
6.1 Managing Volunteer Data

The following steps will help ensure that the data collected by volunteers are well managed, credible, and of value to potential data users.

Review Field Data Sheets

The volunteer program coordinator or designated analyst should screen and review the field data sheets as they are received. This involves some basic “reality checks.” Questions that should be kept in mind include the following:

- *Are the results as might be anticipated, or are they highly unexpected? If unexpected, are they still within the realm of possibility?*

For example, can the kit or technique the volunteer used actually produce results like that? Does the volunteer offer any possible explanations for the results (e.g., a sewage treatment plant malfunction had been recently reported) or corollary information (e.g., a fish kill has been observed along with the extremely low dissolved oxygen readings)? Also check for consistency between similar parameters. For example, total dissolved solids and conductivity should track together—if one goes up, so should the other. So should total solids and turbidity.

- *Are there outliers? (Findings that differ radically from past data or other data from similar sites.)*

Values that are off by a factor of 10 or 100 should be questioned. Follow up on any data that seems suspect. If you can't come up with an explana-

tion for why the results are so unusual, but they are still within the realm of possibility, you may want to flag the data as questionable. Ask an experienced volunteer or program staffer to sample at that site as a backup until uncertainties are resolved, or work with the volunteer to verify that proper sampling and analytical protocols are being followed.

- *Are the field data sheets complete?*

If a volunteer is consistently leaving a section of the sheet incomplete, follow up and ask why. Instructions may not always be easily understood. All sheets should include site location and identification, name of the volunteer, date, time, and weather conditions.

- *Are all measurements reported in the correct units?*

You should minimize the chance for error by including on the data form itself any equations needed to convert measurements, and specify on the form what units should be used. Check the math. All field data sheets should be kept on file in the event that findings are brought into question at a later date.

Review Information in Your Database

Once volunteer data enters a computerized database, it can take on a life of its own. It is a phenomenon of human nature that data suddenly seem more believable once computerized. Therefore, be sure to carefully screen information as soon as you enter it into a database. Then review a printout (preferably with a fresh pair of eyes) against the original field data sheets. One way to minimize transcription errors is to design the computer input screens to look like the field data forms.

As a further check, you can run some simple calculations like determining medians and means to make sure no errors have slipped through. (If the median and the mean are very different, an outlier may be skewing the results.) Again, if you uncover unusual data points that cannot be explained by backup information on the field data sheets or the comment field in the database, flag the data as questionable until it can be verified.

Review Your Final Results

Once volunteer monitoring data has been entered into a database, the next step is to generate reports on the findings of the data. Even at this stage you should continue to look for inconsistencies and problems. For example, you should:

- Review findings against previous years' data.
- Look for outliers on graphs and maps.
- Not remove data just because you don't like it, but do investigate findings that are unusual or can't be explained.

By the time you present your final results to your volunteers or other data users, you should feel fully confident that you have assembled the best possible picture of water quality conditions in your study streams.

Develop a Coding System

A coding system will help simplify the tracking and recording of data. Make sure, however, that the system you create is easily understood and simple to use. Codes developed for sample sites, parameters, and other information on field and lab sheets should parallel the codes you use in your database. If you will be sharing your information with a state or local natural resource agency, you may want your coding system to match or complement the agency system.

Sample Sites: Because sample sites tend to change over time, it is important to have a site numbering system that accommodates change. A good convention to follow is to use a site coding system that includes an abbreviation of the waterbody and a site number (e.g., CtR020 for a site on the Connecticut River). For consistency, you might choose to start the site numbers at the downstream end of the stream and increase them as you move upstream (e.g., the first Connecticut River site would be CtR010, the second CtR020, etc.). Leave extra numbers between sites to allow for your program's future expansion.

Water Quality Parameters: It is also important to develop a coding system for each of the water quality parameters you are testing. These are the codes you will use in the database to identify and extract results. To keep the amount of clerical work to a minimum, abbreviate without losing the ability to distinguish parameters from one another. For example, EC could represent *E. coli* bacteria and FC fecal coliform bacteria.

Spreadsheets, Databases, and Mapping Software

Today's computer software includes a variety of spreadsheet and database packages that allow you to sort, manipulate, and perform statistical analyses on the data you have entered into the computer. For most applications, spreadsheets are adequate and have the advantage of being relatively simple to use. Most spreadsheet packages have graphics capabilities that will allow you to plot your data onto a graph of your choice (i.e., bar, line, or pie chart). Examples of common spreadsheet software packages are Lotus 1-2-3, Excel, and Quattro Pro.

Database software may be more difficult to master and usually lack the graphics capabilities of spreadsheet software. If you manage large amounts of data, however, a database is almost a necessity.

Using a database, you can store and manipulate very large data sets without sacrificing speed. The database can also relate records in one file to records in another file. This allows you to break your data up into smaller, more easily managed files that can work together as though they were one.

If you use database software for storage and retrieval, you may still want to use a spreadsheet or other program with graphics capabilities. Many spreadsheet and database software packages are compatible and will allow you to transport sets of data back and forth with relative ease. Very large data sets can be organized and manipulated in a database. Specific parts of the data (such as results for a particular metric from all stations and all sampling events) can then be transported into the spreadsheet, statistically analyzed, and graphically displayed. Examples of popular database software packages are dBase, FileMaker Pro, and FoxPro.

An effective way to display your data is on a map of the stream or watershed. This clearly illustrates the relationship between land uses and the quality of water, habitat, and biological communities. This type of graphic display can be used to effectively show the correlation between specific activities or land uses and the impacts they have on the ecosystem. Simple personal computer-based mapping packages are available. They allow you to enter layers of data and conduct spatial analysis of that data.

Systems that allow you to map and manipulate various layers of information (such as water quality data, land use information, county boundaries, or geologic conditions) are known as Geographic Information Systems (GIS). They can vary from simple systems run on personal computers to sophisticated and very powerful systems that run on large main frames. For any GIS application, you need to know the coordinates of your sample sites—either

their latitude and longitude, or some alternate system such as an EPA River Reach File identifier. You can also locate your sites on a topographic map that can be digitized on to an electronic map of the watershed. Once these points have been established, you can link your database to the points on the map, query your database, and create graphic displays of the data.

Powerful GIS applications typically require expensive hardware, software, and technical training. Any volunteer program interested in GIS applications should consider working in partnership with other organizations such as universities, natural resource agencies, or large nonprofit groups that can provide access to a GIS.

Many people are capable of writing their own programs to manipulate and display data. The disadvantage of using a “homegrown” software program, however, is that if its author leaves, so too does all knowledge about how the program works. Commercial software, on the other hand, comes with consumer services that provide over-the-phone help and instructions, user’s guides, replacement guarantees, and updates as the company improves its product. Also, most commercial programs are developed to easily import and export data in standard formats. This feature is important if you want to share data with other programs or organizations—all you need are compatible software programs.

STORET

EPA’s national water and biological data storage and retrieval system, STORET, is being modernized and will be available in 1998-1999. Volunteer programs are encouraged to enter their data into the modernized STORET. Individual systems will “feed” data to a centralized file server which will permit national data analyses and through which data can be shared among organizations. A specific set of quality control measures will be required for any data entered into the system to aid in data sharing. For more information, see the EPA web page at www.epa.gov/owow/STORET.