United States Environmental Protection Agency
Office of Solid Waste and Emergency Response (5102G)
EPA 542-R-01-014
October 2001
www.epa.gov
www.clu-in.org

EPA **Current Perspectives in Site Remediation and Monitoring**

CLARIFYING DQO TERMINOLOGY USAGE TO SUPPORT MODERNIZATION OF SITE CLEANUP PRACTICE

D. M. Crumbling[1]

Introduction

The appropriate use of field analytical technologies and dynamic work plans could dramatically improve the cost-effectiveness of environmental restoration activities. EPA's Technology Innovation Office (TIO) has been developing classroom and Internet web-based courses to promote adoption of these tools and strategies. TIO's experience has been that a common language to unambiguously communicate technical concepts is vital if regulators, stakeholders, and practitioners are to negotiate, plan, and implement these projects to their mutual satisfaction.

Systematic planning is critical to the successful implementation of hazardous site characterization and cleanup projects. EPA's "DQO process" has been around for many years, and the "DQO" terminology is used extensively. Unfortunately, over the years the terminology has been used in ambiguous or contradictory ways, and this has resulted in confusion about what terms mean and how they are to be used. It is thus useful to clarify the relationship between DQO-related terms as descriptively and concretely as possible. The discussion provided here has been reviewed by the primary DQO and data quality coordinators within the EPA Headquarters offices of the Office of Solid Waste, the Office of Emergency and Remedial Response, the Office of Environmental Information, and the Quality Staff to ensure that the concepts presented are consistent with EPA's original intent for DQO terminology and with the direction program needs are currently taking. Any questions or comments about this paper should be directed to the EPA Technology Innovation Office through the Clu-In "Comments" form (*http://cluin.org/gbook.cfm*) or to (703) 603-9910.

This paper does not intend to provide all-inclusive definitions that can be found elsewhere in EPA guidances, nor does it attempt to provide all-inclusive coverage of each topic. It is intended to provide, as briefly yet unambiguously as possible, a basic conceptual understanding of DQO-related terms in a way that **facilitates systematic project planning in the context of site cleanups**. A list of descriptions for DQO-related terms and concepts appears first in this paper, followed by a more intensive discussion of the working interrelationships between these concepts. It is entirely possible that other parties use terms other than these to communicate the same concepts. The actual terms used are less important than the ability of parties involved in site cleanup projects to have a "meeting of minds" and clearly communicate the concepts, since the concepts are basic to the scientific validity of environmental decisions and to the data that support those decisions. A common conceptual framework could help all within the hazardous waste community better communicate our goals and results, fostering more cost-effective planning and implementation of projects.

[1] EPA, Technology Innovation Office

Descriptions for DQO-Related Terms

Data Quality Objectives (DQO) Process

This is a *systematic, iterative, and flexible planning process based on the scientific method*. The DQO process was developed by EPA to provide a common structure and terminology to practitioners designing environmental data generation operations. The DQO *process* produces quantitative and/or qualitative statements (called "the DQOs," see below) that express the project-specific decision goals. The DQOs then are used to guide the design of sampling and analysis plans that will be able to cost-effectively produce the "right kind of data."

An important part of the DQO process is developing an understanding of how *uncertainties* can impact the decision-making process. A systematic planning process, such as the DQO process, identifies *what* the goals are and what the consequences may be if the decisions are made in error. It is within the realm of values (not science) for decision-makers (as representatives of society as a whole) to estimate how certain (i.e., confident) they want to be before making decisions that will either impact, or be impacted by, environmental conditions. When technically feasible, an expression of statistical certainty may be desirable because it can be more "objective" (if it is done in a technically valid manner). But in the environmental field, mathematical (e.g., statistical) treatment of "uncertainty" may not always be technically feasible or even necessary. Qualitative expressions of decision confidence through the exercise of professional judgment (such as a "weight of evidence" approach) may well be sufficient, and in some cases, may be the only option available. An important part of systematic planning is identifying the *information gaps* that could cause a decision to be made in error. If the existence of information gaps increases the likelihood of decision error beyond what it acceptable, then it may be desirable to fill those gaps, if it is feasible to do so. Planning how to gathering environmental data that can acceptably fill information gaps is the purpose of the DQO process. Decision-makers should also keep in mind that, on occasional, systematic planning may indicate that it may be more cost-effective to simply go ahead and make the decision to take the most conservative (protective) action, rather than spend the resources needed to scientifically "prove" whether the protective action is absolutely necessary or not.

Sampling and analysis plans lay out the strategy to be used to gather needed data. Steps 1 through 6 of the DQO process provide the structure to help a project team articulate their project goals and decisions, the project's constraints (time, budget, etc.), and how much uncertainty they can tolerate in the final decision. These things must be thoroughly understood before the task of developing the data-gathering plans that can meet those goals within the given constraints is begun. Developing project-specific sampling (i.e., determining the number of samples, their locations, their volume, etc.) and analysis (i.e., selecting and modifying, as needed, the analytical preparation, cleanup, and determinative methods, the analytical QA/QC protocols, etc.) plans is the very last step (Step 7—"optimize the design") of the DQO process.

During Step 7, pre-existing site information should be sought and evaluated so that uncertainties that could impact the sampling and analysis plan can be evaluated as much as possible prior to finalizing the plan. For example, existing information about the mechanism(s) of contaminant(s) distribution and their likely environmental fate (degradation and/or redistribution in the environment) can be used to develop a conceptual model for the variability of contaminant concentrations and the media that should be sampled. Knowledge or suspicion that other contaminants may be present in the samples can guide consideration of alternate analytical methods able to cope with any analytical interferences that might arise. [More details about the development of sampling and analytical plans can be found in the article, "Guidelines for Preparing SAPs Using Systematic Planning and PBMS" (Jan/Feb 2001 issue of *Environmental Testing & Analysis*; also available as a pdf file on http://cluin.org/char1_edu.cfm#syst_plan).]

It should be noted that the DQO process is a *systematic planning process focused on generating project data.* The term "systematic planning" is often used to encompass the broad range of project activities that includes more than just data generation and interpretation activities. (See "Systematic Planning" below.) [More thorough discussions of DQOs and details of the DQO process can be found in various EPA Quality Assurance documents available on EPA's Quality Staff website: http://www.epa.gov/quality/qa_docs.html]

Data Quality Objectives

DQOs are qualitative and quantitative statements that translate non-technical project goals into technical project-specific decision goals. Project planners derive these technical DQOs from the non-technical social, economic, and/or regulatory objectives of the environmental program under which the project is being implemented. **DQOs are goal-oriented statements that establish the (technical) "bar" for overall decision quality or tolerable decision error in accordance with the (non-technical) objectives driving the project.** The DQOs for any particular project may, or may not, be highly specific in naming target elements, target media, and action levels along with the intended uses of the data. Project DQOs may be articulated at different technical levels depending on the intended audience. For communication with public stakeholders or in non-technical settings, DQOs will usually be summarized as simple, less technical statements. For communication between technical practitioners, project DQOs should be articulated in as specific and technically-detailed manner as possible to avoid ambiguities that could cause confusion or misunderstandings. In either case, DQOs summarize the outputs of the DQO (planning) process. Example statements below provide a flavor of simply worded DQO statements that summarize the highly technical statements that would lie behind the simply worded summaries and give them substance. **The most important point to note is that in no case do DQO statements directly set criteria for the quality of data** that will be gathered during implementation of the project. The process of determining the quality of data that will be needed to meet the project decision goals (i.e., "meet the DQOs") must be done *after* the DQOs are established. Quantitative DQOs express decision goals using numbers, such as quantitative expressions of decision certainty. Qualitative DQOs express decisions goals without specifying those goals in a quantitative manner.

Example of a less detailed, quantitative DQO: Determine with greater than 95% confidence that contaminated surface soil will not pose a human exposure hazard.

Example of a more detailed, quantitative DQO: Determine to a 90% degree of statistical certainty whether or not the concentration of mercury in each bin of soil is less than 96 ppm.

Example of a detailed, qualitative DQO: Determine the proper disposition of each bin of soil in real-time using a dynamic work plan and a field method able to turn-around lead (Pb) results on the soil samples within 2 hours of sample collection.

Even when expressed in technical terms, **DQOs should express "what"** (i.e., what decision) the data will ultimately support, **but should not specify "how"** that data will be generated (e.g., which analytical methods are to be used). Despite the name, "*Data Quality* Objectives," DQOs should be thought of as statements that express the *project objectives (or decisions) that the data (and its associated quality) will be expected to support.* As project objectives, DQOs serve to guide the eventual determination of the data quality that is needed to make good decisions, **but DQOs themselves should not attempt to directly define the specifics of that data quality**. Doing so short-circuits the systematic planning process, hindering the ability of project planners to optimize data collection designs to make projects more cost-effective (Step 7 of the DQO process). Various terms have been used that more intuitively express the originally intended concept of "DQO," including "*Decision* Quality Objectives"; "Decision Confidence Objectives (DCOs—used in the context of compliance monitoring—WTQA 2001 short course, "Regulation Writing under PBMS"); and "Project Quality Objectives (PQOs—used by EPA Region 1 in their Quality Assurance Project Plan Guidance)."

A discussion of the analytical flexibility inherent to U.S. EPA's waste programs and to SW-846, the methods manual used by these programs, is found in the paper, *Current Perspectives in Site Remediation and Monitoring: The Relationship between SW-846, PBMS and Innovative Analytical Technologies* [document number EPA 542-R-01-015; available on http://cluin.org/tiopersp/].

Data Quality

Data quality is a term that tends to be rather vaguely understood in the environmental community, despite its importance to the decision-making process. In addition, the term "data" is used to refer to many different kinds of information that is derived from very different kinds of data generation procedures. In the context of the DQO process, "data" generally refers to the measurement of some physical or chemical environmental property. Of greatest concern to the management of hazardous waste and contaminated sites is the measurement of toxic (or potentially toxic) chemicals in environmental media to which receptors may be exposed. In this context, "good" data quality tends to be linked in many minds with using the most sensitive or precise analysis procedures available. However, this view of data quality produces problems because the information value of that kind of data is limited not so much by the analytical procedures used (although that is certainly possible), but by the difficult task of ensuring representative sampling in heterogeneous environmental matrices.

Fortunately, EPA has recently clarified its intended meaning for the term “data quality” in its broadest sense by defining it as “the totality of features and characteristics of data that bear on its ability to meet the stated or implied needs and expectations of the customer “ (i.e., the data user). [This definition appears in the 2000 version of the Office of Environmental Information’s Quality Management Plan, entitled *Management System for Quality*.] Recent EPA guidance reinforces this understanding of data quality by stating that “…data quality, as a concept, is meaningful only when it relates to the intended use of the data. Data quality does not exist in a vacuum; one must know in what context a data set is to be used in order to establish a relevant yardstick for judging whether or not the data set is adequate” [from page 0-1 of *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9 QA00 Update)*. EPA 600/R-96/084; http://www.epa.gov/quality/qs-docs/g9-final.pdf].

Linking data quality directly to the data’s intended use provides a firm foundation for building a vocabulary that distinguishes the various components of data quality. For example, since analytical data are generated from samples, pre-analytical considerations (such as sample representativeness and sample integrity) are crucial when determining whether data are of sufficient quality to meet the user’s need to make correct decisions. Data quality can be broken broadly into the components of analytical quality (how reliable is the analytical procedure) and representativeness (selection of the samples and of the analytical method is appropriate to the intended use of the data). Non-representative sample selection produces “bad” data (misleading or meaningless information), even if the analytical quality on those samples was perfect.

Data Quality Indicators

DQIs are qualitative and quantitative measures of data quality “attributes.” Quality attributes are the descriptors (i.e., the words) used to express various properties of analytical data. DQIs are the measures of the individual data characteristics (the quality attributes) that collectively tend to be grouped under the general term “analytical data quality.” For instance, the data quality attribute of analytical sensitivity can by measured by different DQIs, such as instrument detection limit, sample detection limit, or quantification limit, each of which can be defined somewhat differently depending on the program or laboratory. See EPA QA/G-5 (1998 version) for more discussion on the topic of DQI (http://www.epa.gov/quality/qs-docs/g5-final.pdf). Another guidance document, EPA/G-5i, will explicitly discuss DQIs in much greater detail. EPA/G-5i is currently under development. Look for a peer-review draft to be posted in the future at http://www.epa.gov/quality/qa_docs.html]

Quality attributes (and the facets of data quality that they describe) include (but are not limited to) the following:

- Selectivity/specificity (describes what analytes the technique can “see” and discriminate from other target analytes or from similar-behaving, but non-target, substances);

- Sensitivity [depending whether “detection” or “quantification” is specified, describes the lowest concentration, or increment of concentration, that the technique is able to detect (although quantification may be highly uncertain) or quantitate with greater confidence];

- Bias (describes whether the technique produces results with a predictable deviation from the “true” value);

- Precision (describes how much random error there is in the measurement process or how reproducible the technique is);

- Completeness (describes whether valid data is produced for all the submitted samples, or just some fraction thereof); and

- Comparability (describes whether two data sets can be considered to be equivalent with respect to a common goal).

The familiar “PARCC parameters” have been considered to consist of 5 principal DQIs that include measures of precision, accuracy (used in this context to denote bias), representativeness, comparability, and completeness. Sensitivity (“S”) may also be included as a principal DQI. Precision, bias, and sensitivity describe properties that are measured quantitatively through an appropriate analytical quality control (QC) program. Comparability between data sets generated by different analytical methods can also be established through the use of relevant QC samples, such as standardized performance evaluation (PE) or certified reference material (CRM) samples run by both methods, in addition to other comparisons of each method's performance (sensitivity, selectivity, precision, bias, etc.).

The term "representativeness" can be used to address either the analytical aspect or the sampling aspect of sample analysis. Analytical methods must be selected and designed to be representative of the parameter of interest. Positive (i.e., causing an analytical result to be biased high) or negative (i.e., causing an analytical result to be biased low) interferences and unrecognized non-selectivity for a particular target analyte can result in a non-representative interpretation of analytical results, leading to decision errors. For example, immunoassay tests for environmental contaminants are usually designed to give results that are biased high, and the kits frequently cross-react with daughter products of the parent contaminant or other structurally similar compounds. A potential user of an immunoassay kit who does not recognize these characteristics will risk serious misinterpretation of the kit's test results. On the other hand, users who do understand this will seek to use these characteristics to their advantage, or will manage the inherent uncertainties through a demonstration of method applicability (see below) and an appropriate quality control protocol.

The representativeness of sample selection and collection is complicated by the extreme heterogeneity of many of the matrices encountered in the environmental field. The concentrations of contaminants in soils, sediments, waste streams, and other matrices can vary tremendously on even small scales in both space and time. Samples must be representative of the "true" site conditions *in the context of the decision* to be made based on those samples. If the decision is not specified, a representative sampling design cannot be selected. Sample representativeness also includes sample preservation and subsampling issues.

Comparability and representativeness are critically important to the scientifically valid interpretation of analytical data, but estimating both requires the exercise of professional judgment in BOTH the science generating the data (e.g., analytical chemistry) and in the science involved in interpreting and using the data (e.g., using the data to model contaminant extent or migration or to design a treatment system).

As noted above, there may be more than one DQI for a single data quality attribute. For example, the attribute of precision can be measured using mathematical formulas for relative percent difference (RPD), relative standard deviation (RSD), standard deviation (SD), variance (SD^2), and a variety of other calculations that can quantitatively express the degree of random fluctuation in a measurement process. The selection of a particular DQI to measure a specific data quality attribute (for example, selecting RPD to measure precision) is a matter of:

- Convention (what are people used to seeing or using);
- The characteristics of the analytical method (for example, does the method generate continuous or discontinuous data?);
- The data set being evaluated (for example, the formula for RPD cannot handle more than 2 values, whereas the formula for RSD can handle multiple values); or
- The intended use for the data (which determines how extensively the quality of a data set must be documented, and what form of documentation is most useful to the data user).

The language of "data quality attributes" and "data quality indicators" provides data generators and data users with the ability to establish the comparability of different data sets and whether data are of "known and documented quality" commensurate with intended data use.

Measurement Quality Objectives

MQOs are project-specific analytical parameters derived from project-specific DQOs. MQOs include acceptance criteria for the data quality indicators (DQIs—see above) that are important to the project, such as sensitivity (e.g., what detection or quantitation limit is desired), selectivity (i.e, what analytes are to be targeted), analytical precision, etc. **MQOs can be used to establish the "bar" for data performance parameters**. MQOs are derived by considering the level of analytical performance needed to actually achieve the project goals (as expressed in the DQOs).

However, project MQOs are **not** intended to be technology- or method-specific. As with DQOs, MQOs **specify "what"** the level of data performance should be, but **not "how"** that level of data performance will be achieved. In other words, although MQOs provide the criteria for how good the data must be, MQOs do not specify exactly how the data must be produced, and so MQOs do not specify what analytical method or technology is to be used.

In actual practice during project planning, the planning team's analytical chemist will naturally be considering which specific technologies may be applicable even in the early stages of project planning. Evaluating and refining analytical options is a significant part of the iterative nature of systematic planning which seeks the most resource-effective work strategy that can achieve the stated project goals (i.e., the project DQOs). The project chemist should explore whether available innovative analytical technologies might achieve the project MQOs (i.e., the needed data quality) to the same degree as the conventional technology, yet be able to do so in a way that is more resource-effective for the project because of lower per-sample costs, economies of scale, or more rapid turnaround times that could support real-time project decision-making.

The following are examples of what MQOs "look like":

- *An MQO for one project might read:* "The overall precision of lead measurements taken on the soil in the bins must be less than 50% RPD when at least 10 samples are taken from each bin."

- *An MQO for a different project might read:* "The measurement method to be chosen must be able to detect the presence of compounds X, Y, and Z in groundwater at a quantitation limit of 10 Fg/L with a recovery range of 80-120% and a precision of <20% RSD."

A large part of the variability in environmental data (and thus in overall decision uncertainty) stems from sampling considerations. MQOs should be developed with this fact in mind, and requirements for analytical MQOs should be derived in conjunction with the development of the sampling design. The team or individual setting the MQOs should balance the relative contributions from analytical uncertainties and from sampling uncertainties. In many environmental media (especially solid media), matrix heterogeneity causes sampling variability to overwhelm analytical variability. Insisting on perfectly precise analyses on a few tiny samples taken across a large heterogeneous matrix is meaningless since two adjacent samples will probably provide very different results. Which sample is selected, and hence the project decision influenced by that sample's results, will be a matter of chance. This "luck of the draw" can only be controlled by obtaining a better understanding of the contaminant distribution, and that is dependent on increasing the density of sample collection.

Depending on how the term is being applied and the sources of uncertainty that impact an environmental decision, *measurement* quality objectives may be interpreted to include assessment of the performance of the entire measurement system, including the uncertainties in the data introduced by sampling. This is especially true if there are more sources of uncertainty in making the actual decision than just evaluating the immediate data package. For example, making risk management decisions is based not only on site-specific data sets, but also on the non-site-specific toxicological data sets used to derive the various reference values, etc., all of which have their own associated uncertainties. However, in some usage, the term MQO is restricted to the analytical side of the measurement process, and the broader concept of DQO or decision confidence objective (DCO) is used to include the sampling considerations. This terminology may be used in activities such as permit compliance monitoring where there is no perceived "uncertainty" in the regulatory limit itself (once it has been established by the permit). In this case, the "project decision" involves demonstrating only that the permitted material is in compliance to some specified level of decision confidence. **If usage of terminology such as DQO, MQO, DCO, etc. in a particular situation is ambiguous (as many times it is), parties should strive to clarify what meaning is intended. Parties should also strive to clarify how sampling uncertainties are accounted for in data generation, assessment, and interpretation.**

Whether sampling considerations are evaluated as part of MQOs (as the entire measurement system) or as part of DQOs (or some other term expressing the overall decision uncertainty), the importance of including the sampling component when assessing overall data quality cannot be overemphasized. It is possible to isolate the performance of various parts of the measurement system, and to determine the relative contributions from the various sampling components versus the various analytical components. [Discussions about the partitioning of decision uncertainty can be found in various statistical or sampling documents available on http://cluin.org/chartext_edu.htm#stats. Since soils tend to illustrate a "worst case scenario" for non-gaseous environmental media, the following documents present valuable guiding principles: the 1990 *A Rationale for the Assessment of Errors in the Sampling of Soils*, and the

1989 *Soil Sampling Quality Assurance User's Guide*. This topic is also discussed in the paper, "Applying the Concept of Effective Data to Environmental Analyses for Contaminated Sites," EPA 542-R-01-013; available from http://cluin.org/tiopersp/].

Demonstration of Method Proficiency

A Demonstration of Method Proficiency shows that a particular operator or laboratory has the appropriate training and equipment to accurately perform a method. The demonstration may be done by using Performance Evaluation (PE) samples, or by using known concentrations of analytes spiked into a clean matrix. The purpose of a demonstration of proficiency is to ensure that the performance of the operators and equipment is capable of producing data of known quality. [Proficiency demonstrations are discussed in Chapter 2 of the SW-846 Manual, available at http://www.epa.gov/epaoswer/hazwaste/test/chap2.pdf]

Demonstration of Method Applicability

A Demonstration of Method Applicability involves a laboratory study, pilot study, field trial, or other kind of activity that establishes the appropriateness and performance capability of a particular method for a site-specific matrix and application. The purpose of a demonstration of method applicability is to ensure that a particular method or method modification can produce data of known quality, able to meet the project's decision goals, on the site- or project-specific samples to be tested.

Systematic Planning

Systematic planning for project decision-making is the process of clearly defining and articulating:

- What the goals (i.e., primary decisions) of a project will be (including how much uncertainty will be tolerated in those decisions);
- Identifying what potential sources of error and uncertainty could lead to an erroneous decision; then
- Developing strategies to manage each of the identified uncertainties and avoid decision errors; and
- Planning the most resource-effective means for implementing those strategies.

Strategies for managing uncertainties include identifying information or knowledge gaps and deciding how to fill those gaps. Locating and interpreting historical information or pre-existing data is one possible way to fill certain knowledge gaps. Collecting new data is another way to fill information gaps. Systematic planning then evaluates:

- What types and amounts of data will be needed to address the information gaps; and
- What mix of sampling and analytical technologies can address both sampling and analytical uncertainties to optimization the data collection design to maximize overall cost-effectiveness for the project.

Once decisions are made, follow-up actions (such as remedial activities) may be indicated. Systematic planning evaluates how data gathering, decision-making, and follow-up activities may be efficiently ordered or merged to minimize expensive and time-consuming remobilizations of staff and equipment back to a site. A dynamic work plan approach develops decision trees or other articulations of decision logic that guide real-time decision-making in the field to allow sequential activities to be performed in fewer mobilizations. More information about dynamic work plans can be found in *A Guideline for Dynamic Workplans and Field Analytics: The Keys to Cost-Effective Site Characterization and Cleanup*, available from http://cluin.org/download/char/dynwkpln.pdf .

The DQO process is a systematic planning approach that EPA has articulated to aid *data collection* activities. The DQO process does not address other aspects of project planning that are included under the broader term "systematic planning." Systematic planning also includes developing the work plans that will coordinate and guide site operations related to cleanup, worker safety, waste removal and disposal, public involvement and other activities needed to achieve project goals. Key to successful systematic planning is the involvement of sufficient technical expertise, generally provided through a multi-disciplinary team, that represents the scientific and engineering disciplines needed to adequately address all project issues. For example, the U.S. Army Corps of Engineers uses a systematic planning process called Technical Project Planning (TPP) that encompasses many of the project activities that extend beyond just data collection. The TPP manual can be accessed at http://www.usace.army.mil/inet/usace-docs/eng-manuals/em.htm, refer to E[ngineering] M[anual] 200-1-2.

EPA has policy requirements that mandate the use of systematic planning for all projects performed under EPA direction. EPA does not mandate *the type* of systematic planning to be done, since this necessarily will vary depending on a wide variety of factors. EPA policy statements on systematic planning can be found in *Policy and Program Requirements for the Mandatory Agency-Wide Quality System* (EPA Order 5360.1 A2), available at http://www.epa.gov/quality/qs-docs/5360-1.pdf.

Triad Approach

A strategy for cleaning up hazardous waste sites that relies on the integration of systematic planning, dynamic work plans, and real-time results (usually provided through rapid turnaround on-site measurements) to reduce costs and move site work along faster while maintaining or increasing the reliability and protectiveness of site decisions. [Discussion about the triad approach can be found in the paper, *Current Perspectives in Site Remediation and Monitoring: Using the Triad Approach to Improve the Cost-Effectiveness of Hazardous Waste Cleanups,* EPA 542-R-01-016 available from http://cluin.org/tiopersp/].

The Relationships Among Decision Goals, DQOs, MQOs, and QC Protocols

During project planning, there should a logical *conceptual* progression in the development of decision goals, DQOs, MQOs, and QC acceptance criteria. *In practice*, however, this will be a non-linear, iterative process where various options for implementing a project are explored, dissected, and recombined, the feasibility and costs for various options are estimated and weighed, and then the most promising option is selected and fully developed into project work plans that will actually be implemented. As a project's planning documents (such as work plans, sampling and analysis plans, quality assurance project plans, health and safety plans) are developed and finalized, there should be a clear presentation of (and the reasoning behind):

- The general project decision goals;

- The more detailed, technical expression of the project goals (the DQOs), and the decision rules that will guide project decision-making;

- An expression of how much uncertainty decision-makers are willing to tolerate in the final project decisions;

- An evaluation of the uncertainties (information gaps) that could potentially lead to decision errors; and

- A discussion of the strategies that will be used to manage each of those uncertainties to the degree needed to needed to accommodate the desired decision certainty.

No doubt at least one of those strategies will include the generation of analytical chemistry data from environmental samples to fill information gaps. (In contrast, it may be possible to manage uncertainty without generating data simply by "assuming the worst" and taking the most protective actions. In highly specialized instances, this might be the most cost-effective strategy when the cost of sampling and analysis to reach a "definitive conclusion," and the likelihood that action will be required anyway are both high.) When data generation is planned, the planning document should discuss:

- The roles these data are expected to play within the context of the project or how they will be used to support project decision-making;

- A description of how data will be assessed and interpreted according to the decision rules (e.g., how will the results be reduced, treated statistically, mapped, etc.);

- The goals for *overall data quality* (the overall MQOs, where "data" are measurements generated from samples and sampling uncertainties must be considered);

- How the representativeness of sampling will be ensured or assessed (how the various aspects of sampling uncertainty will be managed);

- A list of the analytical technologies and methods that were selected, and a description of the data attributes (analytes, detection/quantitation limits, requirements for accuracy as bias and precision) that is expected to be generated from the listed methods; and

- The analytical QC protocols and criteria to be used with the methods to demonstrate that analytical data of known quality are being generated that are suitable for the described intended uses.

At designated completion points in the project, project reports that summarize work accomplished to date should clearly reiterate the project goals and the means by which these goals would be achieved. Important uncertainties (that is, those information gaps that bear directly on decision-making confidence) in the decision-making process should be identified. The success of the project plan in managing those uncertainties to the degree desired, and an estimation of the overall decision uncertainty should be assessed in the project report.

In the beginning of a project, high-level program managers often set the **broad, non-technical goals for projects**: For example, "Given a budget of $X, we want to clean up this lead contaminated soil in accordance with all environmental regulations and to the satisfaction of the residents in the neighborhood." The next question, of course, is "How do we do that?" So the next step for the project manager or the planning team is to translate these broad, non-technical goals into more technically oriented goals that can address specific considerations such as:

- Regulations: What are the applicable environmental regulations? Are applicable action levels already in place in regulations, or do site-specific action levels need to be derived based on risk-drivers? If there is more than one possible regulatory action level, which one should be used?

- Confidence in the outcome: How certain do we need to be by the end of the project that we have indeed achieved goals such as risk reduction or regulatory compliance? How will we demonstrate to regulatory agencies or stakeholders that this level of certainty has in fact been achieved (i.e., what evidence will be used to argue that goals have been achieved)?

- Constraints: What are all the constraints that need to be accommodated (like seasonal weather, budget, property access, etc.)?

Making sure that no important details are left out of consideration is the purpose of a systematic planning process such as EPA's 7-step DQO process" [Detailed explanation of the DQO process as applied to hazardous waste sites can be found in the document, *Data Quality Objectives Process for Hazardous Waste Site Investigations (QA/G-4HW),* available through http://www.epa.gov/quality/qa_docs.html, and will not be duplicated here.] Statements that summarize the answers to these and other questions constitute "**the project DQOs.**" As noted earlier in this paper, the project DQOs consist of the unambiguous technical expressions of the overall project decision goals.

The next level of technical detail geared toward data collection involves translating the project DQOs into project **MQOs** [i.e., a general characterization of the kind of information (what parameters or analytes need to be measured, and what level of overall data quality for those parameters is needed) that will be needed to achieve the project DQOs]. Analytical data quality is most often only a very small part of the uncertainty that needs to be controlled in order to have sufficient confidence in the actual project decisions. An honest examination of the "weak" links contributing to overall decision certainty may reveal that paying for expensive "definitive" analyses contributes nothing toward decreasing the overall uncertainty in the project decisions when there are larger uncertainties due to the limitations of sampling very heterogeneous media.

Sampling uncertainty is decreased when sampling density is increased. Composite sampling may sometimes be used to increase sampling density while lowering analytical costs. [Refer to EPA Observational Economy Series Volume 1: Composite Sampling, EPA/QA G-5S, and other statistical documents, all available from http://cluin.org/chartext_edu.htm#stats]. Although composite sampling is undesirable in some situations and its use should be carefully considered in the context of how the data will be used, composite sampling can be a highly cost-effective and informative sampling strategy.

Another way to cost-effectively increase sampling density is by using less expensive analytical methods (perhaps, using screening methods) in association with a well-planned QA/QC design and limited traditional analyses to provide data of known quality matched to the decision needs of the project. As long as the data quality can be demonstrated to be compatible with the project's decision rules, the confidence in the overall decision reliability that is gained by increasing the sampling density will not be lost by the use of a screening method. For more details, see "Guidelines for Preparing SAPs Using Systematic Planning and PBMS" in the January/February 2001 *Environmental Testing & Analysis*. The article is available through http://cluin.org/chartext_edu.htm#planning. Additional discussion can also be found in the issue paper, *Current Perspectives in Site Remediation and Monitoring: Applying the Concept of Effective Data to Environmental Analyses for Contaminated Sites*, available at http://cluin.org/tiopersp/.

When project planners wish to express desired decision confidence objectively and rigorously in terms of statistical certainty (that may have been specified in the project DQOs), statistical expertise is required to translate that goal into strategies that blend the number of samples, the expected variability in the matrix (i.e., heterogeneity), analytical data quality (e.g., precision, quantitation limits), the expected contaminant concentrations (i.e., how close are they expected to be to regulatory limits), sampling design (e.g., grab vs. composite), and costs into an interlocking whole. Since sampling design and analytical strategy interact to influence the statistical confidence in final decisions, the interaction between an analytical chemist, a sampling expert, and a statistician is key to selecting a final strategy that can achieve project goals accurately, yet cost-effectively. Software tools can assist technical experts to develop sampling and analysis designs. [See http://cluin.org/chartext_tech.htm#imp.]

The **statistician** is concerned with managing the overall (or summed) variability (i.e., uncertainty) in the final data set, and with the interpretability of that final data set with respect to the decisions to be made. The statistician does this during project planning by addressing issues related to "sample support" (a concept that involves ensuring that the volume, shape, and orientation of extracted specimens are representative of the original matrix under investigation), by selecting a statistically valid sampling design, and by estimating how analytical variability could impact the overall variability. The **field sampling expert** is responsible for implementing the sampling design while managing contributions to the sampling variability as actual sample locations are selected and as specimens are actually collected. The sampling expert does this by selecting and using sampling tools in ways that ensure that the sample support designated in the sampling plan is met in the field. The **analytical chemist** is responsible for managing components of variability that stem from the analytical side (including aspects of sample preservation, storage, homogenization, subsampling, analyte extraction, concentration, and instrumental determinative analysis). The analytical chemist should select analytical methods that can meet the analytical variability limits estimated by the statistician, and design an analytical QC program that defensibly establishes that those goals were met in the final data set.

Managing the various sources of analytical and sampling uncertainties (assuming no clerical or data management errors) ensures that data of known quality are generated. Sometimes there may be only a single option available for a certain task, so the selection process is simple. Other times there may be more two or more options and cost/efficiency considerations can drive selection of the equipment and/or the design. It should be obvious that staff expertise (training and practical experience directly relevant to the techniques under consideration) is very important to project success.

The data characteristics that will **control analytical and sampling uncertainty** are articulated in the MQOs. Thus the MQOs specify "how good" the data must be *at a general level*. MQOs are contrasted with DQOs, which specify "how good" the *decision* must be. DQOs certainly are the ultimate drivers of how good the data must be, but DQOs themselves do not directly express data quality characteristics. Sometimes, as project planning progresses or as project implementation proceeds, it is discovered that a DQO is unattainable given the realities of the site conditions, the availability of suitable technology, and financial constraints. In collaboration with regulators and stakeholders, revision of the project DQOs may be required. For example, it may be discovered that current technology for a certain analyte is unable to provide the data needed to support risk decisions at a desired 10^{-6} cancer risk level. When a risk-based DQO is unachievable with current technology, an MQO known to be achievable with currently available technology may be substituted for the DQO. In other words, if it is clear that the ideal decision goal (the DQO) is unattainable, data quality goals (MQOs) based on the best available technology may be substituted for the ideal DQO until a time when newer technologies become available. It is important to note that the technology or method itself is NOT specified by the regulatory MQO. This allows the flexibility required for market incentives to encourage the development of technologies that can meet or exceed that same level of data quality more economically.

Although project MQOs are not meant to specify particular methods or technologies, they do serve to *guide* the selection of the technologies that can most cost-effectively meet the DQOs. As instrumentation is selected (based on factors such as the type of data needed, the turnaround time needed to support project activities, the expertise and infrastructure required to operate it, and costs), and as the analytical strategy for the project is perfected (perhaps including a "demonstration of method applicability"), analytical method SOPs and QC protocols are developed that are both method- and project-specific (i.e., tailoring an analytical method's performance to meet the specific data needs of the project). A QC protocol identifies the analytical parameter or DQI to be controlled, the limits within which results for that parameter are acceptable, and the corrective action procedures to be followed if those acceptance limits are exceeded. **QC acceptance criteria** should be very specific and should be designed such that if the QC acceptance criteria are consistently met, the project MQOs will be achieved, which means that the resulting data will be sufficient to meet the project DQOs and support the project decisions.

For example, an overall MQO for precision [for example, a statistically derived objective of less than <50% RPD between side-by-side (collocated) samples] may be partitioned into the primary components of variability that contribute to the overall variability. [Discussions about the partitioning of variability can be found in the *Rationale for the Assessment of Errors in the Sampling of Soils* document, available at http://cluin.org/chartext_edu.htm#stats .] In the QC protocol, QC samples are used to monitor and document these measures of variability. The QC acceptance criteria are used to specify the maximum allowable variation in each component, and they might be expressed something like this:

- Analytical (instrumental) precision: "XRF instrument precision shall be determined using no fewer than 7 replicate analyses of a homogenized sample with a lead concentration near 400 ppm (the action level). The resulting RSD should be less than "20%."

- Combined analytical and sample preparation precision: "Laboratory duplicates (prepared from a single sample with at least 150 ppm lead) should have RPDs less than "35%."

- Combined analytical, sample preparation, and sample collection precision: "Field duplicates (collocated samples collected from a single location with at least 150 ppm lead, with each sample collected, prepared, and analyzed separately) should have RPDs less than "50% (unless matrix heterogeneity is demonstrated to exceed the anticipated variability)."

The figure below serves to illustrate the conceptual progression that comprises the development of a design for generating data based on well-defined project goals. As stated earlier, while conceptually this process is linear, in real-life, the development of a design is highly iterative, as portrayed by the circular arrows. The figure shows that the conceptual progression starts with the project-specific decision goals, and then moves "downhill" from broader, higher level goals to narrower, more technically detailed articulations of the data quality needs. Project decisions are translated into project-specific DQOs; then into project-specific MQOs; then into the technology/method selection and development of a method-specific QC protocol that blends the QA/QC needs of the technology with the project-specific QA/QC needs of the project. Finally, data are generated.

Then the process reverses. The actual raw data must then be assessed against the project MQOs to document that the quality of the data generated do indeed meet the decision-making needs of the project. The final step in the chain is interpreting the data into meaningful information (such as a statistical expression of a contaminant concentration as an average across an exposure unit) that is fed into the decision-making process (e.g., further action is or is not needed). If the "downhill" process has been conscientiously followed, there is a very strong likelihood that the "uphill" process of data assessment and interpretation will show that the data are of known and documented quality, and are fully adequate to support the project decisions.