon analytical results. Whenever the reliability of the laboratory data can be called into question, an independent, third party assessment of the data (i.e. data validation) will play an important role. Data validation is the process by which data are evaluated on an analyte- and sample-specific basis, that extends beyond method, procedural or contractual compliance, to determine the analytical quality of a specific data set. Data validation confirms whether or not the results meet the overall quality requirements of the intended use.

Various levels of data validation can be performed, ranging from a limited review of the sample analysis reports and QC summaries (Stage 2A) to an in-depth review of the raw data, transcriptions, and calculations (Stages 3 and 4). The level of validation effort for a particular project depends largely on the data quality objectives (DQOs) or how the data will be used. However, other factors can also come into play, such as quality of the laboratory, site history, and time and budget constraints.

The cost of data validation may exceed the cost of the laboratory analyses, which can be difficult for the consultant/engineer to understand. While some components of data validation can be automated (assessment of quality element performance [spike recoveries and duplicate precision] and comparison to the sample results, and application of data qualifiers) using the electronic data deliverable (EDD), much of the review must be performed manually. For example, most EDDs do not contain the calibration information. Calibration is an important component of accuracy. The lower concentrations in the calibration must support the reporting limit for an analyte. As laboratories are required to meet lower and lower reporting limits, review of the calibration data is critical to determine whether or not adequate sensitivity and selectivity have been achieved to support the project-required reporting limits.

The process of data validation also depends upon a thorough understanding of analytical chemistry, the methods employed, and their application to reported sample results. A few of the elements this process includes are:

- Review and verification of sample chromatograms and mass spectral data for volatile and semi-volatile analyses,
- Review and verification of sample chromatograms for PCB patterns and identification, and
- Inter-parameter relationships; do the data make sense?

Commercial environmental laboratories analyze thousands of samples each month, and the processes for sample handling will directly affect the costs of data validation. It is important to understand why this is the case. Some examples of laboratory practices that lead to such increases are discussed below:

Example 1 – Validating a sample delivery group (SDG) with only one groundwater and one soil for volatile organic compounds (VOCs) may cost almost the same as an SDG containing more samples.

Water samples and soil samples typically are analyzed on different instruments because they are different matrices and calibrations will differ. As a result, there will be two sets of method blanks, initial and continuing calibration data, laboratory control sample (LCS) data, and potentially, two matrix spike (MS)/MS Duplicate (MSD) samples to review. As a result, the time to perform the review for each single sample isn't much different than if the SDG contained five water samples and five soil samples.

Example 2 – Meeting short hold times may require the laboratory to analyze across multiple instruments, even if the samples are all the same matrix.

Because of the high volume of samples that laboratories must analyze on a daily basis, meeting analytical hold times can become an issue. As a result, samples within a given SDG are often split up and analyzed on multiple instruments. If five samples are analyzed on three different instruments instead of a single instrument, the amount of time to review and assess the data (separate calibrations and instrument-specific QC samples) has, at a minimum, tripled.

Example 3 – Sample preparation or extraction batching creates another level of complexity.

Samples for many parameters must undergo extraction or other preparation steps prior to analysis. The extraction (preparation) step can add another layer of effort for validation. It is not uncommon for laboratories to split samples from a single SDG among multiple preparation batches. Each preparation batch must include batch-specific QC samples (at a minimum, a method blank and LCS). In this scenario, the level of effort for validation has at least doubled before the samples have even been loaded on an instrument for analysis. If the samples from a single extraction batch are then spread out over multiple instruments, the level increases even more, for the reasons discussed before.

Example 4 – Samples in a single sampling event are collected over multiple days.

The sampling schedule also affects the level of effort for validation. If samples are collected over several days and shipped to the laboratory each day of collection, it is likely that the analyses will be performed over several days, unless the field and/or laboratory team has been instructed to do otherwise.

Working with both the field crew and the laboratory to minimize some of these issues, where possible, is beneficial to the project as a whole. Taking into consideration sample analysis hold times, fewer shipments to the laboratory as well as working with the laboratory to maximize preparation and analytical batches would help keep project costs down.

These are only a few examples of the many issues that can impact the level of effort and costs incurred in data validation. Advance planning and effective communication of project requirements are essential to keeping quality up and costs down.

Data from laboratory analyses are essential to risk assessments, extent-of-contamination studies, feasibility studies, measurement of natural attenuation, and other compliance monitoring. Tremendous amounts of resources and capital are expended to produce the laboratory results. Without the benefit of data validation, the user may take data at face value, without understanding the implications of quality control deficiencies that would be identified and interpreted by validation. The work product of data validation is scientifically meaningful. It determines if the data reported by the laboratory are valid and

usable for their intended purpose, legally defensible, and technically valid. And where there are deviations, the potentials for bias and variability are identified. While the cost of validation may seem high by itself, keep in mind that it is generally only a fraction of the costs of the sampling and analytical expenditures. It is more than justified by ensuring that decisions based on invalid and/or indefensible data are avoided, and by the savings in re-sampling and re-analysis costs it can prevent.