# AN APPLICATION OF USEPA'S DATA QUALITY OBJECTIVE PROCESS

Karen A. Storne
O'Brien & Gere Engineers, Inc., 5000 Brittonfield Parkway, Syracuse, NY 13221

**ABSTRACT**
The United States Environmental Protection Agency (USEPA) states that all collected data have error, no one can afford absolute certainty about the data, and uninformed decisions associated with data collection tend to be conservative and expensive.[1] The USEPA proposed that, before an environmental data collection project begins, criteria should be established for decision making that is defendable. To accomplish this, the USEPA developed the data quality objective, or DQO, process. This is a systematic planning tool used to establish criteria for data quality, to define tolerable error rates and to develop a data collection design. Gathering the information for the DQO process is time-consuming and may negatively impact the project budget and schedule. Therefore, a computerized worksheet that summarizes the DQO steps was developed and distributed for review by a team of consultant specialists.

Based on comments received from the consultant specialists, the limitations of the DQO process, from the consultant's aspect, were outlined. This paper presents a streamlined approach to the DQO process, involving use of a computerized worksheet to aid a project team through the DQO process. Comments pertaining to the worksheet and the DQO process, which were solicited from the consultant specialists, are described, including the limitations outlined by the consultant specialists.

**INTRODUCTION**
The Quality Assurance Management Staff (QAMS) of the USEPA developed the DQO process to improve effectiveness, efficiency, and defensibility of decisions related to environmental data collection, while minimizing expenditures by eliminating unnecessary duplication or overly precise data.[2] The DQO process is presented in the USEPA's Guidance for the Data Quality Objectives Process, EPA QA/G-4, EPA/600/R-96/055, September 1994. The DQO process results in qualitative and quantitative statements that are developed through a multi-step process that includes the following:

- Step 1. State the problem to be resolved. Identify the team members, the general problem, the project budget, the time for the study, and the social/political issues that may impact the project.
- Step 2. Identify the decision to be made. Identify the main issue to be resolved, the alternative actions that would result from each resolution, and the specific decision statement that must be resolved to address the project problem.
- Step 3. Identify the inputs to the decision. Identify the variables to be measured and the basis for the action level.
- Step 4. Define the boundaries of the study. Define the geographical area, the media of concern, the homogeneous strata, the time frame, the start and ending time periods, the scale of the decision, and the practical constraints for the project.
- Step 5. Develop a decision rule. The decision rule involves the population parameter of interest, the scale of the decision making, the action level, and the alternative action. Develop the test of the hypothesis and decision error.
- Step 6. Specify the tolerable limits on decision errors. Determine the consequences of each decision error, the quantitation limits of the error, the range of the parameter of interest, the grey region, and the acceptable probability of committing decision errors, or how much error is acceptable before the data becomes unusable.
- Step 7. Optimize the design for obtaining the data. Choose a sampling design that meets the DQO requirements and the budget.

The statements from the DQO process are summarized and presented in the Project Management Section A5 of the Quality Assurance Project Plan (QAPP).

**CONSULTANT SPECIALISTS FEEDBACK**

Since each step of the DQO process is critical in choosing a sampling design, electronic worksheets that prompt team members for responses were developed, in order to efficiently and cost effectively gather the information for each step from busy and remotely located consultant specialists. Examples of appropriate responses to each request were included in the worksheets. The electronic worksheets were distributed to a team of consultant specialists in the environmental consulting firm. The consultant specialists consisted of project managers, risk assessors, quality control officers, project officers, hydrogeologists, field samplers, and data validators. Comments, which were based on practical experience in the environmental field, were obtained from each of the consultant specialists.

In general, the initial response from the team to the DQO process worksheets was positive. The team indicated that the process of gathering project information together in a form that can be shared with the project team early on is very critical, and is not always done properly or completely. This worksheet could be used to effectively accomplish this task. Several comments received from the team requested clarification of some of the steps or that additional information be requested in the steps. Based on these responses, the worksheets were modified. However, the team had significant concerns with the DQO process, as formulated, since this process anticipates having an idealized situation for an environmental project. The process appears to be relatively in-flexible with respect to application of the process to real-life situations involving consent order schedules, information gaps, large number of target constituents, and tight budgets. The team indicated that picture-perfect projects that neatly fit the requirements of this process rarely occur.

The following sections present some of the comments received from the various consultant specialists.

**Comment**: The consultant specialists were unsure about how the worksheets would be used if little background data, such as the target compounds to be measured, is known.
**Resolution**: The worksheets may not be applicable to projects where information is not known. A statement to this effect was added to the introduction.

**Comment**: Decision errors are typically not evaluated. Rather, if sample data is questionable, the data is validated, and samples are recollected and reanalyzed. In addition, sufficient number of samples is collected to support the project decision.
**Resolution**: The goal of the USEPA is to minimize costs related to data collection by decreasing unnecessary duplication samples, and overly precise data. Utilizing the DQO process may help to decrease the number of samples collected, thereby decreasing costs. A statement to this effect was added to the introduction.

**Comment**: The consultant specialists commented that following the worksheets alone to develop a sampling design for a project could potentially leave out important issues. The worksheets attempt to put the real world in an organized box. This structured approach typically does not work in environmental projects.
**Resolution**: Environmental professionals, who can use a broad breath of knowledge, experience, and complex data to solve DQO problems, are needed to evaluate the information in the worksheets. The worksheets are to only be used as a guide to gather the information and the DEFT software is only used to evaluate the feasibility of the chosen sampling design. The professionals must choose the sampling design that meets the DQO needs based on the information gathered. A statement to this effect was added to the introduction.

**Comment**: Some of the consultant specialists may not be able to provide information requested by the DQO Steps.
**Resolution**: The worksheets would be distributed to a core group of consultant specialists, consisting of the project manager, the risk assessor(s), and the Quality Assurance Officer, for completion. After the worksheets are completed, the remaining consultant specialists would receive a copy for information purposes. Asterisks indicated the consultant specialists identified as responsible for completing the worksheet.

**Comment**: Step 1 should include the regulatory agencies and the client name.
**Resolution**: These requests were added to Step 1.

**Comment**: Less time should be spent on the alternative actions requested in Step 2B since these actions are often not relevant until a basic understanding of the site has been developed.
**Resolution**: The assumption is that the background information is available to the project team. A statement to this effect was added to the introduction.

**Comment**: There may be several variables identified in Step 3.
**Resolution**: The worksheets are intended to be used for only one constituent. Separate worksheets must be used for each constituent for a project. A statement to this effect was added to the introduction.

**Comment**: The action levels may not be defined until the risk assessment has been performed.
**Resolution**: It is assumed that the action levels are fixed such as regulatory thresholds and standards. A statement to this effect was added to Step 3.

**Comment**: The information requested in Step 3B, the basis for the action levels require prior agreement between the consultant and the client before the action levels can be presented in a QAPP.
**Resolution**: The action levels that will be used to evaluate the sample data are critical to a project, and should always be included in the QAPP. If the action levels are not established, the methods that can provide method detection limits that are appropriate for the action levels may not be chosen. In addition, the data user may compare the sample results to incorrect action levels, resulting in incorrect decisions being made and the need for resampling. The importance of the action level was noted in Step 3B.

**Comment**: The information requested in Step 4 fails to consider the complexities of sampling soil, groundwater, sediment, and surface water, potential sources, and how contaminants reside in subsurface soil. Without considering these observations, the quality of the investigation may be low.
**Resolution**: The DQO software makes assumptions that there are no temporal issues associated with the project, and that the sample locations can be randomized. The DQO process and software is to be used only to evaluate the feasibility of a sample design. Statements to this effect were added to Step 7.

**Comment**: The significance of the y-axis in the decision performance goal diagram and how the limits of tolerable probability are established are unclear.
**Resolution**: The y-axis represents the probability that a decision error will be made; deciding that the parameter of interest is on one side of the action level when the true value is on the other side of the action level. The grey area is where the consequences of a decision error are minimal. Below the action level, a decision error will result in unneeded actions and increased costs. Above the action level, error will result in human health and environmental hazard issues. The probability of decision error is set above and below the grey area to indicate the tolerable error limits. The limits of tolerable probability are established by the project team. Clarification of these issues was presented in Steps 6A and 6B.

**Comment**: The worksheets are not clear with respect to how the DQO outputs are incorporated into the sampling design.
**Resolution**: Step 7 was expanded to demonstrate how the DQO outputs were utilized to choose the sampling design.

**Comment**: The worksheets don't explain how information from the DQO process is added to the QAPP.
**Resolution**: The DQO process results in qualitative and quantitative statements summarizing the project objective, which are added to Section A5, Problem Definition and Background, of the QAPP. An example of the information added to the project QAPP was added to Section 7, and a statement to this effect was added to the introduction.

Based on the previously discussed comments, the worksheets were edited. The final version of the worksheets, with edits in bold print, is presented at the end of the paper. An example of the Decision Performance Goal Diagram from the DEFT program is presented at the end of the paper.

## CONCLUSION

The DQO process, as presented in USEPA Guidance for the Data Quality Objectives Process, EPA QA/G-4, is a good planning tool for environmental projects. Electronic worksheets that summarize the various inputs required for the DQO steps help to decrease the time required from each team member for the information gathering process.

After review by the consultant specialist team, it was determined that there are limitations associated with the DQO process. All consultant specialists agreed that the process of gathering and clarifying important project information, including the action level, that is requested by the DQO process, and having this summarized for consultant specialists before the project begins, is advantageous. However, the remaining steps of the DQO process may not be applicable to all projects. In some cases, historical background is not available, and there is an abundance of target analytes. The application of the remaining steps of the DQO process under these circumstances would lead to increased time and budget demands which would not be beneficial to the overall project.

The team also concluded that only a core team of consultant specialists would be responsible for filling out the worksheets. Also, the DQO process makes assumptions, including that there are no temporal issues associated with the project and that the sample locations can be randomized. The team also noted that the DQO process is only to be used as a guide to determine the sampling design. Environmental professionals, who can use a broad breath of knowledge and experience, are needed to evaluate the information in the worksheets. The DEFT software is only used to evaluate the feasibility of the chosen sampling design. The professionals must choose the sampling design that meets the DQO needs based on the information gathered. These limitations must be considered when implementing the DQO process.

### Data Quality Objective Worksheet 1999

This worksheet is a project-planning tool, based on the Data Quality Objective (DQO) process, presented in USEPA QA/G-4. This process is used to establish criteria for data quality and sampling designs **for each constituent at the site**, so that decisions made are reasonable, defendable, and represent a logical approach to solving the project problem**, while minimizing unnecessary duplication or overly precise data**. This worksheet should be used to organize project information and is intended to be **used in projects for which the basic site problem is known and background information is available to the project team. The DQO process results in qualitative and quantitative statements that are presented in the project QAPP. This worksheet, along with experience, should be used by professionals to establish the data quality and sampling design.**

The steps of the DQO process are presented as well as examples of appropriate responses to each request. Please fill out appropriate steps and return to K. Storne within 5 working days.

Example provided: Investigation of possible soil contamination with trichloroethene (TCE). Early sampling activities indicate that there is a low concentration area (0-50 ppm) and a high concentration area (0-80 ppm); TCE is not detected off-site; Future land use is residential; Total budget is $100,000; Remediation must take place within one year.

Step 1. State the problem to be resolved:

A. Who are the team members? (* indicates member responsible for completing worksheet)

Project Manager*_________________________Risk Assessor*___________________________

Quality Control Officer*___________________Data Validator__________________________

Data User_______________________________Laboratory Project Manager_________________

Field sampler____________________________**Client** _________________________________

**Regulatory agencies**____________________________________________________________

B. What is the general problem? (Contamination of TCE in soil. Affects human health and the environment. Low activity area is 0-50 ppm and high activity area is 0-80 ppm) _______________

C. What project budget is available? ($100,000) _______________________________________

D. What time is available? (One year for remediation) ___________________________________________

E. What social/political issues have an impact? (Future land use is residential.) __________________

Step 2. Identify the decision to be made:

A. What is the main issue to be resolved? (Does the TCE contamination pose unacceptable danger to human health or the environment?)

________________________________________________________________

B. Specify alternative actions that would result from each resolution. (ActionA – Remediate soil; Action B – Do not remediate soil) ____________________________________________________________

C. Combine main issue and the alternative actions into a specific decision statement that must be resolved to address the problem: (Determine whether or not TCE contamination in soil poses a danger that requires remediation) ________

Step 3. Identify inputs for the decision:

A. What are the variables/characteristics to be measured? (TCE)

____________________________________

B. What is the basis for the action **level (regulatory threshold or standard), that must be established and included in the QAPP before sample collection?** (Risk assessor/toxicologist set site-specific exposure assessment at 50 ppm) _______________

Step 4. Define the boundaries of the investigation:

A. What are the spacial boundaries?

1. What is the geographical area? (property boundary; none detected off site) ___________________
2. What is the media of concern? (TCE in surface soil to depth of 15cm) __________________
3. What are the homogeneous strata? (Area of high concentration to 80 ppm, area of low concentration to 50 ppm)

B. What are the temporal boundaries?

1. What is the time frame? (Results represent future conditions at sites) ______________________
2. When will the investigation start and end? (Starts in 1 month and ends in 1 year) _____________

C. What is the scale of decision to be made? (For each residential lot-sized acre.) _________________

D. What are the practical constraints on data collection? (Existing structures exist) _________________

Step 5A. Develop a decision rule or if/then statement that includes:

1. The population parameter of interest (do not consider sample depth) (average mean)
2. The scale of the decision making (resident lot size)
3. The action level (50 ppm)
4. The alternative action (remediate / do not remediate) (If the true mean TCE concentration in the residential lot is greater than 50 ppm, the soil is remediated. If not, the soil will be left in place)

Step 5B. Develop a test of hypothesis and decision error:

1. If the assumption is that the site is clean:
   Null Hypothesis: *Site is clean; true mean level <50 ppm*
   Alternative Hypothesis: *Site is not clean; true mean level >50 ppm*

- False positive (F+) Type 1 Error: *Decide that the site is not clean when it is which results in action when none was required, which is an overreaction to a situation, wasted resources, unnecessary expenditure and cleanup.*
- False negative (F-), Type II Error: *Decide the site is clean when it is not which results in no action when some was required, which is a missed opportunity for correction, allows a hazard to public health or environment.*

Step 6A. Specify limits on decision error; how much error is acceptable:

1. Determine consequences of each decision error; how sensitive is each decision? (health/ecological/political/social/resource risk)
2. Set quantitation limits of false positive/negative error (0-20ppm, 20-35ppm, 35-50ppm, 50-60ppm, 60-100ppm, 100-200ppm, 200-250ppm)
3. Determine range of parameter of interest; should fall within range of possible concentration (0-250 ppm)
4. Specify grey region (see table - *), **where consequence of decision/error are minor**; grey area is bounded by:
   A. the action level (50 ppm)
   B. The value **where the consequences of making decision begins to be significant** (60 ppm)

Step 6B. Develop the "what/if " table:

1. Specify limits on probability of committing decision errors**. (For 0.3 tolerable probability, at 30% of the time a wrong decision will be tolerated)**; (50ppm- 30%, 35ppm – 20%, 20ppm – 10%, 60ppm – 30%, 100ppm – 20%, 200ppm – 10%)

| What/If Table | | | | | |
|---|---|---|---|---|---|
| Reported TCE Concentration | Decision Made | True Concentration# | Error Type | Aversion | **Tolerable Probability |
| >50ppm | Cleanup | 0-20ppm | F(+) | Severe **(Cost high)** | 10% |
| >50ppm | Cleanup | 20-35ppm | F(+) | Moderate | 20% |
| >50ppm | Cleanup | 35-50ppm | F(+) | Minor | 30% |
| <50ppm | No action | 50-60ppm | F(-) | Minor | *Grey Region |
| <50ppm | No action | 60-100ppm | F(-) | Moderate | 30% |
| <50ppm | No action | 100-200ppm | F(-) | Severe | 20% |
| <50ppm | No action | 200-250ppm | F(-) | Very Severe **(Risk to human health and environment)** | 10% |
| Note: Null hypothesis - the site is clean | | | | | |

This completes the question section of the worksheet. The information gathered and professional experience is then used to generate the sampling design. This process is described below.

Step 7. Based on the DQO outputs and historical information develop a sampling design. The sampling design must be cost-effective and balance sample size with method performance and decision error tolerance. For example, it may be more cost effective to use less expensive and less precise methods in cases of high variability in samples exist, so that a large number of samples can be taken and so that the sample design error can be controlled. If less variability in samples exists, more expensive and precise methods can be used to collection fewer samples to control the measurement error.

The USEPA DEFT software is used only as a guide to develop the sampling design alternatives. DEFT does not account for the difference between media, spacial or temporal boundaries. Inputs for the DEFT program include:

A. Parameter of interest; assumption is that the population mean is used (mean)
B. Limits on probability of committing decision errors (50ppm- 30%, 35ppm – 20%, 20ppm – 10%, 60ppm – 30%, 100ppm – 20%, 200ppm – 10%)
C. Action level (50ppm)
D. Possible range of parameter (250ppm)
E. Unit cost of sample collection and analysis per sample ($30, $220)
F. Location and width of grey region; range of possible parameter values where consequences of F(-) error are minor; bounded by action level and parameter value where consequences of F(-) begin to become significant (50-60ppm)
G. Estimated standard deviation (default is used; max concentration – minimum concentration /6)
H. Null hypothesis; which error is F(+) and which is F(-) (Site is clean)

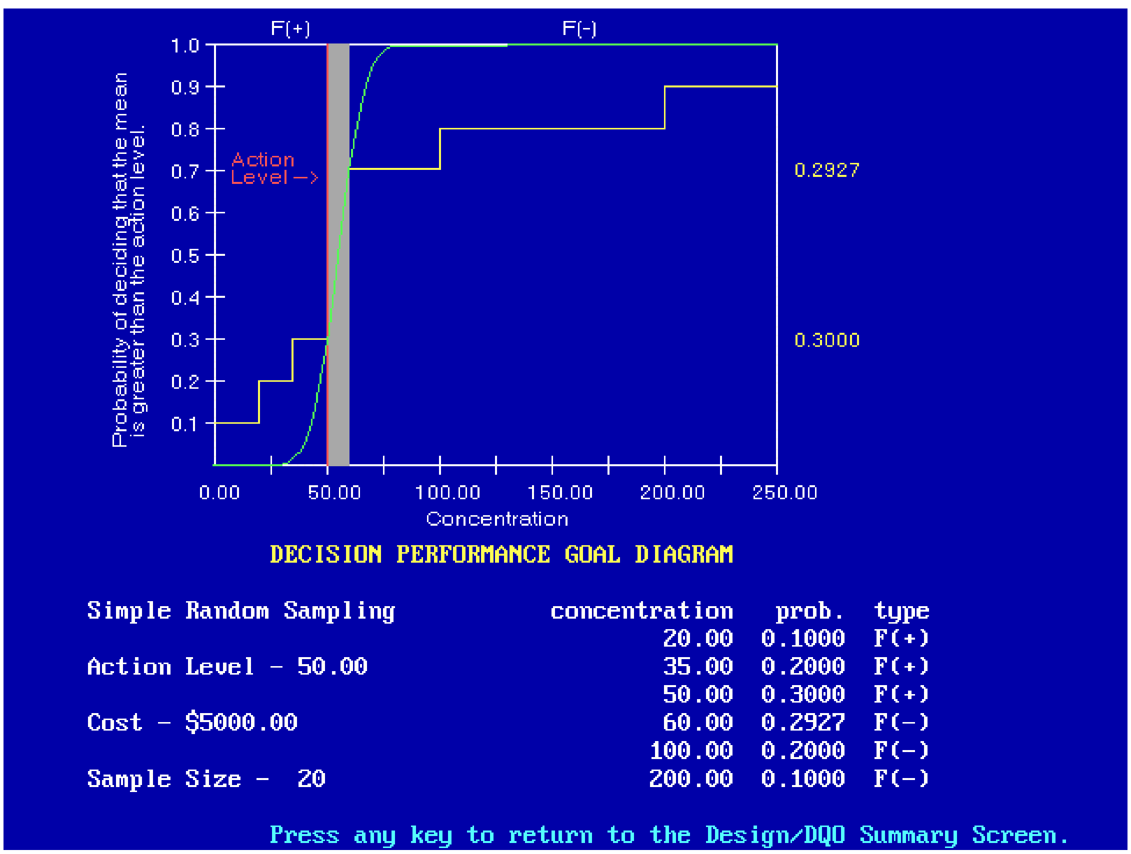
Three basic sampling designs available in the DEFT program include:
A. Simple random; Many samples are taken and total costs are high. Every possible point at the site has equal chance of being sampled. Simple random is used when variability is small and field and analytical costs are low to detect peak concentrations.
B. Composite; Multiple samples are collected and combined; subsamples are collected for additional analysis. Composite is used when the average concentration and sampling of a large number of sample sites at a reduced cost is desired.
C. Stratified random; The site is divided into two or more subsets. Each subset is sampled separately with one of the designs previously described. Stratified random is used to improve the precision of the design.

The previously listed inputs, the initial sample design, and the sample size are entered into the DEFT program, and a performance goal diagram is drawn. Altering the inputs, the design, or the sampling size may change the decision performance goal. The performance of the design is evaluated by the performance curve, which is based on the graph of the power function, and which is overlaid onto the performance goal diagram. The design that produces a very steep performance curve is preferred over one that is flatter. The power function is the probability that the null hypothesis is rejected when the null hypothesis is false. Ideally, the power function would be zero if the null hypothesis were true and one if the null hypothesis were false. Due to imperfect data, it is not possible to achieve the ideal power function. However, the power function will yield values that are small when the null hypothesis is true and large when the null hypothesis is false.

If the design fails to meet the DQOs, increase the budget, increase the width of the grey area, or increase the tolerable decision error rates.

The statements resulting from the DQO process are presented in the Project Management, Section A5, of the QAPP. (A simple random sample design should be used to compare concentrations of samples collected for TCE analysis from the site to the action level of 50 ppm. 20 samples shall be collected from each sample location. Each sample location will be generated randomly.)

The Decision Performance Goal Diagram for the example provided in the DQO worksheets



**Footnotes**

1. USEPA 1997. *Introduction to Data Quality Objectives,* Quality Assurance Division, Washington D.C., page 4.
2. USEPA 1994*.* *Guidance for the Data Quality Objectives Process*, EPA QA/G-4, Washington D.C., page 1.

**References**


USEPA. 1994a. *Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT)*, Version 4.0, EPA QA/G-4D, Washington, D.C.

USEPA. 1994b. *Guidance for the Data Quality Objectives Process*, EPA QA/G-4, EPA/600/R-96/055, Washington, D.C.

USEPA. 1997a. *Introduction to the Data Quality Objectives, Quality Assurance Division*, Washington, D.C.

USEPA. 1997b. *Overview of EPA Quality System Requirements for Environmental Programs*, Quality Assurance Division, Washington, D.C.