Section 3:

Data quality

This Section will show you how to:

- **Plan so your data can be used by others, based on your specific purpose.**
- **Interface with primary data users to establish your data collection strategies.**

This Section may also be of value to you if you want to enhance an education program by teaching the importance of quality assurance and quality control (e.g., learning the value of duplicate field samples, teaching scientific processes or understanding variability of results).

Collecting credible data

Assuring data credibility is the primary challenge you may face if you want your data to be used by others. It's also a primary challenge to show how to do this in one guide, because there are as many different approaches for quality assurance and quality control (QA/QC) as there are different types of monitoring. Keep in mind that the level of data quality you need is relative to *your* purpose and the uses of *your* data. Data used for one purpose is not "higher quality" than for another purpose; you should select a level of data quality that is appropriate for *your* particular purpose.

This Section is about building QA/QC into your project, or how to ensure the data you collect is usable. If you are spending time and resources to make the effort to collect data, you want to be sure you don't compromise the results by not following basic accepted procedures. If you expect decisions to be made based on the data you collect, the data will need to meet criteria accepted by the ultimate users. Remember, collecting data is time sensitive. In other words, if you make a mistake, you can't go back and correct it, as conditions will never be the same at any

MPCA Guidelines for 305b and 303d Assessments

See *Appendix D* for a copy of the Minnesota Pollution Control Agency's Monitoring Guidelines to meet the Clean Water Act's 305b (use-support assessments) and 303d (list of impaired waters) requirements. This document is a compilation of information from various resources at the MPCA.

other time. If you don't do it right the first time, the data may not be usable for your purposes.

It is important to note, too, that the data you collect will undergo greater scrutiny as the use moves from awareness to regulation and also with the number of people and institutions affected.

Communicating with data users

The best way to ensure you will collect usable data is to *check with the primary data user who will use your information*. We cannot emphasize this too strongly. You can waste time and resources putting together a water monitoring project, only to discover that you did not use appropriate methods or equipment that will make your data usable. Some users may not require rigorous data, but the level of rigor needed rests with the ultimate user.

It is important to note that in this context "data user" refers to the *primary* user of your data, whom you identify up-front and consult while developing your monitoring plan. Once you finish your monitoring effort and the data is public, there may be many other groups and individuals who wish to use your data. For these "secondary" users, it's up to them to decide if your monitoring purpose and QA/QC practices meet their needs. It would be impossible to plan for all the potential uses of your data. What you *can* do is identify up-front who *you* want to use your data and then consult with that primary data user to ensure the data you collect meet their needs.

General QA/QC concepts

In this guide, we will discuss the concepts of building QA/QC into any volunteer water monitoring project and the general parameters that scientists look for when setting up QA/QC objectives. If you set data quality objectives and/or develop a Quality Assurance Project Plan *before you begin monitoring*, you can help ensure all your data is usable for its intended purpose. Building QA/QC into your project, up front, will put you on the right track from the beginning.

Your project may include some or all of these parameters. Again, *talk to your primary data user* to see which ones are appropriate for your project. And remember, too, when you are establishing QA/QC objectives, there are many professionals available to help you. If you are working with an organization, for example, it is likely that the group has QA/QC objectives already established.

Help your data user

Your data users may not be sure of the monitoring protocols and QA/QC procedures they need to be able to use your data. If that is the case, the following are some things to consider that will help you and your primary data users determine acceptable protocols:

- If the primary data users are not sure about data quality needs and QA/QC protocol, try phrasing the question differently and ask what their data quality concerns might be for the parameters you are considering. Then use this Section and the examples to identify QA/QC protocols that address those concerns. You can then present suggested QA/QC protocols to the users to assess their comfort level.

- If the data uses are potentially controversial or involve resource management decisions with significant financial implications, you want to have especially high confidence in your data. In this case, the protocols and QA/QC procedures in *Appendix D* may be a good model. In general, these protocols have been reviewed and recognized by scientists.

- Consider your audience or the people who must accept the credibility of the data. In general, people will be more likely to accept results that come from

accepted methods or protocols. In other words, do some research and find out the generally accepted scientific methods for sampling the parameters you are interested in, and then reference the source of your methods. *Section 4* of this guide provides some references for specific methods and sampling design considerations for Minnesota.

- Consider the variability of parameters you are monitoring. For example, bacteria counts in streams can vary widely and bacteria sampling can be easily contaminated. So you will probably want to have some QA/QC samples, such as field or sampler blanks that help determine whether or not accuracy has been compromised by contamination.

- You might choose to use QA/QC sampling to assess laboratory accuracy and precision with field kits as well as for use with a contract lab. It is always good practice to run standards and duplicates when using field kits. You can complete duplicates for assessing the precision of physical parameters such as temperature or stream flow. Taking duplicate Secchi disk readings only takes a few minutes.

- Consider the questions you might get regarding the data you are collecting. Then use this Section and the examples to identify QA/QC protocols that address those questions.

- When in doubt, reach for the highest level of quality you can and build into your program all the QA/QC protocols you can afford. Err on the side of more/better data, using the highest level of QA/QC you can.

Another option is to look for existing volunteer monitoring efforts that are tackling questions similar to the one(s) you hope to address, and ask participants about the procedures they follow and who uses their data. If you can bring an example to your potential primary data user of how similar data has been gathered and used elsewhere in Minnesota, you may be able to build a level of understanding and confidence that will allow you to work through data quality questions.

There are many examples of local individuals and organizations using volunteer monitoring data for a variety of purposes (see examples in *Appendix H*). No one magic formula will ensure your data will be used for local decision-making. However, by clearly identifying your monitoring purpose, talking through data quality questions with your intended primary data users and sharing examples from other parts of the state, you will be well on your way to assuring yourself and your primary data users that the data you generate will be usable for the intended purpose.

Quality assurance/quality control

Quality assurance refers to the overall management system, including the organization, planning, data collection, quality control, documentation, evaluation and reporting activities.

Quality control refers to the routine technical activities that help you minimize errors. Together, establishing QA/QC helps you produce data of known quality, enhances the credibility of your monitoring activities and ultimately saves time and money. To ensure quality data, both sample collection and laboratory analysis have QA/QC responsibilities.

You must collect samples according to the needs of primary data users and the Standard Operating Procedures (SOPs) you have selected, being aware of:

- sample containers (sizes and materials)
- preservation
- sample holding times
- sampling methods
- documenting methods and materials used
- sample handling before and after use to eliminate contamination

The lab must also follow the analytical SOPs and assure that:

- it is using proper analytical procedures
- it is documenting calibration procedures/results, analytical results and lab QA/QC analyses
- its instruments are calibrated according to manufacturers' direction and tested with known standards; calibrations should be recorded on lab sheets

The primary data user has the final responsibility of determining validity based on the monitoring program and analytical QA/QC procedures.

Setting data quality objectives

There are two basic ways to establish data quality objectives: 1) from your primary data users; and/or 2) from experimentation. Keep in mind that if you fail to meet your objectives, you can learn and improve, change your methods or change your data use goals. Five major parameters are typically used to measure the quality of your monitoring results and to use in building your data quality objectives.

- **Precision** – How closely repeated measurements of the same characteristic agree. You determine precision by calculating the difference between samples taken from the same place at the same time. Minimizing human error plays an important part in assuring precision.

- **Accuracy** – How close your results are to a true or expected value. You determine accuracy by comparing your analysis of a known standard or reference sample to its actual value.

- **Representativeness** – How closely samples represent the true environmental condition or population at the time a sample was collected.

- **Completeness** – Whether you collect enough valid, or usable, data (compare what you originally planned to collect with how much you actually collected). For example, if 100 samples were to be collected, but only 90 were actually collected, then 90% completeness is documented.

- **Comparability** – How data compares between sample locations or periods of time within a project, or between volunteers.

Precision is usually assessed with field and/or laboratory duplicate samples. Field duplicates are made by collecting two or more samples from the same place at the same time. This simply means you collect a duplicate sample in the exact same manner as the first sam-

ple (using the normal sampling equipment, cleaning procedures, etc.). Each duplicate is analyzed and the results theoretically should agree. Results not in reasonable agreement suggest a quality problem in the field. Laboratory duplicates consist of running analyses twice from one particular sample. Results not in reasonable agreement for laboratory duplicates suggest a quality problem in the laboratory.

How many duplicate samples do you have to collect to ensure you meet the precision parameter? *It is typically 5% to 10% of the samples collected.*

Here's how precision enters into whether your data is credible: you typically calculate the relative percent difference (RPD) (a calculation based on the percent difference of the samples) between the samples. The smaller the RPD, the more precise your measurements are. Based on the data quality objective set for the parameter you are measuring, a decision will be made about whether the data is usable or not.

Accuracy reflects how close your results are to a true or expected value. For the purposes of volunteer water monitoring, you will use procedures to determine whether or not your equipment is giving accurate results, or if contaminants are being introduced in the sampling and analysis process that may bias results and provide less than accurate results.

Accuracy in water chemistry monitoring.
QA/QC sample analyses often include blanks and spikes, as follows:

- **Sampler blanks (analyzing a blank sample with a zero value)** A sampler blank (sometimes called rinsate blank or equipment blank) is a sample of distilled or deionized water that is rinsed through the sampling device and collected for analysis. Results will determine if equipment was properly rinsed or decontaminated from one site to the next and if equipment was properly handled in the field.

Calculating relative percent difference

Data quality objectives for precision are typically expressed as the relative percent difference (RPD). Relative percent difference is calculated using the following equation:

RPD = (Result 1 – Result 2)/((Result 1 + Result 2)/2) x 100

EXAMPLE:
On May 9, 2002 the Prior Lake-Spring Lake project staff collected a field duplicate at site CD-1 on County Ditch 13, which was analyzed for Total Phosphorus (TP) with the following results:

Duplicate 1 = 0.271 mg/L TP
Duplicate 2 = 0.276 mg/L TP

RPD = (0.271 – 0.276)/((0.271 +0.276)/2) x 100 = 1.8%

This meets the field precision objective set by the project of ± 30%.

If significant concentrations of the water quality parameter being measured are found in sampler blanks, it could suggest that field equipment is not being properly cleaned between sites. In this case, you will need to determine whether to change/improve field procedures, and whether or not the problem could have affected results of other samples collected that day.

- **Field blanks** Field blanks are "clean" samples produced in the field. They are used to test for problems with contamination from the time of sample collection through analysis at the laboratory. A field blank is created by filling a clean sample container with distilled or deionized water in the field using the same procedures used to collect the site water samples. When the field blank is analyzed, it should be at least a factor of 5 below all sample results (i.e., little of the substance being analyzed should be found in the field blank sample).

- **Spiked samples (also known as matrix spikes)** One way to assess accuracy of water chemistry samples in the laboratory is to add a known concentration of the parameter to a portion of the sample to get a "spiked sample." The difference between the original measurement of the parameter in the sample and the measurement of the spiked sample should equal (or be close to) the added amount. The difference indicates your ability to obtain an accurate measurement.

- **Method blanks** A method blank consists of deionized water that is run through the normal analytical method. The method blanks should be clean water and the water quality parameters being assessed should not be detected above the reporting limits. If the water quality parameter being analyzed for is detected in this "clean" water sample, it may suggest that the analytical equipment is not accurate since it did not read the true value.

Using field blanks

In 1999 and 2000, citizen volunteers from the Vermillion River Watch Council worked with state and local agencies to monitor fecal coliform bacteria levels in the Vermillion River, Dakota County, as part of a Total Maximum Daily Load (TMDL) Study. Agencies provided training, sampling protocols, clean buckets and distilled water for rinsing.

Volunteers collected weekly samples from various sites, along with occasional field blanks. Samples were kept on ice and immediately delivered to a central location where a contract laboratory picked them up for timely analysis. By analyzing the field blanks against the samples taken from the site, they were able to assess potential contamination from the sampling method, shipping and laboratory process. Bacteria were not found in any of the field blanks, which increased the confidence that the accuracy of the measurements on the river water samples was not compromised by bacteria contamination from other sources.

**Partnering with the Vermillion River Watch Council for this project were: Dakota County, Dakota Soil and Water Conservation District, Dakota County Environmental Education Project and the Minnesota Pollution Control Agency.*

Accuracy in biomonitoring.
For biological (plant and animal) monitoring, accuracy is commonly assessed during sample processing and identification.

Processing: Typically, samples are processed in a laboratory, after they have been collected and preserved. In the lab, organisms are removed from the excess sediment or vegetation that was collected during sampling. Usually a lab technician will use a microscope to sort through samples, but most volunteer monitors pick through samples with the naked eye. To ensure that the final group of identified organisms accurately reflects the sample, an independent person should check the matrix of sorted material to ensure all organisms were found. Ideally, you will find 95 percent of the target organisms.

Identification: Usually all volunteers' invertebrate identifications must be verified by an expert. Typically, an expert verifies entire samples, but as the volunteers' skills increase, they can assemble a "voucher collection" to use as a primary means for verification. A voucher collection is a collection of invertebrates, all verified by an expert, that is preserved for use a "true value" to which taxonomic comparisons can be made. Even with the use of a voucher collection, there will always be difficult organisms that must be checked by an expert.

Repeat sample: To ensure that the individual or individuals responsible for collecting the field sample are doing so properly and consistently, two samples should be taken at a minimum of 10 percent of all sites sampled. The second sample can be collected concurrently with the first sample, or within a relatively short time from the collection of the first sample (i.e., one to three weeks). Wetland and stream samples can generally be collected concurrently, but care must be taken to collect the second sample in an area that was not disturbed while taking the initial sample. If concurrent sampling is not possible, take

Matrix spike calculations

Percent recovery for matrix spikes is calculated with the following equation: **% recovery = (C1 – C2) /C3) x100**
C1 = Concentration of spiked sample **C2** = concentration of unspiked sample **C3** = Concentration of spike added

Assessment of laboratory accuracy for the Prior Lake–Spring Lake Improvement Project

The contract laboratory used for this project included the following results in their laboratory report for May 9, 2002 samples. Review shows that these results meet data quality objectives, since concentrations were not detected in the method blanks and the matrix spike percent recovery results were within the guidelines of 90 to 110 percent.

Analyte/Parameter	*Method Blank Results*	*Matrix Spike Results*
Ortho Phosphate as P	<0.006 mg/L	98% recovery
Phosphate as P, Total	<0.010 mg/L	99% recovery

spring samples at a close interval, as this is a time when the invertebrate community can change rapidly in a short time frame (i.e., one week). Fall samples can be spaced up to three weeks apart.

Representativeness. A number of factors may affect the extent to which measurements actually represent the true environmental condition or population at the time a sample was collected. For example, data collected from a backwater area of a stream may not be representative of the primary flow in the stream. Making sure the data you collect is representative of the water body is typically addressed with sampling program design (see *Section 4: Designing Your Monitoring Program*).

Completeness is a measure of the number of samples you originally determined you would need, compared to how many you actually collected. For example, if your monitoring purpose is *problem investigation* with the intent to provide data to the MPCA for assessing the impairment status of a lake, you need to meet MPCA's data needs. That means, if you were assessing the narrative eutrophication standard, you would need to collect 12 total phosphorus samples, 12 chlorophyll-*a* samples and 12 Secchi disk measurements. If, at the end of your project, you had collected only 10 measurements of each parameter, it would mean you did not meet your data quality objective for completeness. Since there are many reasons why samples are not collected as planned, a general rule of thumb is to plan to collect more samples than you actually need.

Comparability is the extent to which data can be compared between sample locations or periods of time within a project, or between projects. This is a useful data quality check that essentially asks how your data compares with data that others have found for the same site or for similar conditions. It is good practice when reporting your data to include comparisons with other data.

Other data quality considerations

Although incorporating the above parameters will help ensure credible data, you will also need to do the following: follow instructions; provide documentation; inspect, maintain and calibrate equipment; and manage data.

Following instructions. It's easier to follow instructions that are developed using clear Standard Operating Procedures (SOPs) (the detailed procedures for the methods you will use). You should develop SOPs for your project before you go to the field. Many SOPs are already available for sampling and analytical procedures. Section 4 of this guide references a number of existing methods manuals, which include SOPs.

"Comparability" in action

For a quality check, the Metropolitan Council, as part of its Citizen-Assisted Monitoring Program (CAMP), routinely has a professional limnologist on its staff collect samples from the same lakes at approximately the same date that volunteers are monitoring. This professionally collected quality check is compared with CAMP volunteer collected data. Data generated by the CAMP program has been accepted by the MPCA and used as part of its impaired waters assessments.

Reporting laboratory QA/QC results

Data quality objectives are typically established for both field and laboratory efforts.If you decide to use a contract laboratory, we suggest making the reporting and assessment of laboratory QA/QC parameters a required part of the laboratory report. Guidance for retaining laboratory services is included in *Appendix B*.

Documentation. It is important to use and completely fill out data sheets. The same holds true for sample bottle labels, lab sheets (if applicable) and sample drop-off sheets (e.g., chain of custody).

Inspecting, maintaining and calibrating equipment. Keep field and laboratory equipment in good working condition. You should regularly inspect equipment and perform maintenance as suggested by the manufacturer. You should calibrate equipment before each use according to manufacturers' directions and test with known standards.Record all calibrations on lab or field sheets.If equipment is used to collect analytical samples, decontaminate the equipment between sample collections and analyses.

Data management. The subject of managing data is covered in detail in *Section 5*. As you collect data, it is a good idea to check it against your data quality objectives throughout the project, so if corrective actions are necessary they can be made before the end of the project. Try to identify a QA/QC project manager who can review the data and compare it with the data quality objectives. No data should be entered into a database before the QA/QC manager approves it. If data does not meet the data quality objectives set for your project, a decision needs to be made regarding its use and if it should be flagged when it is entered into a database.

Using data quality parameters in the field

The Dakota County Wetland Health Evaluation Project (WHEP) demonstrated the use of data quality parameters in a project to sample plant and invertebrate (true bugs, beetles and crustaceans) communities in the county's wetlands.

In the project, adult citizen volunteers worked under the direction of local teachers or nature center staff. In 2001, 10 teams (representing 10 cities) sampled 41 wetlands. To implement the program, they held three training sessions for the citizen monitoring teams. At least one experienced person on each team served as the team leader. The teams relied on spot checks to ensure they were adhering to data quality parameters.

- Each city evaluated one wetland in another city, as a means of providing a duplicate analysis and assessing whether repeated measurements agree (i.e., are precise).
- A technical expert spot-checked 10% of the wetlands sampled to assess accuracy, representativeness and completeness.

The expert reviewed the vegetation sample plot already evaluated by the citizen team to check if it was *representative* of the wetland and the vegetation was *accurately* identified. The expert also reviewed the insects collected by the team to check for *accuracy* of identification and to ensure they completely filled out the data collection sheets.

Taking the next step: developing a Quality Assurance Project Plan (QAPP)

A QAPP is a written document that outlines the procedures you would use to ensure that the samples you collect and analyze, the data you store and manage and the reports you write are of high enough quality to meet the desired data uses. A QAPP is a plan required for all USEPA- funded monitoring efforts.

A QAPP is very thorough and detailed, with elements prescribed and formatted to meet the needs of reviewers and provide some standardization across the county. A QAPP has the following elements:

1. Title and Approval Page
2. Table of Contents
3. Distribution List
4. Project/Task Organization
5. Problem Identification/Background
6. Project/Task Description
7. Data Quality Objectives for Measurement Data
8. Training Requirements/Certification
9. Documentation and Records
10. Sampling Process Design
11. Sampling Methods Requirements
12. Sample Handling and Custody Requirements
13. Analytical Methods Requirements
14. Quality Control Requirements
15. Instrument/Equipment Testing, Inspection, and Maintenance Requirements
16. Instrument Calibration and Frequency
17. Inspection/Acceptance Requirements for Supplies
18. Data Acquisition Requirements
19. Data Management
20. Assessment and Response Actions
21. Reports
22. Data Review, Validation, and Verification Requirements
23. Validation and Verification Methods
24. Reconciliation with Data Quality Objectives

A QAPP can be extremely valuable to you and the data users to ensure that the data collected is of a certain confidence and meets the objectives of the project. You can use the QAPP to make sure you are following proper procedures and collecting data that meet the project objectives and will be credible to decision-makers.

The ability to reference a QAPP and show how it was followed can also help you answer questions from other groups concerned about the reliability of your data. However, QAPPs are not necessary in every situation, and it does take some time to put one together. Unless you are required to do a QAPP, you may want to start with a monitoring plan (see *Section 4*). And, once you have completed a study design, it's easier to move up to a QAPP, as most of the elements required by a QAPP will be a part of your study design.

For more on QAPPs

For additional information on Quality Assurance Project Plans, see *The Volunteer Monitor's Guide To Quality Assurance Project Plans* by the USEPA, Doc. number EPA 841-B-96-003

http://www.epa.gov/owow/monitoring/volunteer/qappcovr.htm

and *The Massachusetts Volunteer Monitor's Guidebook to Quality Assurance Project Plans*, Doc. number DWM-CN61.0