Analytical Data Quality Review: Verification, Validation, and Usability

ITRC has developed a series of fact sheets that summarizes the latest science, engineering, and technologies regarding environmental data management (EDM) best practices. This fact sheet describes:

- the importance of verification in relation to data quality
- what makes validation of analytical chemistry data different from other types of validation tasks
- how verification and validation are used to assess overall data quality and usability
- various federal and state resources related to data quality review

1. INTRODUCTION

Some degree of data review should be performed on all data that is used for decision-making. How extensive that review needs to be depends on the intended use of the data and any regulatory requirements.

A clear understanding of the data needs and expectations is crucial. Establishing data quality objectives (DQOs) and requirements to be met before data is used in project decisions, evaluations, and conclusions is essential to managing defensible environmental data. For some projects, these DQOs are formally developed in various planning documents, such as quality assurance project plans (QAPPs), where data quality indicators of precision, accuracy/bias, representativeness, comparability, completeness, and sensitivity (PARCCS) are well defined. Less formally, DQOs might be established in various data management standard operating procedures (SOPs) or documented best practices.

This document will discuss standard terminology and concepts associated with analytical data quality review and its use in assessing overall data quality and usability

2. DEFINITIONS AND OVERVIEW

2.1 Verification

The USEPA Guidance on Environmental Data Verification and Data Validation (USEPA 2002) defines verification as “the process of evaluating the completeness, correctness, and conformance/compliance of a specific data set against the method, procedural, or contractual requirements.” This definition focuses on environmental analytical data and would include activities such as reviewing sample chains of custody (COC), comparing data in electronic data deliverables (EDDs) to paper or electronic laboratory reports, and reviewing laboratory data packages against project PARCCS criteria.

While the focus of this document is on analytical data review, all environmental data collection efforts will have some verification needs. This verification could include review of metadata, proper sample naming conventions, location data, lithology, water levels, field observations, or keyword tags on media files, for example. This type of verification is crucial for assessing overall data quality. The Using Data Quality Dimensions to Assess and Manage Data Quality subtopic sheet provides a table of key considerations for assessing the data quality dimensions of integrity, unambiguity, consistency, completeness, and correctness for most types of environmental project data, including analytical and non-analytical data, throughout the project lifecycle.

Data verification needs will vary depending on program/project DQOs and regulatory requirements, but all data collection activities should have a standard, and preferably documented, verification process. Verification tasks should be assigned to an individual familiar with the data being verified and ideally performed as soon as possible to identify errors while there may still be opportunities to correct and improve the overall quality of the data.

If laboratory QC data is provided in an EDD with the sample data, it can be valuable to perform a data quality screening using an automated review tool during the verification process. These tools utilize user-defined PARCCS criteria and/or historical data to provide quick, reliable, consistent review of data sets. These screening tools can help identify outliers, flag potentially unusable data, or in the case of trend charts (USEPA, 2011), even show changes over time.

2.2 Validation

Validation, in the context of environmental analytical data quality, is a formal analyte/sample specific review process that extends beyond verification to determine the analytical quality of a specific data set (USEPA 2002). Due to variability in field collection and laboratory analysis, it is highly unlikely that analytical data collected for any regulatory program is going to always meet established PARCCS criteria. Verification can determine if your data met various data quality indicators, but validation goes a step further in determining (and documenting) how failure to meet method, procedural, or contractual requirements impacts the quality of the associated data.

While verification, if performed promptly, may provide opportunities for improving data quality, validation in the above context cannot change the quality of an analytical data set, it only defines it. In other words, validation cannot turn low quality data into high quality data, but rather identifies areas where data may lack the quality needed for a project DQO and/or regulatory standard.

It is important to acknowledge that the term “validation” is a general term that can have different meanings depending on the industry and can even have different definitions within the environmental data community. Software engineers, database developers, GIS professionals, statisticians, analytical laboratory personnel, and environmental chemistry validators might all perform validation on data in some capacity. In some work environments, a project team could include all of these individuals, and miscommunication becomes more likely. In very general terms, “validation” is the process of ensuring accuracy and quality of data. In practice, how each group of people performs validation, and the goals of that validation, can differ. It is extremely important when different groups are interacting with each other that everyone understands what type of validation is being discussed, who will be performing the task, and what the goals of the validation task are.

For the context of this document, we are discussing validation of analytical data performed by an environmental chemistry validator after the receipt of a final laboratory data package and before final data analysis. When validation status and/or review level is recorded within an environmental data management system (EDMS), those entries are typically related to this specific analytical data validation process, since analytical chemistry validation is often a component of regulatory compliance. Other types of validation may be performed on the same data before (such as validation performed at the analytical laboratory) or could be performed after analytical chemistry validation (during statistical analysis), but this document is narrow in focus. For the remainder of this fact sheet, we will refrain from using the term “validation” outside of the specific meaning defined for

Documenting Quality of Non-Analytical Data

Analytical data typically has well defined data quality objectives that can be easily measured and a defined system of documenting quality using validation qualifiers. When these qualifiers are stored within an EDMS, they are easily accessible to all data users alongside their associated data. In contrast, documenting the quality of non-analytical data is often less defined within the environmental protection industry. If information on the quality of non-analytical data is stored within an EDMS, it is much more likely to vary between organizations. It is possible that information related to quality for non-analytical data is only stored in metadata or other supporting project documents outside of the EDMS, and data users may or may not know that it exists or where to find the supplemental information. While developing a standardized process for documenting the quality of non-analytical environmental chemistry data is beyond the scope of these documents, the ITRC EDM Best Practices Team recognizes this as an area for improvement within the industry.

analytical chemistry data at the beginning of this section, and instead use “verification” or the generic term “review” for all other data review activities.

2.3 Overall Program / Project Data Quality

Verification is the first step in determining overall data quality and compliance with DQOs and any defined PARCCS criteria. Validation builds off verification and helps define analytical data quality through the assignment of validation qualifiers. You can omit formal validation from the overall data quality review process, but the analytical quality of the data will be less defined and could add additional time and effort to the overall data quality assessment. As shown in Figure 1, verification and validation are important steps toward assessing data quality, but do not form a complete picture of overall data quality alone. All data quality dimensions (see Using Data Quality Dimensions to Assess and Manage Data Quality subtopic sheet), including the quality of supporting non-analytical data, should be considered before data are used for final reporting or

decision-making. For example, if analytical data are assigned to an incorrect sample location in a data table or figure, the quality of analytical data becomes largely irrelevant. Even small data errors or omissions can have drastic impacts on overall data quality.

2.4 Data Usability

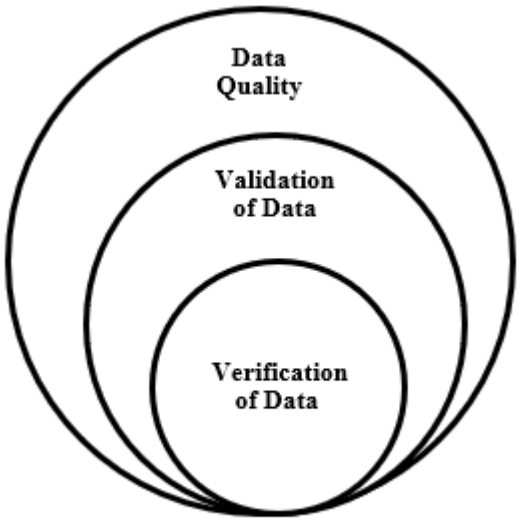


Figure 1. Data verification and validation in relation to data quality.