Data Quality Objectives

The Data Quality Objectives (DQOs) Process is used to systematic plan for collecting environmental data of a known quality and quantity to support decisions. This guidance describes a seven-step process that is used for developing performance and acceptance criteria that will be used to establish a collection design for a project. The U.S. Environmental Protection Agency (EPA) has written a document that provides guidance on the subject. The document is "Guidance on Systematic Planning Using the Data Quality Objectives Process", EPA QA/G-4, EPA/240/B-06/001, February 2006 (http://www.epa.gov/quality/qs-docs/g4-final.pdf). Questions or comments about the Data Quality Objectives Process may be addressed to any Quality Assurance (QA) Coordinator listed on the MPCA Quality System webpage (http://www.pca.state.mn.us/index.php/about-mpca/mpca-overview/agency-strategy/mpca-quality-system.html?expandable=1).

DQOs are qualitative and quantitative statements derived from the outputs of the first six steps of the DQO process. These statements will clarify the study objectives, define the most appropriate type of data to collect, determine the appropriate conditions from which to collect the data, and specify tolerable limits on decision errors which will be used as the basis for establishing the quantity and quality of data needed to support the decision.

The seven steps used to the DQO Process are used to generate the optimal design for sampling and analytical work. Each step must be used to complete the process, but depending on the site, may or may not be in as much depth. These steps will be clearly described in the Site Quality Assurance Project Plan (QAPP) or Site Work Plan. This document must be approved prior to any work being done on the site.

State of the problem

The purpose is to clearly define the problem so the focus will be unambiguous. The following steps need to be completed to begin the process.

1. Identify the members of the planning team. This will be driven by the size and complexity of the problem. There should be representation from all stakeholders. Potential individuals include; samplers, chemists, engineers, modelers, hydrogeolist, etc.
2. Identify the decision maker. The decision maker is the leader of the team. This person has ultimate authority for making final decisions.
3. Give a concise description of the problem. Describe the conditions and circumstances that are causing the problem. Discuss how they affect human health or the environment, and potential non-compliance with regulation. Describe how the problem is currently understood. If the problem is complex, consider breaking it into pieces.
4. Specify the available resources and relevant deadlines for the study. Discuss budget, personnel, vehicles, and any other resources that may be used.

Identify the decision

The purpose of this step is to define the decision statement that drives work being done on a site. A decision statement links the principal study question to possible actions that will solve the problem. The principal study question is what is the reason for doing the work? The principle reason for doing the study is to answer a question, such as, “is pollution from the site contributing to groundwater degradation?”After finding the principal study question the different actions that could be taken to answer the question need to be written down such as, “buried solvent barrels on site are leaking and causing the groundwater degradation.”

1. The last step is to combine the principal study question with the possible actions to form a decision statement. An example from above would be, “determine if solvent barrels buried on site are causing the degradation of the groundwater”. This is a simplified example, but shows the logic involved in developing the decision statement.
2. When several separate decisions are needed, list them, and define what the sequence is in which they must be solved for the site. A flow chart is a possible method of tracking the levels of decisions to be made.

Identify the inputs to the decision

Identify the inputs that are needed to resolve the decision statement and determine which inputs rquire environmental measurements.

1. Determine the variables or other information that are needed. Answer questions on whether chemical or physical properties need be studied to answer the decision statement.
2. Determine the source of the information needed (e.g., historical records, previous data, regulatory guidance, etc.).
3. Identify the information needed to establish the action level. This could be driven by risk assessment of regulatory thresholds or limits.
4. Confirm the appropriate analytical methods exist to provide the necessary data. Consider the chemicals of concern and limits that are required to find methods that would be appropriate for measurement.

Define the boundaries of the study

Define the spatial and temporal boundaries that are covered by the decision statement.

1. Give a detailed description of the boundaries of the problem. This would be physical and temporal (the timeframe when the data are taken).
2. When appropriate, divide the population into strata that have homogeneous characteristics.
3. Determine when to collect data. Weather, humidity, time of year, etc., can affect the data.
4. Define the scale of decision making. This deals with focusing your sampling and analysis in an area that is most affected by your study. For example, the top six inches of soil in children’s play area.
5. Identify any practical constraints on data collection. This includes weather, consent to sample on someone’s property, equipment, time, and personnel.

Develop a decision rule

Define the parameter of interest, specify the action level and integrate previous DQO outputs into a single statement that describes a logical basis for choosing among alternative actions.

1. Specify the statistical parameter that characterizes the parameter of interest. This means what is to be measured must be defined. Is the mean of the concentrations what drives the cleanup? This may be defined by a regulatory guidance.
2. Specify the action level for the study. Choose what level would cause a different action. For example, if the soil has <1 mg/kg of a contaminant, the soil is left on site, vs. >1 mg/kg the soil goes to a landfill vs. >50 mg/kg the soil goes to a hazardous waste landfill. Confirm that all action levels can be met by the laboratory analytical methods.
3. Develop a decision rule. This takes into account the prior steps in the following example. "If the soil (average value from sampling a grid) within the site (defined by the boundaries) is less than the action level (e.g., 1 mg/kg) then take action A (leave the soil in place) else action B (remove the soil).

Specify tolerable limits on decision errors

This step allows the tolerance to be set up for decision errors. This is based upon the consideration of the consequences of making an incorrect decision. This step involves setting the limits on a false positive or false negative and realizing the impacts that these errors could cause. There will also be a gray region that is defined as where the consequences of decision errors are relatively minor. Part of the gray region also deals with the error built into any method, either analytical or sampling, that is used. A planner can go as far in depth as they wish in this step to determine what level of data and number of samples they need to feel confident, they have a site picture that is "good enough". It is recommended that at a minimum this step should generate the level of analytical data needed, the number of samples needed to have a statistically valid set of data, and the type of sampling needed (e.g., how precise).

Optimize the design for obtaining data

1. Review the outputs from the prior six steps and determine if they are consistent with the requirements of the site.
2. Develop the general data collection design alternatives. These could be simple random sampling, stratified random sampling, systematic sampling, etc.
3. Select the optimal sample size that satisfies the DQOs for each data collection design. This includes a cost function of number of samples to total cost of sampling and analysis. If the cost (or other constraint) compared to the needed data do not match, then one of the constraints will need to be relaxed. These include increasing the budget, increasing the tolerable error rates, easing the schedule changing the boundaries, etc.
4. Select the most resource effective data collection design that satisfies all of the DQOs. Choose the option that gives the best balance of costs and ability to meet project requirements.
5. Document the details, theory, and assumptions behind choosing the option to be used in the sampling and analysis plan and in the QAPP. An experienced statistician can review the results of the work and make suggestions that could reduce the sample load or costs in other areas.

The DQP process is iterative and may take several iterations to determine the best course to take. Each time the evaluation of the potential methods of site work or laboratory analysis will get more detailed.