

DIVISION OF ENVIRONMENTAL RESPONSE AND REVITALIZATION

DATA QUALITY OBJECTIVES PROCESS SUMMARY

DERR-00-DI-32

Internal Guidance Document: January 2002

DERR DQO PROCESS SUMMARY

Table of Contents

Step Page

Foreward..... ii

 Introduction..... 1

1. Step 1: State the Problem..... 4

2. Step 2: Identify the Decision..... 5

3. Step 3: Identifying the Inputs to the Decision..... 6

4. Step 4: Define the Boundaries of the Study..... 8

5. Step 5: Develop a Decision Rule..... 10

6. Step 6: Specify Tolerable Limits on Decision Errors..... 13

7. Step 7: Optimizing the Design..... 16

The Division of Environmental Response and Revitalization (DERR) has developed a summary of the US EPA Guidance for the Data Quality Objectives Process (EPA QA/G-4) (EPA 1994). This interpretation of the DQO Process is meant to serve as a tool for DERR site coordinators and project-specific DERR staff (support staff). By following the prescribed steps of the DQO Process, DERR staff (and other project personnel) will gain insight and assistance into addressing, formulating and documenting the type and quality of data needed for sound environmental decisions. A description of the methods for collecting and assessing those data are also a part of the DQO Process described herein. The development, review and successful implementation of the DQO Process is a mandatory part of the DERR planning process and the DERR Quality Assurance/Quality Control (QA/QC) Program. This process will help DERR develop, plan and implement projects which will ensure that data generated and used in decision-making are of the type and quality needed for their intended purpose.

This document provides a summarized version of the DQO Process and includes the specific requirements necessary for appropriate project planning within DERR and for organizations that conduct environmental data collections operations on behalf of the Ohio EPA/DERR.

This document is one of the Ohio EPA/DERR QA/QC documents which describe DERR policies and procedures for planning, implementing, and assessing the effectiveness of DERR's quality system.

Background

With the creation and implementation of the Division of Environmental Response and Revitalization (DERR) Quality Management Plan (QMP) (October 2000), the DERR is now required to follow the Data Quality Objective Process (DQO) as it has been established by the US EPA document, ***Guidance for the Data Quality Objective Process (EPA QA/G-4)***. This change to the project planning process is one component in the overall Quality Assurance/Quality Control Program within the division. DERR engages annually in a number of data collection activities. This data is used for environmental decision making and for support of project-specific activities and actions. The need to collect data of known quality and quantity assists in the overall confidence of the data being collected. For this confidence to exist, and for the data to be used in support of decisions, there is therefore a need for an established, structured process for this data collection activity.

II What are Data Quality Objectives (DQO)?:

DQO's are Qualitative and Quantitative criteria for clarifying project objectives, defining the appropriate types of data needed, and defining the tolerable levels of potential decision errors for the project. It is a systematic planning process to generate environmental data appropriate and sufficient for its intended use. The process is designed to answer four basic questions:

- 1). What data is needed?
- 2). Why is it needed?
- 3). How will the data be used?
- 4). What tolerance does the user have for decision errors?

The DQO process is a method to ensure that the collection and analysis of data for a project meets the requirements for the specific project goal. The conceptual site model is an integral part of the DQO process and can provide direction to define tolerable limits and data needs.

III Data Quality Objective Process

The structured process for the planning of all data collection activities within DERR will be the DQO Process. The DQO Process is DERR's recommended planning process for environmental decision making. DQOs provide consideration for a number of issues during the planning stages of a project.

Some of these items/issues include:

- Early involvement of the decision maker(s)
- A graded approach to data quality requirements
- Effective sampling and analysis programs
- A basis for judging the usability of the collected data

The DQO Process is essentially a systematic planning process for generating environmental data that will be sufficient for their intended use.

From consideration of the qualitative and quantitative statements and the general questions indicated above, formation of specific underlying principles for environmental data collection can be realized through implementation of the DQO Process. These principles (or items for consideration) include, but are not necessarily limited to:

1. Collected data may have the potential for error.
2. Absolute certainty comes with a high price
3. The DQO Process defines tolerable error rates
4. Without DQOs, the quality of decisions are unknown
5. The DQO Process is based on the scientific method

The DQO Process involves a seven step process that is considered a “dynamic” process through the life-cycle of a project. The term “dynamic” is used because it affords project personnel the ability to re-visit the planned goals and objectives for a project and make necessary adjustments where and when needed for successful completion of the project.

Before the seven step DQO Process is started, several issues/items must be taken into account which help in preparation for consideration of each step of the DQO Process. These issues/items include getting the right people involved in the right way with the project. Who are the stakeholders, the decision maker, the technical experts, and the personnel with appropriate statistical training? Other items proven to be useful when implementing the DQO Process include gathering existing site data/knowledge, giving consideration to the overall project objectives, and being realistic about resource and sociopolitical constraints.

The seven steps of the DQO Process are as follows:

1. State the Problem
2. Identify the Decision
3. Identify Inputs to the Decision
4. Define Boundaries
5. Develop a Decision Rule
6. Specify Limits on Decision Errors
7. Optimize the Data Collection Design

Data Quality Objective Process - DERR Summarized Overview

The following pages in this document include a summary of the DQO Process used by DERR staff for environmental data collection activities. The material and/or language have been summarized to provide a usable, general understanding of this process for all DERR staff. The key points for each step of the DQO Process are put into general terms for ease of understanding. This summary will help facilitate appropriate implementation of the process and will further afford DERR with the collection of data that is of sufficient quality and quantity to support DERR decision making.

STEP 1 - STATE THE PROBLEM

- I. Establish the planning team including the decision makers. This should include the project manager, technical staff, data users and stakeholders.
- II. Describe the problem and develop a conceptual model of the environmental hazard to be investigated. Describe the conditions that may pose a threat to human health or the environment or circumstances of potential noncompliance with regulations and the reasons for undertaking the study. Conduct literature and ongoing/past study searches to help define the problem. Development of the conceptual model at this time is critical.

The conceptual model is often a diagram that shows:

- A. known or expected locations of contaminants,
- B. potential sources of contaminants,
- C. media that are contaminated or may become contaminated, and
- D. exposure scenarios (location of human health or ecological receptors).

- III. Identify available resources, constraints and deadlines. Limitations on resources and time constraints for collecting data need to be identified. Available personnel, contracts (if applicable) and intermediate and time-frames for collecting data should be identified.

Outputs of this step are as follows:

- 1). A list of the planning team members and their roles.
- 2). Identification of decision makers.
- 3). A concise description of the problem and a conceptual model of the environmental problem to be investigated.
- 4). A summary of available resources and relevant deadlines for the study including budget, availability of personnel and schedule.

Reference: US EPA "Guidance for the Data Quality Objective Process" (EPA QA/G-4 August 2000) Chapter 1 - pages 1-4 through 1-6.

STEP 2 - IDENTIFY THE DECISION

The purpose of this step in the DQO process is to develop a “Decision Statement” that identifies both the purpose of your study and the potential actions that may be taken when the data is analyzed. If the principal study question is not obvious then the study may fall in the category of exploratory research, in which case this particular step may not be needed.

- I. Identify the principal study question (p. 2-2) A. What is the unknown condition(s) or unresolved issue(s) being studied? or What do you know vs. what you don't know?
Ex. Does the concentration of contaminants in groundwater exceed acceptable levels?
- II. What are the alternative actions that may be taken to solve the problem? (p. 2-2) A. The purpose of this question is to develop a list of possible remedial actions that may be taken depending on the answer to the study question.
Ex. 1. Design and implement a ground water cleanup action 2. Hook up residents to City water supply 3. No action
- III. Combine the principal study question and the alternative actions into a decision statement. This statement is the output of this DQO step. (p. 2-3) A. The following template may be helpful: Determine whether or not [unknown environmental conditions/issues/criteria from the principal study questions] require (or support) [taking alternative actions].
Ex. Determine whether or not the concentration of contaminants in groundwater exceed acceptable levels and if they do design and implement a groundwater cleanup action, and/or hook up residents on private wells to the City water supply.
- IV. If multiple decisions are to be resolved during the study, you should examine how the decisions relate to one another and then prioritize each separate decision step in order of importance and the sequence in which they should be resolved. (p. 2-3)

Output of this step are as follows: 1). A Decision Statement that links the principal study question to possible actions that will solve the problem.

Reference: US EPA “Guidance for the Data Quality Objectives Process” (EPA QA/G-4 August 2000) Chapter 2 - pages 2-1 to 2-4.

STEP 3 - IDENTIFYING THE INPUTS TO THE DECISION

The following represents the basic elements required to identify inputs to the decision for the Data Quality Objectives Process:

- I. Identify to the kinds of information needed
- II. Identify the sources of information
- III. Determine the basis for setting the Action Level
- IV. Confirm the appropriateness of proposed sampling and analyses methods

I. Identify to the Kinds of Information Needed

Determining the kinds of information needed can be accomplished by answering the following questions:

- A. is information on the physical properties of the media required;
- B. is information on the chemical or radiological characteristics of the matrix needed;
- C. can existing data be used to make the decision; and
- D. are new measurements of environmental characteristics needed.

If it is decided that new measurements are needed, a list of characteristics required to be measured to make the decision should be developed. If the decision can be based on existing data, the sources of data should be examined to ensure acceptability. In the event consideration is given to integrating new data with existing data, parameters with the existing data need to be examined so that new samples are collected and/or analyzed in a similar manner.

II. Identify the Sources of Information

It is necessary to identify and document the sources of information needed to resolve the decision. The sources of information may include results of previous data collections, historical records, regulatory guidance, professional judgment, scientific literature, or new data collections.

III. Determine the Basis for Setting the Action Level

Action Levels are concentrations of contaminants that are based on regulatory requirements, risk assessments, performance criteria for analytical methodology, or a reference standard. It is important to understand how the Action Level will be derived or rather what information will be used to determine the Action Level.

Establishing an actual numerical value for the Action Level is not required at this stage; however, a potential Action Level should be established.

It is important to note that if the Action Level (used here as an example) is based on a regulatory requirement, the numerical value of the Action Level will be known at this stage in the process. If the Action Level is based on a risk assessment or other performance criterion, specification of the numerical value of the Action Level should be deferred until after the study boundaries have been determined.

IV. Confirm the Appropriateness of Proposed Sampling and Analyses Methods

A list of sampling and analytical methods, derived from evaluating pertinent environmental characteristics, should be developed that are appropriate for the scope of the problem being investigated. It is imperative to specify sampling considerations required for detecting analytes at low concentrations and procedures required to collect these sample quantities as well as identifying analytical methods that have appropriate detection limits. Importance should be given to the problem of minimizing bias as this is an important performance characteristic of sampling and analysis. Methods known to exhibit large biases should be avoided if possible. Additional considerations include requirements for certification of personnel and laboratory accreditation. Laboratories analyzing environmental samples should follow standard protocols and procedures or use performance-based methods.

Outputs of this step are as follows:

- 1). A list of environmental characteristics that will be measured to enable decision making.
- 2). A list of information sources or methods/assumptions that indicate how each Action Level will be derived
- 3). A list of information that may be applicable to uses of the data in future investigations
- 4). Confirmation that sampling and analytical methods exist to meet the detection limit criteria required given the appropriate magnitude of the Action Level 5.

Reference: US EPA "Guidance for the Data Quality Objectives Process" (EPA QA/G-4, August 2000) Chapter 3 - pages 3-1 through 3-5.

STEP 4 - DEFINE THE BOUNDARIES OF THE STUDY

I. Define the target population of interest.

1 The target population is the set of all environmental samples about which the decision maker wants to draw a conclusion and include the various media to be sampled.

2 In environmental studies the target population is often the set of all possible samples from various media that taken together constitute the geographic area of interest.

II. Determine spatial boundaries that define the physical area to be studied and generally where samples will be collected.

1 Define the geographic area applicable for decision making. A metropolitan city limit, soil within a property (to a depth of six inches), or a length of shoreline are examples of defined geographic areas.

2 Divide the population into subsets that have relatively homogeneous characteristics. Dividing the target population into subpopulations that are relatively homogeneous within each area or subunit and appropriate sampling design (Step 7) can reduce the total number of samples required to meet tolerable limits on decision errors.

III. Determine temporal boundaries that describe the time frame that the study will represent and when samples should be collected.

1 Determine when to collect data. Consider time-related phenomena such as weather, seasons, etc., and how they may affect data and data collection.

2 Determine the time frame for decision making. Evaluate the population and determine the optimum time frame for collecting data, given that the medium may change over time.

IV. Determine the practical constraints on collecting data.

1. Practical constraints could include site access, availability of equipment or staff, and conditions which would prevent sampling such as flooding, freezing temperatures, etc.

V. Define the scale of the decision making.

— Determine the smallest subpopulation, area, volume, or time for which separate decisions must be made (e.g., operable units, species age, gender, etc.).

— Decisions may be based on risk, technological considerations, temporal considerations, financial scale, or others.

Outputs of this step are as follows:

- 1). Detailed descriptions of the characteristics that define the population to be sampled
- 2). Detailed descriptions of geographic limits (spatial boundaries) that are appropriate for the data collection and decision making.
- 3). Time frame for collecting data and making the decision.
- 4). List of practical constraints that may interfere with data collection.
- 5). Appropriate scale for decision making.

Reference: US EPA "Guidance for the Data Quality Objective Process" (EPA QA/G-4, August 2000) Chapter 4 - pages 4-1 through 4-9.

STEP 5 - DEVELOP A DECISION RULE

I. Background:

A. Steps 1-5 of the DQO Process are primarily focused on identifying qualitative criteria including:

- 1 The nature of the problem that initiated the study and a conceptual model of the environmental hazard to be investigated.
- 2 The decisions that need to be made and the order of priority for resolving them.
- 3 The type of data needed (i.e., geographic area, environmental medium, overall timing of data collection, etc.)
- 4 A decision rule that defines how the data will be used to choose among alternative actions. (EPA QA/G-4 0-8)

B. Step 5, assume that all information determined in the previous steps (1 - 4 outputs) are "perfect":

- 1 The site problems have been correctly identified
- 2 the conceptual model is appropriate for the site
- 3 the proper decisions to be made have been identified
- 4 the appropriate sources of information were identified
- 5 the appropriate potential Action Levels were identified
- 6 the appropriate possible measurement methods were identified
- 7 the spatial/temporal boundaries of the decision have been decided.

C. Step 5 Activities

- 1 define appropriate statistical parameter(s) (e.g., mean, median, etc.)
- 2 specify Action Level(s) (sets boundaries for outcomes of the decision process)
- 3 develop logic for action (measurement and analysis methods -sufficiently sensitive (and appropriate to media, etc.) so measurement method detection limit is below Action Level
- 4 combine true value of the population parameter selected and the Action Level with scale of decision making (Step 4) and alternative actions (Step 2).

D. Outputs from the 4 previous steps are integrated with the Step 5 inputs.

1. A theoretical if...then decision rule is decided. (This rule determines when to choose possible alternative actions).

II. Activities (Referenced above in background information):

A. Define the population parameter

Select a population parameter ...that summarizes the critical characteristic or feature of the population (e.g., see "b" and "c" below) that will be compared to the Action Level to make a decision.

1. Mean - (e.g., average of all sediment sample concentrations within the boundaries of the site - central tendency)

- a. Compares only the middle of the population (group of all sample sets) to the Action Level - extreme values below detection limits will dilute the population mean ([Cd]) and a few higher values will elevate it.
- b. Unless the sample set concentrations are all close to the same value the mean is not a very useful tool, particularly if there are "hot spots" (skewed or outlier data).
- c. Good for chronic carcinogenic exposures concentrations for comparison to Action Levels
- d. Allows an estimate of the total amount of a contaminant in defined area of soil or water body

2. Median - (e.g., the center most [Cd] value between all sediment samples within the boundaries of the site)

- a. More useful than the mean for skewed (nonsymmetrical, or not normally distributed) data (not influenced by extreme values in sample distribution)
- b. Also preferred if many sample values are below detection limits (must have no more than 50% sample values below detection limits - or a true median cannot exist)

3. Percentile - (percent of samples below given value (e.g., Action Level))

- a. For use in more acutely toxic contaminants.
- b. Useful when many of the samples analyzed are below detection limits.
- c. Can require larger sample sizes than the previous two parameters.

4. Total amount - (All samples used)

NOTE: The Decision rule and data collection design complexity is dependent upon the chosen parameter complexity. (The above parameters are ranked starting with least to most complex)

B. Define Action Level

1. Types of Action Levels

a. Predetermined

1). Fixed standards (e.g. drinking water or technology based)

b. DQO Process determined

1). Background or specific risk-based standards (e.g., conservative or real concern thresholds)

NOTE: more conservative standards can require more stringent detection limits (i.e., more sensitive analytical methods - usually more expensive)

C. Detection Limits

1. Detection limits for measurement methods identified in Step 3 must be below the Action Level. If not, then a more sensitive method or different approach will be necessary.
2. Detection limits are defined for a specific purpose. For quantitative identification purposes they will be more stringent, ruling out most probabilities of false positive results in the matrix while increasing a high probability of an accurate positive identification.
3. When comparing a mean concentration to a threshold action level, detection limits must be determined to define the reliability of quantitation.

Outputs of this step are as follows:

A. Construct the theoretical "if....then.." decision rule

- 1). Combine selected population parameter and Action Level with scale of decision making (Step 4) and the alternative actions (Step 2)

(e.g., If the true mean dioxin concentration in the surficial 2 inches of soil of a decision unit (20' X 100') exceeds 1 ppb, then remove a 6 in layer of soil. If the true mean is not greater than 1 ppb, then do nothing.)

Reference: US EPA "Guidance for the Data Quality Objectives Process"
QA/G-4 August 2000) Chapter 5 - pages 5-1 to 5-5.

(EPA

STEP 6 - SPECIFY TOLERABLE LIMITS ON DECISION ERRORS

In this step of the DQO process, the user no longer imagines that perfect information and unlimited data will be available for making decisions. The purpose of Step 6 is to specify quantitative performance goals for choosing between the alternative actions in order to determine the probabilities of making errors in the decision. To do this, a user should conduct the following activities:

- I. Determine the sources of potential errors in the data set
- II. Establish a plausible range of values for the chemicals of concern
- III. Define the types of potential decision errors and the consequences of making the errors:
 - Determine how to manage the errors
 - Select the baseline condition that will be assumed to be true in any event
- VI. Specify a range of possible parameter values (gray area)
- VII. Assign probability values at several true value points above and below the Action Level that reflect your tolerable probability for the occurrence of decision errors.

The Activities of Step 6:

I. Sources of Error:

A decision error occurs when the sample data set misleads you into making the wrong decision (wrong response action). Sample data is subject to various random and systematic errors throughout the collection process. The combination of all errors is called the "total study error". Typically, there are two types of study errors:

A). Sample Design Error: This occurs when the data collection design does not fully capture the variability in the decision unit. This error leads to random errors, such as variability and imprecision, and systematic errors, such as bias, in estimating a chemical of concern.

B). Measurement Error: This is an error of variability in sample collection, handling, preparation, analysis, data reduction, transmission, and storage. Both random and systematic errors in measurement occur because of imperfections in the measurement and analysis system selected for a project.

II. Establish a Plausible Range of Values:

The user must review the current and historical data and approximate the upper and lower boundaries of the chemical of interest using best professional judgment. An example would be to use the mean of the chemical of interest as the relevant value and the highest and lowest concentration of contamination as the upper and lower boundaries as the range.

III. Defining Decision Errors:

When uncertainty in the data does not allow the user to determine whether the true value is above or below the Action Level, it becomes necessary to assign a de facto decision when there is not enough evidence to refute the decision. An example is “innocent until proven guilty”. When the user examines the data set and makes incorrect decisions based upon a faulty data set, the errors can be either False Acceptance errors (false positives) or False Rejection errors (false negatives).

IV. Managing Decision Errors:

Although the possibility of making decision errors cannot be totally eliminated, a user must learn to manage the errors when making decisions on response actions. The user must determine the final goal of the data and what type of errors would make the biggest impact on the use of this data. If sampling design errors are believed to be the biggest concern, a larger sample size or development of better sampling techniques may be required. If the measurement errors are suspect, multiple analysis of the sample or use of more precise and accurate analytical methods may be required. However, a balance of the consequences against the cost of limiting these errors must also be evaluated. If moderate errors in sample design or measurement would not compromise a project, a user may consider the “screening data” obtained satisfactory for the preliminary stages of the project and reconfirm the decision later based upon confirmation samples.

V. Selection of the Baseline Condition:

The baseline condition may already be selected for the user as established by regulatory considerations. If there is not an established baseline, then one would be defined based upon the relative consequences of decision errors, such as a judgment between the threat to public health and the spending of unnecessary resources. Please refer to Section 3 and Section 5 outline pages for this selection and determination. The baseline condition is the Action Level. The user should evaluate the potential consequences of decision errors at several points within the false positive and false negative ranges. In the DQO Process, the baseline condition is the Action Level determined in Step 5.

VI. Specify Gray Area Values:

The gray area is a region where it is not feasible to control the false position errors at low levels due to the high cost of sampling and analysis and the potential of selecting the wrong course of action is low. The first boundary of the gray area is usually the Action Level.

Generally, the gray area is narrower when significant expense is afforded to sampling and analysis in order to reduce uncertainty. It is recommended that the gray area be widened during the early phases of the study process, but narrowed at latter stages to determine if the chemicals of concern are only slightly different than the Action Level.

VII. Assign Probability Values Above and Below the Action Level to Reflect Error Tolerance:

The user should specify a limitation for false positives and false negatives in the data set with regard to the Action Level based upon the consequences expected from these errors. The rationale for setting these limits should include consideration of regulatory guidelines, potential cost impacts; human health, ecological conditions, and socio-political consequences. USEPA guidance recommends the most stringent limits typically encountered: 1% for both false positives and false negatives. Other guidance can be found in EPA policies and programs if a user has concerns which justify an alternate tolerance.

Two software programs which may be helpful during this step are "**Data Quality Assessment Statistical Toolbox**" (**DataQUEST**), which is very helpful for the statistical-side of the process, and "**Decision Error Feasibility Trials**" (**DEFT**), which can be utilized for the final steps of the DQO Process and helps the scoping team especially in the final step #7 when optimizing your design. These will need to be evaluated based upon the project complexity.

Outputs of this step are as follows:

- 1). Baseline Condition = Regulatory or Risk Limit or Background
- 2). Gray Area = Range of Acceptable Values
- 3). Decision Error Limits = Percent of False Positives and Negatives Allowable.

Reference: US EPA "Guidance for the Data Quality Objectives Process" (EPA QA/G-4 - August 2000) Chapter 6 - pages 6-1 through 6-15.

I. Purpose:

The purpose of this final step is to achieve optimization of the Sampling and Analysis Design. The technical and/or scoping team needs to be re-assembled for this final purpose of the DQO Process. Most typically, there will be more than one design option for the technical/scoping team to review. The end determined decision will incorporate the type and number of samples deemed necessary to characterize a site or an area of the site.

II. Activities:

The scoping team is expected to review each design option and balance the site assessment sampling with such resources as funding, personnel, and temporal constraints while meeting the DQOs established in the previous steps. The technical/scoping team should review the outcomes of the previous DQO Process steps to determine exactly how the limits on decision errors will prescribe number and location of samples to be collected and the types of analyses per sample. From information of the previous DQO steps, the team can then confirm the budget for sampling and analysis and the project schedule.

Once the final design is selected, the team leader and/or project coordinator must ensure that the chosen design is properly documented.

Outputs of this step are as follows:

- 1). The most resource-effective design is expected to achieve the DQOs.

Reference: US EPA, "Guidance for the Data Quality Objectives Process" (EPA QA/G-4 August 2000), Chapter 7 - pages 37 through 40.