# Section 6

# *Data Management*



## This section will show you how to:

- Create an effective system for storing and managing your data.

## Managing data

By now you know that volunteer surface water monitoring requires attention to detail and precision at every stage. The next step is to set up an effective plan for managing your data *before* you begin collecting it. Setting up a data management system before you actually begin monitoring will help to:

- Eliminate/minimize errors in recording and transferring data.
- Prevent loss of data.
- Provide easy access and use of the data.
- Format data in a way that will be useful and acceptable to others.

For some independent groups, storing data in loose-leaf binders may be an acceptable form of management. However, most likely a computer will be used to store and access data in either a spreadsheet or database. *Always talk to your data user to find out how they would like the data submitted to them*. Decide in advance how data sheets will be handled, and how and where they will be stored and then archived.

Computerized data management offers many advantages, especially if you will be collecting data from many sites over a long period of time. With the right computer program:

- A great amount of information can be stored in a small amount of space.
- Data can be viewed in many different ways (e.g., by site, by sampling date, in order from highest to lowest value), allowing you to detect relationships between data points that otherwise could not easily be seen.
- Conversions, calculations (e.g., percent saturation from dissolved oxygen and temperature data), and statistics can be automatically computed in a fraction of the time required for manual calculation.
- Data can be automatically graphed, arranged in a chart or table, or in other ways illustrated for purposes of analysis or presentation.
- Large amounts of data can be readily shared with other users.

### Data Transfer and Data Entry Concerns

Keep the number of times data is transcribed to a minimum. The more times you transcribe the data (from one sheet to another), the more chances there are for errors to occur.

To minimize errors in data entry (from data sheets into a computerized database), always have another person check your work. Be aware of these common data entry errors:

- Entering data in the wrong units (e.g., entering concentrations in milligrams/liter instead of micrograms/liter).
- Reversing numbers.
- Misplacing decimal points.
- Entering in the wrong row or column.

## Database or spreadsheet?

The best way to set up an effective data management system is to ask yourself, "What do we need this system to do?" Think about what information will be put into the database and what needs to be done with it. What types of output will you need? Reports? Graphs? It may be helpful to make a list of all the tasks you want the system to accomplish. Talking with your data user and other volunteer groups to find out what programs they have used and what, if any, problems they have encountered can also be helpful.

The main difference between a spreadsheet and a database is in the functions that they can perform. Each system has its own particular strengths and weaknesses. In general, spreadsheets (e.g., Lotus 1-2-3, Excel, Quattro Pro) can perform a wider range of mathematical and statistical calculations, and can be used to create a variety of useful graphs. Database programs (e.g., Lotus Approach, FileMaker Pro, Access) can store a larger volume of data and are capable of finding or "retrieving" data more efficiently.

More sophisticated database programs are "relational," meaning that you can extract information from two or more separate database files at the same time. As a simple example, suppose you have two separate files, one with volunteer names and addresses, and one with site locations plus the name of the volunteer who monitors the site. A relational database can "link" the two files via their common field (volunteer name), allowing you to produce a report that combines data from both files.

Another advantage of databases is that they can be easily programmed to catch errors in data entry. Fields can be programmed to accept a certain style or range of values- say a pH is entered as 80 instead of 8.0, the database would recognize the error and reject the entry. Databases are also better equipped to handle different types of data, including nonnumeric information. In a database, you are able to create a field to enter comments and observations, whereas spreadsheets primarily accept and analyze quantitative data.

Most groups will find that they need both types of software to optimally manage their data. For example, if your program has many sites, lots of parameters and includes biological (nonnumeric) data, a database will probably be the easiest system to store and manage your data. The database program automatically generates data tables and summary statistics, which can then be copied into a spreadsheet program to generate graphs.

Anytime you are entering data, check it and check it again. Then have another individual

review it yet again. Make certain that you have a record of individuals who check the results and also record the dates they were reviewed.

Be wary of releasing an electronic form of data to users before the numbers are checked and rechecked or before all data are entered. To prevent multiple versions of a database from being circulated:

- Wait to release the results until all data have been entered and checked (i.e., resist the temptation to release draft databases).
- Include a field for dates and initials of the last update or approval so that you can easily tell if you are working with the most current version. Also, to avoid losing your data, make a backup copy and store it in another location.

Using field sheets and lab reports

One of the most important elements to consider in designing a data management system is who will be using the system (inputting the data). If primarily just one person will use the system, it can be designed for speed and power. If the system will be used by a large number of volunteers, you may want to compromise speed and power for ease of use.

Creating a spreadsheet or a database that is set up to look like the field or lab sheets that volunteers have filled out can maximize ease of use. The person entering data should not have to spend time hunting for information in the field sheets or across the computer screen for someplace to input data. An easy way to do this is to use the same headings and field names in the same order in your database as they appear on the field sheets.

## Sharing data

Most groups and organizations store and manage their own data. However, you may want to make your data accessible to other groups, agencies and the public. Many groups are now posting their data or results of their monitoring online. Some examples include:

- The Vermont Monitoring Cooperative. **http://vmc.snr.uvm.edu**
- The Burlington Eco-Info Project. **www.uvm.edu/~empact**
- The White River Partnership. **www.whiteriverpartnership.org**
- Friends of the Mad River. **www.friendsofthemadriver.org**
- Poultney-Mettowee Watershed Partnership. **http://vacd.org/pmnrcd**
- The Shelburne Bay Watershed Programs. **www.shelburnebay.org/home**
- Friends of the Winooski River. **www.winooskiriver.org**
- Huntington Conservation Commission. **http://www.gmavt.net/~aaronw/ecoli/2004_home.htm**

Another option is to submit your data to a state, regional or national database. This usually re-

Send Complete Information: Metadata

Note that it is generally not sufficient to simply send in the monitoring data for inclusion in a database. You must also send "metadata," or information about the data. This should include your reasons for sampling (your "what" question or monitoring purpose), how samples were collected (including where and when), types of gear and containers used, as well as information on quality assurance. Quality assurance information may include a QA plan and who approved it, training level and certification of volunteers, and results of QC samples. Providing this information can extend the uses of your monitoring data, as it will assure credibility and understanding of why and how data were collected to other interested parties now and in the future.

quires registration of your program and communication with the sponsoring organization (or database manager) ahead of time. Submitting data to one of these databases increases the chances that federal, state, and local agencies and organizations are aware of, have access to, and use the data you have collected.

Some databases to consider for housing your data are:

- EPA's STORET Database.
  **www.epa.gov/STORET**
- America's Clean Water Foundation World Water Monitoring Day Database.
  **www.worldwatermonitoringday.org**
- The Great North American Secchi Dip-In.
  **http://dipin.kent.edu**
- Global Rivers Environmental Education Network (GREEN).
  **www.green.org**
- UVM Watershed Alliance Vermont Water Quality Gateway (school group data).
  **www.vtwatergateway.org**

## Helpful hints

Based on your specific purpose, use any or all of the following guidelines to set up your data management system:

- Develop or use data collection sheets and checklists that have already been developed. This will ensure uniform collection and recording of results in both the field and the laboratory. If you are working with an organization, you will probably use sheets and checklists provided by the organization's program.
- Continually review the data sheets and checklists as you collect data to ensure the information is complete. To do this, you may have a signature line for a sampling team captain or third party to indicate the data sheet was checked or approved. If there are any problems, this reviewer will contact any sampler whose field sheets contain significant errors or omissions.
- If using a laboratory try to use a Chain of Custody form (or transmittal letter) to document the transmittal of samples. Most laboratories will supply these.
- If using a laboratory, the laboratory manager should review the QA/QC parameters used and include the results with the laboratory report.
- Review field and laboratory QA/QC results and determine if data quality goals (set in *Section 4*) have been met. Make a decision whether to keep the data or not. Many times, even though the data do not meet QA/QC goals for a particular purpose, they will meet goals for another purpose and may still be usable. In such cases, datapoints may be "flagged" in a spreadsheet or database to indicate how they did not meet their original QA/QC goals.
- If you are using a contract laboratory, you can require the lab to provide results in an electronic format that is compatible with your data management system and/or your data user's requirements.
- Always have a second person review the data entered.

Some information in this Section was taken with permission from the article "Designing a Data Management System," by Fred Lease: *The Volunteer Monitor*, Volume 7, No. 1, Spring 1995.

Now that you have finished reading *Section 6*, return to the Worksheet on pages 5-8 to answer the corresponding questions.