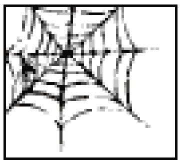
The Volunteer Monitor

The National Newsletter of Volunteer Water Quality Monitoring
Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Special Topic: Managing and Presenting Your Data

Imagine this . . . Your volunteer monitors are doing a splendid job! Data sheets are pouring in. Now what? This issue talks about all the things you need to do after your monitoring information is collected. How do you get those numbers into a format that makes sense to you - and to others? How do you turn your data into a story? The steps are summarized briefly in the schematic below, and discussed in detail in the articles.

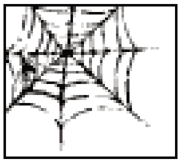
DATA
||
\/
entry and validation
.
preliminary summary
.
interpretation (findings, conclusions, and recommendations)
.
presentation to different audiences
||
\/
USE

Co-Editors: New Hampshire Lakes Lay Monitoring Program (NH LLMP)

This issue was co-edited by the New Hampshire Lakes Lay Monitoring Program, a collaborative effort between the University of New Hampshire and lake and watershed associations across the state. In its 17 years of operation, NH LLMP has helped train volunteers for lake, stream, river, and estuary monitoring and watershed evaluation.

The following articles appear in this edition of *The Volunteer Monitor:*

- Interpreting Your Data
- Using Data in the Classroom
- Using Graphs to Tell Your Story
- "Variability Happens," Basic Descriptive Statistics for Volunteer Programs



About The Volunteer Monitor

The Volunteer Monitor newsletter facilitates the exchange of ideas, monitoring methods, and practical advice among volunteer environmental monitoring groups across the nation.

Subscribing

The Volunteer Monitor is published twice yearly. Subscriptions are free. To be added to the mailing list, write to the address below. Your subscription will start with the next issue.

Reprinting Articles

Reprinting of material from *The Volunteer Monitor* is encouraged. Please notify the editor of your intentions, and send a copy of your final publication to the address below.

Participating

Let us know what topics you want to read about, and what information you have to share.

Address all correspondence to: Eleanor Ely, editor, 1318 Masonic Avenue, San Francisco, CA 94117; telephone 415/255-8049.

Rotating Co-editors

The Volunteer Monitor has a permanent editor and volunteer editorial board. In addition, a different monitoring group serves as co-editor for each issue. This unique structure ensures stability while allowing a variety of viewpoints to be represented.

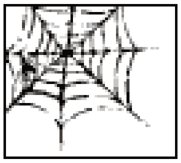
This issue was co-edited by the New Hampshire Lakes Lay Monitoring Program, a collaborative effort between the University of New Hampshire and lake and watershed associations across the state. In its 17 years of operation, NH LLMP has helped train volunteers for lake, stream, river, and estuary monitoring and watershed evaluation.

Back Issues

The following back issues are available:

- Fall 1991 - Biological Monitoring (photocopy)
- Spring 1992 - Monitoring for Advocacy
- Fall 1992 - Building Credibility
- Spring 1993 - School-Based Monitoring
- Fall 1993 - Staying Afloat Financially
- Spring 1994 - Volunteer Monitoring: Past, Present, & Future
- Fall 1994 - Monitoring a Watershed

To obtain back issues, or additional copies of this issue, send a self-addressed stamped envelope, 9 x 12 or larger, to *The Volunteer Monitor*, 1318 Masonic Ave., San Francisco, CA 94117. First-class postage is 78¢ for one issue, $1.24 for two, and $1.47 for three. For $3.00, you can get up to 15 copies. For larger orders, please call for shipping charges.



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

From the Editor

Data Is, Data Are

Anyone who writes about environmental monitoring quickly comes up against a minor but annoying question - is the word *data* singular or plural? Until now, I followed a simple policy: always treat *data* as plural. But sometimes that just doesn't work. Take, for example, this sentence: "Once data is in a database, it takes on a life of its own." You really don't want to change it to "Once data are in a database, they take on a life of their own," conjuring up the image of thousands of individual data points hopping around. Clearly in this case the writer was thinking of the whole body of data as a single unit.

In this particular issue of the newsletter, the word *data* seemed to turn up in nearly every sentence. The singular/plural dilemma was reaching acute proportions. Finally I dragged out the style manuals. For those of you who care about such things, here's what I found. According to *Webster's Dictionary of English Usage* (Merriam-Webster, 1989), *data* "is used in two constructions--plural, with plural apparatus, and singular, as a mass noun, with singular apparatus. Both constructions are fully standard . . . The plural construction is more common."

Next Issue: Urban Monitoring

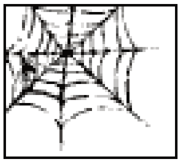
The next issue of *The Volunteer Monitor* will focus on "Monitoring in an Urban Environment," and the coediting group will be the Texas Watch Program, coordinated by the Texas Natural Resource Conservation Commission. Anyone wishing to contribute an article on the theme of urban monitoring is invited to contact the editor.

Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Staff at *The Volunteer Monitor*

This project has been partially funded by the U.S. Environmental Protection Agency. The contents of this document do not necessarily reflect the views and policies of EPA, nor does mention of trade names or commercial products constitute endorsement or recommendation of use.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Great All-American Secchi Dip-In

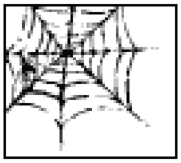
This summer, volunteer lake monitors across the U.S. will be asked to dip their Secchi disks for science as part of the Great All-American Secchi Dip-In.

The Dip-In was piloted last summer in six Midwestern states and was a rollicking success, with 826 volunteers collecting water clarity data on 843 lakes. This year, the Dip-In goes nationwide. Volunteers will take a Secchi disk reading in their lake between July 1 and July 9 and report their results to Dip-In Headquarters. The data will be analyzed, mapped, and reported back to all participants.

The Dip-In is being coordinated by the North American Lake Management Society (NALMS) and Kent State University under a grant from the U.S. EPA.

For more information, contact Dip-In Headquarters at 216/672-3849.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Kids Outdo "Expert"

by Susan Wertz

in a magically beautiful place in southern Puget Sound, the "big D" - as in big development company - proposed to develop a half-mile stretch of beach and 1,000 acres of uplands. The project included a 1,800- foot pier and a manufacturing plant.

An environmental impact statement was prepared. Reams of paper were submitted. A series of public meetings were held. At one especially pivotal meeting, the developer called a special expert witness - a marine biologist from New York City. All attention was on the biologist as he announced to the crowded room that this project would have no impact on the beaches and tidelands because the beach was virtually devoid of life.

A buzz circulated through the audience. How could this be? Our beach dead? Worthless? Many scoffed at this so- called expert - but I had the ways and means to counteract his claim. I had a 7th grade biology class!

My students and I consulted a topographic map, gathered together the tools of the field-study trade - identification books, small shovels, and nets - and headed for the beach in question. We found a mixed mud, sand, and gravel beach with fairly low wave energy and an amazing view of Mount Rainier. Like most south Puget Sound beaches, it lacked the spectacular tidepool life of the open Washington coast; but we knew that beaches like this typically support a rich array of plants and animals, including succulent clams, oysters, and mussels. Many of these animals are well camouflaged, and during low tides they seek shelter in deeper water, in the sand, and under rocks and seaweed.

So, what did the students find on this beach that the expert had declared "devoid of life"? In an hour and a half, working in teams (and following proper beach etiquette), they listed 85 different species on a big sandwich board we had set up on the beach. The students were convinced they could have found more if we could have stayed longer.

That summer, at another critical hearing before the county commissioners, I presented my students' list. The long list, generated by 13-year-olds, certainly got the commissioners' attention! The crux of my testimony was the issue of trust: If the development company was buying experts to give incorrect and unsubstantiated information, what else was being falsified in their proposal?

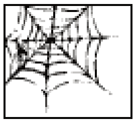
The permits were not granted. The pier was not built. The shores with their steep, hundred-foot banks remain, and Mount Rainier still stands guard over the beach creatures that hide during low tide.

Susan Wertz *is a science teacher at North Thurston High School in Lacey, WA.*

Stream Manual Ready for Field Testing

A draft version of EPA's *Volunteer Stream Manual: A Methods Manual* is now available to volunteer groups who would like to field test the methods described. The manual covers a wide variety of stream monitoring procedures, including watershed surveys, chemical monitoring, and biological monitoring.

This draft manual is one in a series of volunteer methods manuals produced by EPA. It is scheduled to be issued in final form in early 1996. For more information, contact Alice Mayio, EPA Volunteer Monitoring Coordinator, at 202/260-7018.



Geographic Information Systems: A Guided Tour for the Digitally Challenged

Just imagine . . .

To determine where to add a new sampling station for volunteer river monitoring in Montana, the program coordinators visit the state library. There, staff members help them use a computer to create a map that displays existing federal and state monitoring stations in relation to the known range of an endangered fish species as well as the human population density in various subwatersheds. A few hours later, the coordinators leave carrying a copy of the map plus a table of historical data for the existing stations.

Meanwhile, in New Hampshire, decision makers view a three-dimensional representation of the depth contours of their lake, which allows them to visualize the lake as a series of connected basins. Color coding, depicting water quality trends documented by volunteer monitoring, shows that some basins are showing signs of degradation. As a result, new zoning decisions are made to better protect the lake.

Seem like something out of Volume 13 (year 2001) of *The Volunteer Monitor* newsletter? Actually, it's happening today through the use of Geographic Information Systems (GIS for short). The program coordinators in the first example used the state's GIS, the Montana Natural Resource Information System (NRIS), which is housed at the state library and available to the public. The New Hampshire example occurred last year and was the culmination of a six-year study of the Squam Lakes. Volunteer monitoring was very much a part of that study - in fact, the existence of a 10-year water quality dataset collected by New Hampshire Lakes Lay Monitoring Program volunteers was one reason the Squam Lakes were chosen for the study in the first place.

What is GIS?

In simple terms, a GIS is a system that allows us to store, retrieve, integrate, analyze, and display descriptive information, such as water quality measurements, on a locational frame of reference (usually a map). Generally the term GIS refers to the combination of hardware (computer, printer or plotter, etc.) and software designed to work with a collection of information referenced by location. (See Typical Components of a GISfor more on GIS software.)

The different collections of data within a GIS - monitoring data, soil classifications, land cover, etc. - are commonly considered ""layers" of information. In fact, the early GIS systems developed about 20 years ago by landscape architects were crude replacements for the practice of hand-tracing different overlays on the clear plastic Mylar.

The Power of GIS

When most of us think of GIS, the first thing that comes to mind are the nifty color maps that hang on the walls of many an agency office or town hall lobby. The majority of these maps are just selected layers of the GIS database. Yet the potential of GIS is much more far-reaching. The important difference between a true GIS and a mapping or drawing program such as CADD ("Computer Aided Drafting and Design") is the ability of the GIS to take multiple layers and analyze relationships among them to actually create a new layer--as opposed to simply drawing two layers on top of each other. For example, you could ask a GIS to create a layer containing areas that meet a certain condition (e.g., wet soils); or areas that intersect (or don't); or sites at a specified distance from a road, well, or sewage treatment plant. GIS systems also offer additional tools, such as "weighting," in which one criterion can be emphasized over another, and "coding," which groups together a specified range of data as a new layer (so you could code all lakes with summer average chlorophyll levels of 0 to 3 µg/L as "excellent," 3 to 7 as "good," etc.)

The power of a GIS comes largely from its ability to analyze relationships, patterns, and trends and depict the results visually. A GIS can create a map that portrays storm event water quality data in relation to urban centers, or stream macroinvertebrate diversity with respect to locations of discharge sites or landfills. More advanced GIS systems can perform "what-if" scenarios - showing, for example, the predicted impacts of proposed regulations on a natural resource. They can also create three-dimensional models that display elevation or depth along with selected information layers. This technology has been used to create "viewscapes" that show how a proposed clearcut would affect the view from various points on surrounding roads.

Accessing GIS

At this point you may be thinking, "Sure, GIS is impressive, but is it really accessible to volunteer monitoring programs? How much fancy equipment do we need to buy?"

Actually, you don't need to own any technology to interact with GIS. For most volunteer monitoring groups, the gateway to GIS technology will be via a partner, such as a university or a government agency. The partner will have the mainframe computers, UNIX work-stations, or PC networks and the peripheral hardware (scanners, digitizers, and plotters) to support a full-featured GIS system. They will have high-end GIS software: UNIX-based packages like ArcInfo and ArcGrid by ESRI (Environmental Systems Research Institute), or GRASS (public-domain software that can be downloaded for free from the Internet). Most important, the partner will have staff capable of operating the system, and many data sets ("layers") on soils, land cover, demographics, and so forth that can be combined with your monitoring data for sophisticated analyses.

Where can you find such a partner? A good start would be to check whether your state has a central GIS depository. Often these centers are located at the state university, the statewide planning office, or the state natural resource or environmental protection agencies. The collaboration between a GIS partner and a volunteer monitoring program is by no means a one-way street. In New Hampshire, the Lakes Lay Monitoring Program provided our Lake Winnipesaukee data to a regional planning agency, which used their GIS to combine the volunteers' data with additional layers (e.g., zoning information, satellite land cover data). Both parties benefited: The planning agency got the data they needed for their analysis of land use regulations and water quality, and the Lay Monitoring Program got a map and data tables that helped us define which areas of the lake were experiencing development pressure. As a result, we added five new monitoring sites.

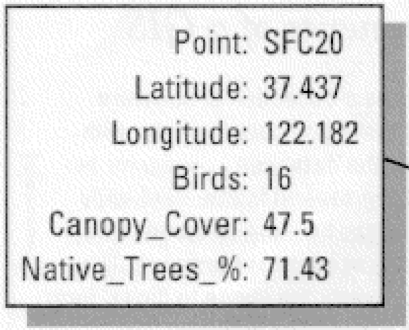
Buying GIS Software

But what if your monitoring program wants to get more directly involved in GIS? If you're in the market to purchase your own GIS software, what options are available? For IBM PCs and PC-compatibles, the most widely used full-featured GIS software is PC-ArcInfo (ESRI). However, to be used efficiently and effectively this package requires a highly skilled programmer and a substantial investment in hardware: high resolution video graphics, a minimum of 16 megabytes of RAM, and 500 megabytes to a gigabyte of storage.

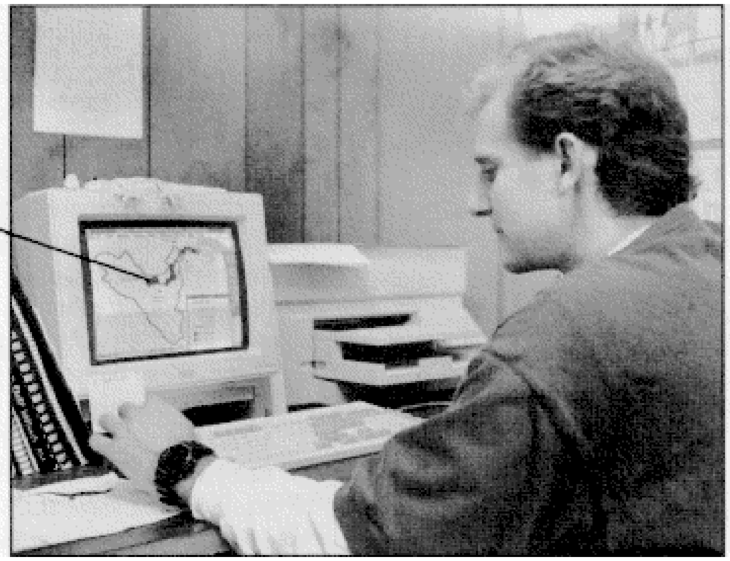
For many volunteer monitoring groups, the most practical alternative is probably a program like ArcView 2 (recently released from ESRI) or MapInfo (from Map Information Systems). ArcView 2 allows you to take information created in ArcInfo (or from other sources) and view it on screen in a variety of formats; add, delete, and manipulate data; query the database; and generate maps. MapInfo has similar capabilities; in addition, you can purchase an add-on feature that allows you to input data from a digitizer. ArcView 2 is most useful if you have a partner with ArcInfo who can provide you with the information layers you need. Otherwise, you'll need to purchase geographic files and databases. Both ArcView 2 and MapInfo are more user-friendly than PC-ArcInfo, but neither has the capabilities for 3- dimensional analysis or predictive modeling that a full- featured GIS offers. "I think ArcView 2 is the wave of the future," says Mike Rigney, Director of Coyote Creek Riparian Station. "It's really inexpensive, it has fairly sophisticated querying capabilities, and it's very good at drawing pretty maps."

GIS and Volunteer Monitoring

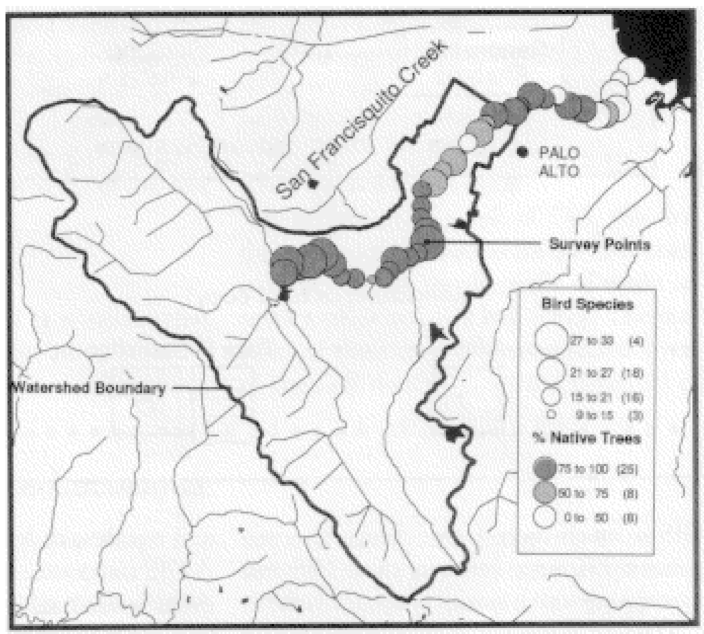
The volunteer monitoring program that Mike Rigney coordinates--Community Creek Watch, near San Jose, California--is on the cutting edge of GIS involvement for a volunteer group. Creek Watch has its own in-house GIS, using both MapInfo and GRASS software on IBM-PC computers. Rigney estimates the program's costs for GIS- related hardware and software at about $8,000. Monitoring site location latitudes and longitudes are determined with a Global Positioning System receiver (see Determining Site Locations in this issue). The volunteers' data - including results from water chemistry tests, bird counts, riparian vegetation surveys, and habitat assessments - are entered into the GIS, which can generate maps showing relationships between the various parameters (see the second illustration below).



Clicking on any monitoring site location along the creek brings up an onscreen box displaying data for that site.



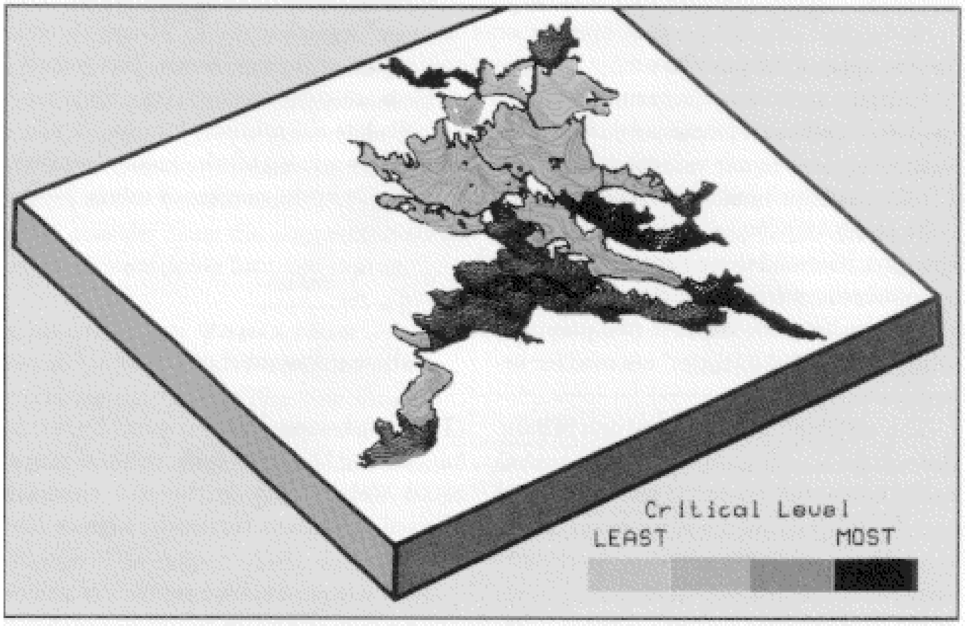
Creek Watch data manager Charles Preuss uses a GIS to create a map that shows two variables—number of bird species and percent native trees—in a geographic context. A hard copy printout of the map is shown below.



Why did Creek Watch consider it worthwhile to invest substantial time and money in GIS technology? Quite simply, they wanted to boost the chances that their data would be used by local decision makers. As Rigney puts it, "Stressed and harried public employees are much more likely to look at a map than a database."

Getting data into a format that local planners could use was also an important goal for the Squam Lakes study in New Hampshire. Under the direction of the New Hampshire Office of State Planning, a statewide task force used GIS to create a resource inventory of the Squam Lakes watershed. One of the data layers input into the GIS consisted of 10 years of water quality data collected weekly during the ice-free season by Lakes Lay Monitoring Program volunteer monitors. Other layers contained information on hydrology, elevation, zoning, land cover, fisheries, and loon habitat.

Using the extensive GIS data collected by the task force, we conducted further analyses at the University of New Hampshire. GIS modeling allowed for the visualization of the lakes' multibasin character, and GIS analysis integrated the various data layers to show which basins were most critical (see illustration below). For example, a basin with extensive fisheries and declining water quality would be defined as "critical." Through GIS analysis we also discovered relationships between characteristics of subwatersheds and water quality in the basins they drain into.



Several data layers were combined in a GIS to produce this 3-dimensional image of critical basins in the Squam Lakes.

The towns in the Squam Lakes watershed don't have their own GIS technology - nor do they need it in order to make use of the study results. We used the GIS to generate low-tech products, such as maps and tables, that were provided to local planners. One map shows at a glance where the critical basins are (critical areas are colored red, and less critical ones in cooler tones) as well as which subwatersheds drain to a given basin.

Moving Toward GIS

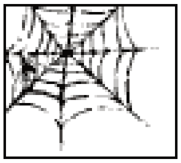
It seems as if everyone's using GIS these days - agencies, scientists, local planners, and corporations. Even though it's relatively new, the technology is fast assuming a central role in environmental management.

Volunteer monitoring groups are also moving toward GIS. Several examples were mentioned above, and there are many more. In a "nature mapping" project coordinated by the Washington Department of Fish and Wildlife and the University of Washington School of Fisheries, schoolteachers and students collect data on birds, amphibians, reptiles, and mammals and send it to the School of Fisheries, where it's entered into a GIS. ESRI has donated 100 free copies of ArcView 2 for participating teachers to use in their classrooms. And in Deerlodge, Montana, the state Department of Health and Environmental Sciences is using its GPS receiver to determine exact site locations for a volunteer group of senior citizens who are inventorying contamination sites around public supply wells. Eventually the volunteers' data will be fed into a GIS that will analyze the findings and produce maps.

The future looks even more promising. GIS technology is getting cheaper and easier to use - computer prices are dropping, the software is getting more user-friendly, more supporting data layers are being compiled, and efforts are under way to standardize data formats. So if your group hasn't yet given much thought to how you can participate in this new technology, now may be the time to start looking ahead. Even if you have no immediate plans for using GIS, you can be prepared by finding the exact locations of your sites. That way your data will be ready to plug into a GIS whenever the opportunity arises.

In many ways, volunteer monitoring and GIS seem tailor- made for each other. Because the power of a GIS is directly dependent upon its information base, the technology has a notoriously voracious appetite for data. And data is what volunteer monitoring programs have. In fact, volunteer monitors may soon find potential GIS partners knocking at their doors. That's exactly what happened to two Rhode Island programs, Rhode Island Watershed Watch and River Rescue. A group of Rhode Island Sea Grant researchers wanted to use GIS to create a map defining which water bodies in the state were at high risk for zebra mussel invasion. In fresh water, calcium - which the mussels need for their shells - is a limiting factor. Water with low calcium levels is not at risk for infestation. So the researchers needed up-to-date calcium data, and there were only two places to get it - the databases of the two volunteer monitoring programs. Watershed Watch had calcium data on lakes and ponds, and River Rescue had the data for rivers. Karen Tammi, one of the researchers, says, "We could not have done the study without the volunteers' data."

Jeffrey Schloss *is a Water Resource Specialist with UNH Cooperative Extension and the Coordinator of the New Hampshire Lakes Lay Monitoring Program, UNH Cooperative Extension, 109 Pettee Hall, 55 College Rd., Durham, NH 03824; 603/862-3848.*



Determining Site Locations

Before your monitoring data can be used in a GIS, each piece of information must be linked with a specific geographic location, usually defined by latitude and longitude. A description like "the Elm Street bridge across from the high school" won't do. So how do you pinpoint the exact location of your monitoring sites?

A low-tech approach is to estimate latitude and longitude from a USGS topo map. However, it can be tricky to figure out exactly where on the map your site is. The task is easier if you can relate your site to a nearby landmark (road, cemetery, etc.) that is marked on the map.

If you're monitoring a stream, another option is to use EPA's River Reach system, which assigns a unique number to every stream segment in the U.S. Your regional EPA office can provide you with a map of your area showing the River Reach numbers. Most GIS systems can accept this number as a locator.

The most accurate method is to use a device that receives satellite data from the U.S. Department of Defense's "Global Positioning System" (GPS) - 24 orbiting satellites that constantly transmit their position. At any given time and place, at least four satellites will be in "view." A GPS receiver can pick up the signals and use them to determine latitude and longitude (and altitude, if desired). GPS receivers range in price from about $300 to $6,000 with accuracy increasing along with price - but you don't necessarily need to buy one. Since determining site locations is a one-time task, you may be able to borrow a receiver from a cooperating agency.

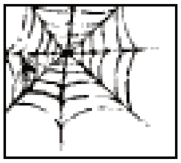
In the Creek Watch program near San Jose, California, volunteers use a GPS receiver called GeoExplorer that is accurate to within about 3 meters (see photo below). This device sells for about $3,000 (fortunately, Creek Watch received theirs as a donation from the manufacturer, Trimble Navigation). Receivers at the lower end of the price scale are accurate to about 30 meters.

ELEANOR ELY

Creek Watch intern Diane Kodama demonstrates the "GPS stance." The GPS receiver must be held well away from the body to avoid blocking satellite transmissions.

Creek Watch intern Diane Kodama demonstrates the "GPS stance." The GPS receiver must be held well away from the body to avoid blocking satellite transmissions.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



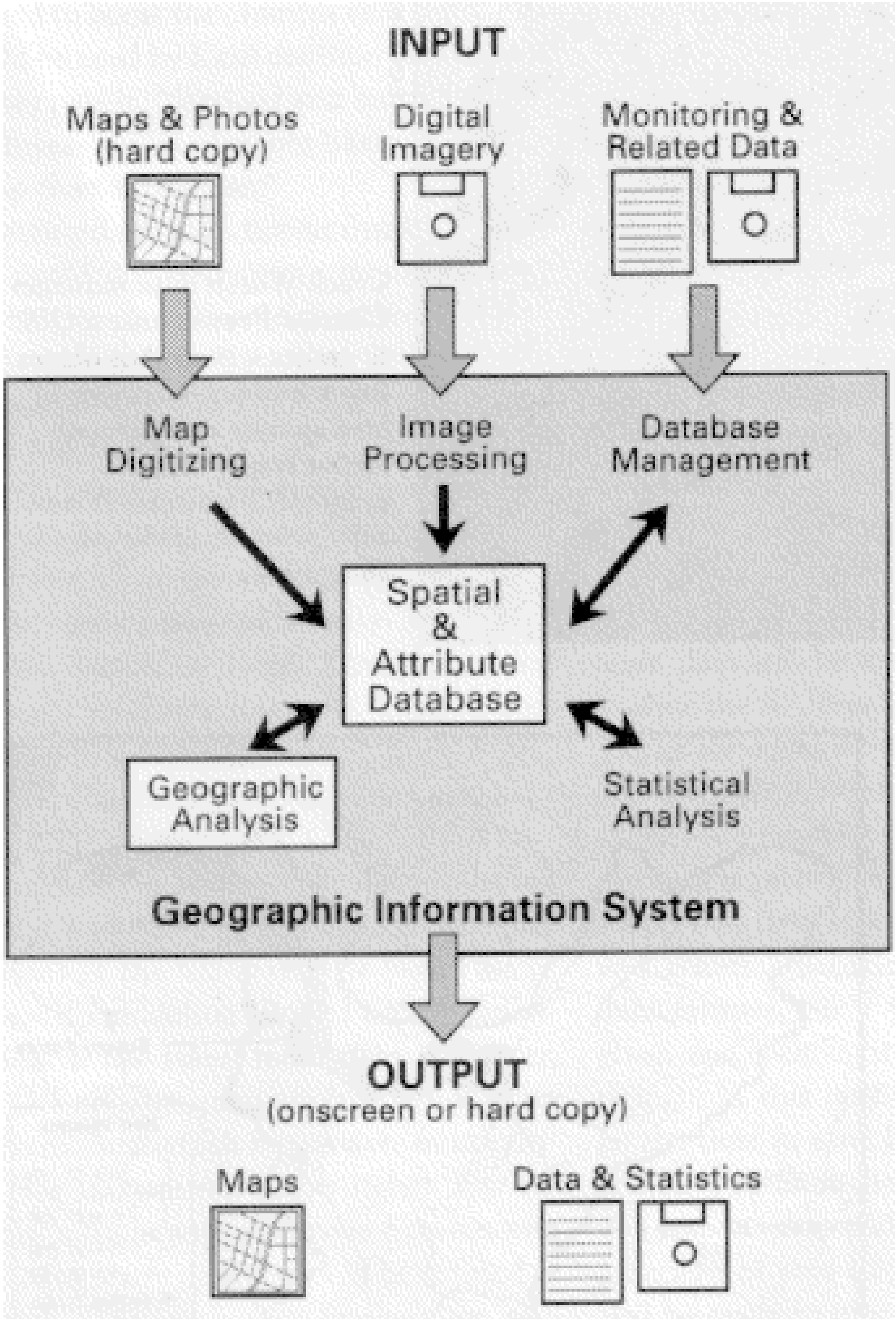
Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Typical Components of a GIS

GIS software can be thought of as a collection of software modules that perform different tasks and communicate with each other to create and modify the database.

A GIS can accept information in several forms. Hard copy images such as maps or photos may be input by means of a digitizer (electronic tracing table) or scanner hardware. The **map digitizing** software then converts the image to digital form. Some GIS systems have **image processing** software that converts digital imagery from remote sensing techniques (e.g., satellite data) to a usable format. Data (e.g., water quality data, sample site descriptions) can be input into the **database management** software via a keyboard or a floppy disk.

A GIS **database** has two components: a *spatial* component that describes the shape and position of geographic features, and an *attribute* component that contains the characteristics of these features. For example, for a volunteer lake monitoring program the spatial component would define the location and shape of the shoreline and the attribute component would contain such information as lake name, maximum depth, and monitoring results.

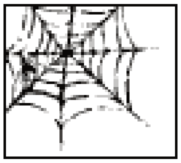


The heart of the GIS is the *geographic analysis* software, which is able to analyze the data according to the referenced locations. This capacity is the defining factor that makes a true GIS.

The **database management** software links the attribute data to the spatial data and allows for manipulation of the attribute data.

The products from the GIS analysis can be displayed as maps or data tables, either on the computer monitor or as hard copy output to a plotter or printer.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Seize the Data

by Steven Hubbell

As a civic activity, volunteer environmental monitoring is exciting and fun, personally meaningful and rewarding. As an environmental management tool, it must also include efficient procedures for handling the information volunteers collect. But since people are generally far more enthusiastic about data collection than data management, it's easy to convince ourselves that tasks like verifying numbers, entering data into a database and generating reports will somehow take care of themselves. Somewhere in the background, we imagine, a reclusive slide-rule wizard will wave a wand; then all we'll have to do is push a button and get a printout that explains at a glance whether the water is healthy or not.

Obviously, the reality of successful data management is rather more labor intensive. But what does it entail? How does a volunteer monitoring program turn the observations of monitors into a useful tool for protecting the environment? It's helpful to think of data management as a two-tiered process. One tier is the series of actions performed on the data--recording, verifying, evaluating, and so on. Underlying and supporting these actions is the organizational infrastructure, which should clearly specify the sequence of events, timelines for data entry and reporting, and staff responsibilities for each stage of data management. Both tiers must function together smoothly for a monitoring program to take full advantage of the information volunteers collect.

Boxes in the Corner

I first became aware of the complexity of managing volunteer monitoring data while doing research for my Master's thesis (on the topic of "assessing the credibility of volunteer-collected data"). When I visited the Colorado River Watch Network, I found that volunteers' data were being entered on a computer, but not in a form that could be integrated with data collected by the staff of the sponsoring agency (Lower Colorado River Authority). At the Texas Watch office, when I asked about the data, hands were waved in the general direction of some boxes. Texas Watch was almost two years old at the time (1992), with monitors on about 100 sites. There were quite a few boxes.

Steven Hubbell demonstrates the dire consequences of ignoring data management.

I suspect that similar stacks of boxes lurk in the offices of a number of volunteer monitoring programs. Once this pattern of data mismanagement begins, it's a challenge to turn the tide. But it can be done, and it must be done if the data are to become useful. What prevents volunteer programs from getting on with the job? Below are a few common misconceptions, excuses, and rationalizations that may be standing in your way.

"It's Not My Job"

Wrong. If you are involved in volunteer environmental monitoring, then data management is your job. The data collection efforts expended by volunteers deserve to be answered by an uncompromising commitment to organizing the information so it is readily available to volunteers and environmental managers alike. The following list of job categories is based on Texas Watch, a statewide volunteer monitoring network sponsored by the Texas Natural Resource Conservation Commission (TNRCC) and managed by an eight-member staff. In smaller programs, one person will perform more than one role. But regardless of where you fit into the picture, there are ways you can contribute to successful data management.

- **Program designers, planners, steering committee members, advisors:** This is where data management should begin. As the excitement of developing a citizen monitoring program begins to settle in, everyone involved in planning should ask themselves, "What are we going to do with the data?" ("We'll deal with it later" is not an acceptable answer.) The entire program's credibility can suffer if a good data management program is not planned before outreach and recruitment efforts start. Chances are that once the word is out, public response will exceed expectations.
- **Data managers:** If your program has the luxury of having a data manager on staff, don't expect him or her to personally perform the entire gamut of duties associated with data management. Other staff can help with checking data sheets, entering data, and proofing printouts. Also, avoid the temptation to ask the data manager to spend an inordinate amount of time tutoring colleagues

in basic computer skills. Texas Watch's first data manager not only was tasked with designing data management software and establishing a comprehensive mailing list database, but also was on call to assist staff from other agency programs with computer problems.

- **Program coordinators:** Each data sheet represents a volunteer's time. If a backlog is growing, the coordinator would be well advised to take action before the situation gets out of hand. This may mean restructuring priorities.
- **Trainers:** In addition to demonstrating proper sampling protocol, trainers should explain to volunteers what range to expect for each water quality variable, what to do if an unusual value is found, how to record the data, and when and to whom the data are to be submitted.
- **Volunteers:** Data review - an integral part of data management - begins on site with the volunteers. Volunteers should double-check their field sheets to ensure that the form is completely filled in and values are within a reasonable range. Also, volunteers should become so familiar with their sites that unusual findings leap out at them.
- **Supporters, sponsors, partners:** In many cases, sponsors such as river authorities, universities, municipalities, or industries are in a position to assist with data entry, reporting, and analysis.

"It's So Boring"

Well, maybe, but who knows what's hidden in those boxes of data sheets? There may be a site where the dissolved oxygen is in a steady decline. A data management system that includes reviewing the data sheets as they arrive might save the day, impress the governor, or win an award.

One Texas Watch group became so familiar with the warning signs (a slight rise in pH accompanied by a slight drop in dissolved oxygen) that they successfully predicted fish kills on Oyster Creek near the city of Sugar Land. Recently over $100,000 was committed for identifying potential solutions for the creek.

"Nobody Cares"

Guess again. One of the most frequently asked questions in volunteer monitoring programs is "What happens to my data?" People who consistently go out and collect water quality information aren't doing it just to collect numbers. They want to know what this information is telling us about the water.

Even where no water quality changes are evident, the data can be used to document normal, cyclic variations at a site. Often volunteer-collected data are the only data available for a stretch of water, and can be used to establish baselines. For example, as Amy Picotte reported in the Spring 1994 issue of this newsletter, a 15-year data set collected by Vermont Lay Monitoring Program volunteers was recently used as the basis for establishing numeric standards for phosphorus on Lake Champlain, Vermont's largest lake.

"We Can Always Do It Later"

When? Consider this: A monitoring program with just 10 sites monitored monthly will produce 120 individual data sheets in a year. Weekly monitoring of those 10 sites would yield 520 data sheets. And the more successful your program, the more data will come pouring in. Texas Watch volunteers now monitor more than 300 sites, most monthly, some weekly. A conservative estimate puts the program's yearly total at over 6,000 individual data sheets.

Data sheets pile up fast. And there are few sights more daunting to program staff or more disheartening to volunteers than boxes of backlog data.

Texas Watch Today

Texas Watch turned four in February 1995. (I joined the program as communications coordinator in 1993.) We knocked out the data backlog last August to produce the *1994 Texas Watch Report*. It took an idealistic intern, a supportive team effort, a competent and committed data-entry temporary employee (our data manager had accepted a position elsewhere in TNRCC), and an EPA grant deadline to get the job done. A lot of tense moments, overnight delivery expenses, and overtime hours could have been avoided if a data entry regimen had been established and followed in the years preceding this report.

As for the Colorado River Watch Network (a Texas Watch partner), the program now uses a data-entry system that is compatible with LCRA's staff-collected data.

Where We're Going

One of the most promising new directions for data management at Texas Watch is "remote access data exchange." This will mean data gatherers and data users can enter and retrieve Texas Watch data from computers around the state. Texas Watch is also working to refine and distribute its DRIPS (Data Retrieval and Information Processing System) software, developed to help volunteers and partners store, print, graph, and report data in a format compatible with the Texas Watch and TNRCC databases.

The possibilities are virtually unlimited. To say data management is tedious and endless is to see only half the truth. It can also be thoroughly rewarding. Imagine saving a species by your timely and diligent attention to your data. It could happen.

Note: A limited number of copies of the *1994 Texas Watch Report* are available. For a copy, please contact Steven Hubbell (address below).

Steven Hubbell is the communications coordinator for Texas Watch, Texas Natural Resource Conservation Commission, P.O. Box 13087, Austin, TX 78711; 512/239-4743.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995

Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Data Screening and Common Sense

by Janice K. Miller

The data-screening tips covered in this article are extremely simple. We are not talking about computer programming or advanced chemistry. We're talking about volunteers and program coordinators taking a look at their data and asking common-sense questions like, Is the field sheet completely filled out? Are the calculations correct? Are the right units reported? Do the results seem reasonable?

Perhaps it's because these questions seem so obvious that they're often overlooked--which is a shame, because after all the time and effort that goes into water quality testing, the small final step of checking these details can make the difference between data that will be valuable for years to come and data that can't be used.

Data collection involves a chain of responsibilities, and the data will only be as good as the weakest link in the chain. The people who do the field testing and sampling (usually the volunteers) are responsible for filling out forms correctly. The person who manages the data (usually the program coordinator) is responsible for seeing that it's correctly entered and interpreted.

No matter how carefully measurements are made, the data will only be as good as what's recorded on the data sheet. Above, Community Creek Watch volunteers record data on a California creek.

1. Designing a Good Field Sheet

Coordinators can get data collection started off on the right foot by designing data sheets that make it easy for volunteers to record all the necessary information in the proper format. Field sheets should include simple instructions and examples for calculations, and provide ample space for the following:

- Site name and exact location (latitude and longitude, directions to the site, or site code).
- Time and date of sampling.
- Volunteer's name and phone number.
- Weather conditions (recent as well as current).
- Name and model number of equipment or test kits used.

- Actual readings, including replicate readings, from the kits or equipment (don't just provide a space for the final answer). For example, for dissolved oxygen titrations, record the actual number of milliliters titrated as well as the final concentration of dissolved oxygen.

- Comments. Leave enough room for volunteers to record anything unusual they see (spills, new construction, dead animals, etc.) as well as any problems they may have in performing the tests.

2. Recording Data In the Field

Once volunteers leave the monitoring site, the field sheet is the only record of their efforts. No matter how carefully the tests were performed, the data will only be as useful as what's written on the form. Therefore it's essential that volunteer training include detailed instructions on filling out forms. Here are some points to go over with volunteers during training:

- Record any unusual conditions at the site. (When in doubt, write it down.)

- Record all instrument or kit readings, including units, on the form.

- Do not report a value of zero. Instead report "less than ___," filling in the blank with the lowest value that can be read with the equipment.

  Example: If the range of a test is 0 - 1 mg/L, the smallest increment is 0.02 mg/L, and the test result is zero, report "less than 0.02 mg/L" or "< 0.02 mg/L."

- If calculations are performed, show all formulas, calculations, and units.

- When reporting results of calculations, do not report excess decimal places. Use the following rule of thumb (this is a simplified approach to handling significant figures and decimal places, which works in most cases): Look at all the values that were used in the calculation, and find the measured value with the fewest decimal places. The final answer should have that same number of decimal places.

  Example:

  - Measured value Y = 7.7
  - Measured value X = 5.32
  - Constant value C = 12

  Calculation formula is (Y * X) / C

  Report the answer as 3.4 - **not** 3.41 or 3.413 (because the measured value with the least decimal places is Y, which has one decimal place).

- When rounding a final value, follow these two rules:

    1. Consider only one digit to the right of the digit you are rounding.
    2. If the digit to the right is a 5, "round to an even number." (Some people always "round up" a 5, but over time this method will introduce a slight positive bias into the data.)

  Example: Say you are rounding each of the following to one decimal place:

    - 3.548 rounds to 3.5. (Ignore the 8; just look at the 4.)
    - 3.758 rounds to 3.8. (Ignore the 8; just look at the 5. Round the 7 to the even number of 8.)
    - 3.458 rounds to 3.4. (Ignore the 8; just look at the 5. The 4 is not changed because it is even.)

If data sheets are to be mailed in, encourage volunteers to keep a copy in case the originals are lost in the mail. Having a reference copy also comes in handy if the program coordinator calls with questions.

3. Reviewing Data Sheets

The program coordinator should review the completed field data sheets as soon as volunteers turn them in--ideally, with the volunteer present (if your volunteers deliver their forms in person). Don't wait until six months later, when someone wants to use the data, to discover discrepancies. By that time it will probably be too late to fix problems that could have been easily resolved earlier.

In reviewing the sheets, check for the following:

- Is the sheet completely filled out? If not, make an effort to get any missing information as soon as possible.

- Were calculations performed correctly? Make sure the volunteer followed the rules listed above regarding decimal places, rounding, etc.

- If a value of "zero" is reported, convert it to the "less than ___" format.

- Check to be sure the values are possible. For example:

    - pH > 14 is not possible.

    - If water temperature is reported as "60 ° Celsius," a Fahrenheit thermometer was probably used. Convert to Celsius.

    - Dissolved oxygen readings over 13 should be checked. (A value over 13 is possible, if there is a lot of algal growth.)

- If replicates or other QA tests were done, are they within the limits of your data quality objectives? For smaller volunteer monitoring programs who don't have formally defined data quality objectives, use your judgment. Are the replicates reasonably close? For example, if the volunteer did three replicates and got 1, 18, and 20, don't use the 1. Explain your decision on the field data sheet.

- Were the proper chemical constituents reported? This is often an issue with tests for nutrients. Phosphorus results can either be reported "as phosphate" or--more often--"as phosphorus," and results for ammonia, nitrate, and nitrite can be reported as the respective compound or "as nitrogen." Double-check the constituent that the test kit or procedure measures, and make any necessary conversions.

  Example: The kit measures and reports total phosphate (PO4) and you want to report it as phosphorus (P). Convert as follows:

  Atomic weight of P = 30.9738
  Molecular weight of PO4 = 94.9714
  Measured value = 3.8 mg/L PO4
  To convert:
  (3.8 mg/L PO4) * (30.9738 / 94.9714) = 1.2 mg/L P

- Are the results in the same units that are used in the database? Volunteers will record the units from the kit or equipment they use. If these don't match what's in the database, make conversions as necessary. Remember that for water samples, mg/L is equivalent to parts per million, and æg/L is equivalent to parts per billion.

A careful review of the field sheets yields an extra benefit - it can alert you to problems with your procedures. For example, if you note a lot of variability in the results for a certain test, this could indicate that many of your results are in the lower range of the test procedure where precision is not as good. Ideally, you want to use a test procedure for which most of your results are in the midrange. So if the test range is, say, 0 to 100, and many of your values are less than 10, you may need to switch to a procedure whose range is 0 to 10.

4. Screening Data After It's In the Computer

Data will be far more useful to others once it has been entered into a computer database or spreadsheet. After entering the data, print it out and proofread it against the original field data sheets. (Preferably, the proofreader should not be the same person who entered the data.) After proofing, use the methods below to screen for possible problems and ensure that the data "makes sense."

- Look for outliers, using any or all of the following techniques:
  1. Compare with past data from similar sites;
  2. Graph the data;
  3. Visually scan for numbers that are off by a factor of 10 or 100;
  4. Calculate the mean and median and see if they are significantly different.

  If an outlier is found, check for problems with calculations or data entry.

- Look for consistency between similar parameters. For example:
  - Total solids should be greater than suspended solids.
  - Total dissolved solids and conductivity should track (go up or down) together.
  - Total suspended solids and turbidity should track together.
  - If the field data sheet says the sample was "dirty," solids and turbidity should be high.

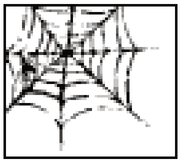
5. Fixing Problems

If you find inconsistencies in the data, follow up on the problem. A professional lab or another volunteer group doing similar work may be able to help interpret the data or suggest possible sources of error. If you can track down the error, fix it in the computer, and alert others to whom you have given the data so they can fix it also. The correct answer and the reason for the correction should be noted on the field data sheet. If you suspect a result is incorrect but can't figure out the correct result, remove the suspect result from the computer and note on the field data sheet why it was removed. But don't take out data just because you don't like it or it does not support your viewpoint. A "Q" code can be used to flag questionable data that has been checked for obvious problems.

Data in a computer has enhanced credibility, so do everything you can to make your monitoring database accurately reflect the conditions in the environment. This will help ensure that any decisions based on the data are good ones.

Janice K. Miller *is Water Resources Analyst for Unified Sewerage Agency, 155 N. First Ave., Suite 270, Hillsboro, OR 97124; 503/693-4493.*

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Designing a Data Management System

by Fred Lease

At some point in any monitoring program the problem of data management will inevitably rear its ugly head. Naturally, the larger the project, the uglier the data management tasks can seem. But even in smaller programs, data can be difficult to keep under good control.

Today, with reasonably priced computers and software available, a simple yet effective data management system can be developed to fit the needs of most monitoring programs. However, many headaches await those who charge blindly into the task. This article will provide some basic information on data management tools and techniques and will also address some of the problems encountered when setting up a monitoring database for the first time.

Why Computerize Your Data?

Simply put, a good data management system provides for efficient storage of information, wherein any piece of that information is accurately retained and readily accessible. Does data management always require a sophisticated computer program? Not at all. Many adequate databases reside between the covers of loose-leaf binders. However, a computerized data management system does offer many advantages, especially if you will be collecting data from many sites or 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., salinity from hydrometer 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.

Spreadsheet or Database?

Once you've decided to use a computer for managing data, the next step is deciding between a spreadsheet program or a database. These two types of software are not as different as they once were - newer versions of most spreadsheet programs offer some of the capabilities of databases, and vice versa. Still, each system has its own particular strengths and weaknesses.

In general, spreadsheets (e.g., Lotus 123, Excel, Quattro Pro) can perform a much wider array of mathematical and statistical manipulations, and are better at graphing. Database programs (e.g., dBASE, FoxPro, Paradox) can store a larger volume of data and are better at finding or "retrieving" information. Furthermore, the more sophisticated database programs - including the three just mentioned - 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 a database with two separate files: a volunteer file with volunteer names and addresses, and a site file 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. ("Flatfile" database systems, such as FileMaker Pro, do not have this relational capability.)

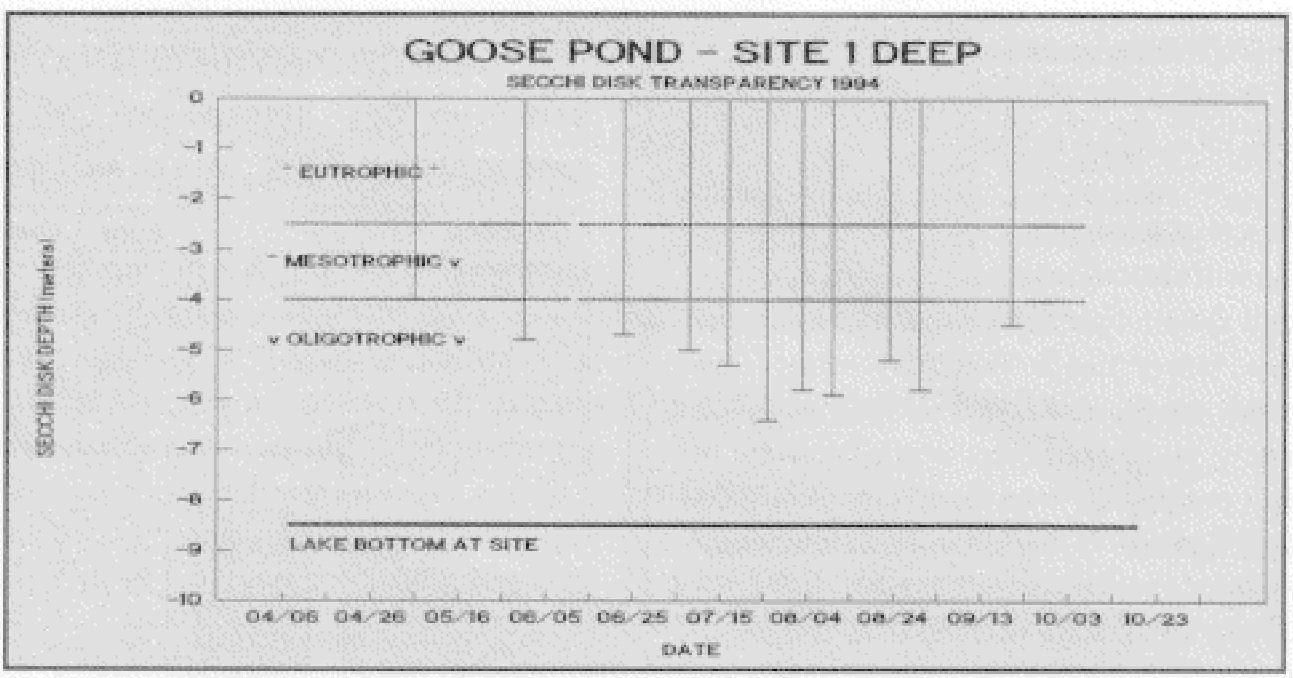
Vermont-based River Watch Network is now in the process of converting from a spreadsheet system to a relational database (FoxPro). One reason for making the change was the sheer volume of data (RWN handles data from dozens of river monitoring groups). "After a certain point, spreadsheets simply run out of gas," says Geoff Dates, Science Coordinator for RWN. In addition, RWN needed the relational capability of a database. "We've been using spreadsheets for years," says Dates, "and now we have all these separate spreadsheets, for different groups and different years, and it would be very time-consuming and tedious to query across them. For example, we can't easily extract all the sites in New England where bacteria counts exceed water quality standards, because the information is stored in different spreadsheets."

Another advantage of databases is that they can easily be programmed to catch errors in data entry. If someone tries to enter data that doesn't conform to a specified range or style - say, a water temperature of 75°C rather than 7.5°C, or a nonexistent code - the program will not accept it. Databases are also better equipped to handle different types of data, including nonnumeric data, within a single file, whereas spreadsheets are primarily designed to handle quantitative data that is all of a kind - dollars or inches or pounds.

SarahRose Werner, who recently worked on the MURPHY database program for volunteer groups in Maine (see article in this issue), notes that "in general, more time and expertise are required up-front to set up a database in a database program than in a spreadsheet program, but once this is done, it's easier and quicker to train inexperienced data entry operators (e.g., volunteers) to use a database program."

With all the advantages offered by databases, why are so many volunteer monitoring programs using spreadsheets? Perhaps the main reason is that historically, spreadsheets were easier for novices to use. Older versions of database software required a thorough knowledge of programming commands and offered little or no on-screen help, whereas spreadsheets have been menu-driven since the early days of Lotus 1-2-3. Today's database programs are much more user-friendly; but by this time spreadsheet command structures are more widely known and understood. Further, the use of spreadsheets for business purposes has required a wide array of business people to work with them.

Some monitoring groups may find they need both types of software to manage their data optimally. For example, the New Hampshire Lakes Lay Monitoring Program uses a database program for data entry and storage, and a spreadsheet for statistical analysis and graphing. Given the large amount of data collected by Lakes Lay Monitoring Program volunteers (16 years of monitoring and about 1,000 lake site visits per year), a database program was essential. The database is set up with input forms that check the data as it is entered. Automated programs created in the database software generate data tables and provide summary statistics for each site. Then the data are imported into the spreadsheet to take advantage of the superior graphing and statistics capabilities that spreadsheets offer (see following illustration).



NH Lakes Lay Monitoring Program's spreadsheet program automatically produces Secchi disk graphs like this for each site. Automated features include scaling the graph based on the minimum and maximum reading; indicating interpretive ranges for oligotrophic, mesotrophic, and eutrophic; and drawing a reference line for the lake bottom at the site.

Which Brand of Software is Best?

When a volunteer group considers what brand of software to buy, cost may be the first factor that comes to mind (see Note at end). Can you get away with a less expensive program, perhaps one that was supplied with your computer? Maybe; but be careful. Cheaper spreadsheet programs are often limited in their mathematical, statistical, and data-sorting capabilities. Inexpensive databases may not be relational, and will not have the wide assortment of retrieval capabilities found in more sophisticated programs. Also, less expensive programs may be weak in graphing capabilities and printing options, whereas effective data presentation is often extremely important for monitoring groups. Jeff Schloss, coordinator of the New Hampshire Lakes Lay Monitoring Program, cautions, "In choosing a software package, don't rely on the expertise of one volunteer. If that volunteer leaves, you will be up a creek!" He recommends choosing software that is already supported by an agency or university that you work with, so that you will be able to receive training and assistance.

Another important consideration when purchasing software is how easily data files can be translated by another software program. For instance, Paradox database software was specifically designed by Borland for use with Borland's Quattro Pro spreadsheet software. According to Borland, file conversion from Quattro Pro to dBASE is also "easily" accomplished - but I am struggling with problems now because I did not question this claim. Easy it may be, but time-efficient it is not! Extra steps are required that can take up considerable time and require some additional computer savvy.

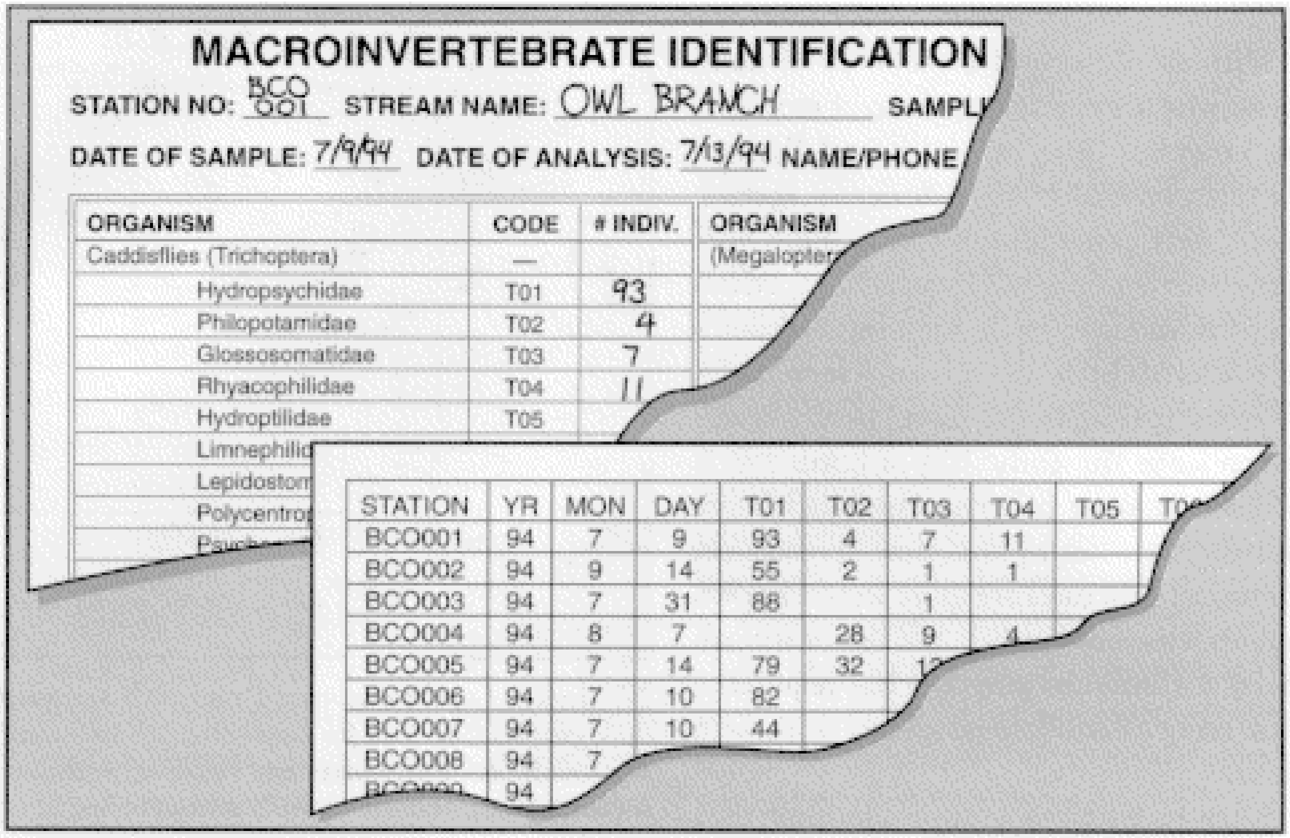
Take time, now, to ask other monitoring groups which software programs they are using and what problems they may be encountering with file conversion. Check with any groups or government agencies with whom you plan to share data, and make sure the software you are considering will easily convert to the file formats they require. If you have data analyzed by an outside lab that provides you the results in electronic format, you will also want to check compatibility with their system.

Basic System Design

When designing any data management system, begin at the beginning by asking yourself, "What do we need this system to do?" Analyze what information flows through the organization and what needs to be done with it. Make a list of all the tasks you want the system to accomplish. What types of output will you need? Reports? Graphs? What data will need to be in the system to support these outputs?

It's especially important for a volunteer monitoring group to think about who will input the data and make the reports. If the system will be used mainly by just one person, it can be designed primarily for speed and power. But if it will be used by a large number of volunteers, ease of use becomes the paramount goal. Speed would be sacrificed in favor of features like data validation, onscreen help, and user-friendly forms for data input.

Ease of data entry is always an important element to consider. Ideally, the spreadsheet or database input screen should be set up to look like the field or lab data forms filled out by your monitors. The person keying should not have to spend a lot of time hunting for information on field forms or searching back and forth across the computer screen for some place to input a data item. (See the figure that follows.)



MACROINVERTEBRATE IDENTIFICATION

STATION NO: BCO 001 STREAM NAME: OWL BRANCH SAMPL

DATE OF SAMPLE: 7/9/94 DATE OF ANALYSIS: 7/13/94 NAME/PHONE

ORGANISM	CODE	# INDIV.	ORGANISM
Caddisflies (Trichoptera)	—		(Megalopter
Hydropsychidae	T01	93	
Philopotamidae	T02	4	
Glossosomatidae	T03	7	
Rhyacophilidae	T04	11	
Hydroptilidae	T05		
Limnephilid			
Lepidostom			
Polycentro			

STATION	YR	MON	DAY	T01	T02	T03	T04	T05	T0
BCO001	94	7	9	93	4	7	11		
BCO002	94	9	14	55	2	1	1		
BCO003	94	7	31	88		1			
BCO004	94	8	7		28	9	4		
BCO005	94	7	14	79	32	[illegible]			
BCO006	94	7	10	82					
BCO007	94	7	10	44					
BCO008	94	7							
[illegible]	94								

The top form is used by Project Heartbeat volunteers analyzing macroinvertebrate samples in the lab. Volunteers record counts for each family next to the appropriate scientific name. A 3-digit code is printed next to each taxonomic name. Only the codes appear on the data input screen (bottom form). The keyer does not have to search for a long scientific name on the screen, or distinguish between similar names like "Hydropsychidae" and "Hydroptilidae."

Plan Ahead

You will also want to design a data management system that will be flexible enough to meet future needs that may arise. This sounds tough, and it is. How can you plan now to meet needs you won't be aware of until months or years down the road? You can't, exactly. But you can build the requisite flexibility into your system.

Take something as simple as site numbering, for instance. Suppose you are conducting 10 sampling projects, each having 10 sampling sites. One possible coding scheme might consist of a simple letter designation for each project, followed by a number for the site location (A1, A2, A3; B1, B2, and so on). This is fine until you ask the computer to sort your data by site number. The computer will place A10 ahead of A2, since it considers each digit individually from left to right. "Okay," you say, "we'll change the numbering scheme to A01, A02, etc." This is fine, too, as long as your database remains small. But what if you end up with more than 99 sites in one project or more than 26 projects?

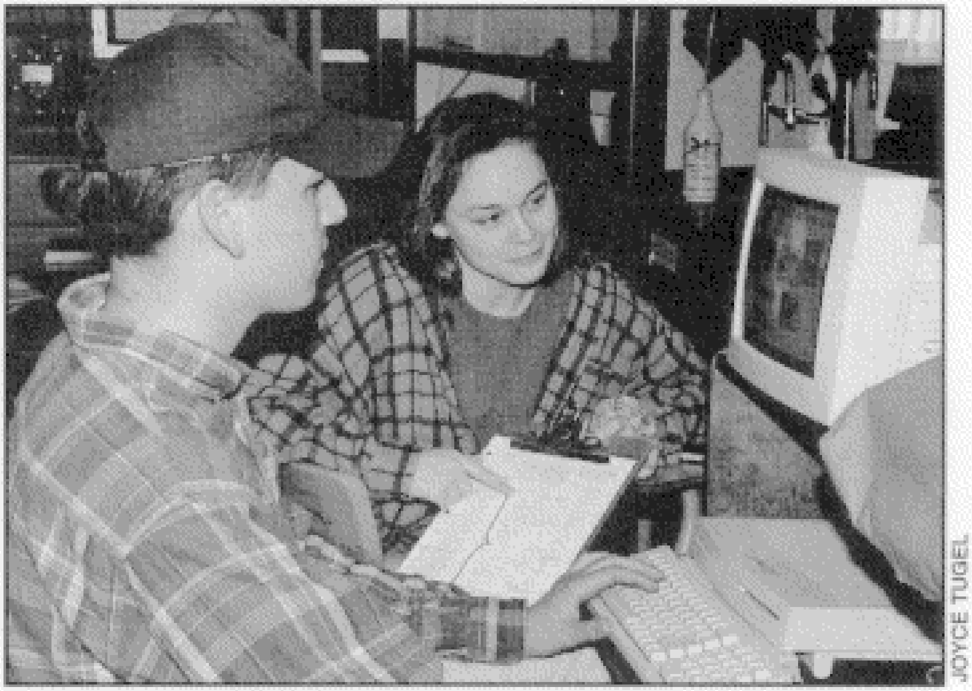
In Maryland Save Our Streams' Project Heartbeat, each site is identified by a three-letter code representing the name of the monitoring project, followed by three digits designating the particular site within the project. Thus, station number SHA001 designates site number 1 (001) of our State Highways Administration project (SHA). We could potentially create thousands of project codes, with up to 999 sites per project - far more than sufficient for our needs. With this coding system, we can pull together sampling data from many different projects and easily identify which samples pertain to which project.

If you are currently using a spreadsheet and anticipate switching to a database program in the future, planning ahead may help you avoid problems. For example, in early versions of Heartbeat spreadsheets Save Our Streams used two-line column headings. But later, when we converted to dBASE format, we found that the column headings had to be changed to one-line field names, which could contain up to 10 letters. Still later, we selected a statistical program that, as luck would have it, would only accept field names of eight letters or less. Each time an earlier field name was scrapped, we had to create new documentation for defining the names, and we had to change all the earlier spreadsheets. Deciding now which database software you will want to use later will enable you to accommodate that software's particular conventions.

Using Templates

Try to keep your data files as uncomplicated as possible. The more you load up a file with equations and fancy cell formatting, the larger it becomes and the more storage space it requires. If you do need to perform complicated calculations in a spreadsheet and find that disk space or computer memory problems plague you, try using a series of "template" files instead of expecting one spreadsheet to do everything.

For example, Project Heartbeat protocols call for calculating 23 metrics (indices such as total taxa, taxa richness, and so forth) to analyze the macroinvertebrate data. To accomplish this task, we use three templates in Quattro Pro. The first template is used for inputting raw "bug" counts from the volunteers' tally sheets. Once all the tallies from a sampling season have been keyed in, they are copied as a single block of data into a second template, which contains all the equations and automation instructions needed to calculate the metrics. This second template is never saved with data in it; it is saved only as an empty calculation spreadsheet. Even empty, it requires a great deal of disk space for storage and ties up a considerable amount of computer memory when it is open. Once the calculations are completed, the metrics are copied to the third, or output, template. The output template, like the input template, contains no equations. When the entire procedure is finished, the calculation template is closed without saving and the newly created raw data files and metrics output files are stored under new names. We save a considerable amount of disk space by storing only the empty copy of the calculation template. Furthermore, since the input template contains no equations and thus uses a relatively small amount of memory, data entry is quicker.



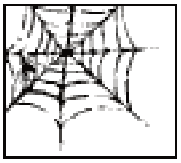
Marshwood High School students Mark Hunter and Aimee Pratt enter their monitoring data into a computer.

As you can see, many things need to be considered when designing a monitoring data management system. First and foremost, you must know what you want to get out of the system before you can decide what design features you need to build into it. If you don't use good planning in the beginning, you could find yourself going back months or years later and repeating all your efforts. So before you buy any software, take the time to assess your needs thoroughly, to ask questions of other monitoring groups, and to consider how to build flexibility into your system. It will be time well spent.

Note: If you are an environmental organization that is local, state, or regional in scope, is organized on a permanent basis, and has 501(c)(3) status, you can get Lotus software (including 1-2-3) from the Environmental Support Center for a modest shipping and handling charge. For more information, contact ESC at 1825 Connecticut Ave. NW, Suite 220, Washington DC 20009; 202/328-7813.

Fred Lease *is the Director of Project Heartbeat for Maryland Save Our Streams, 258 Scotts Manor Dr., Glen Burnie, MD 21061; 800/448-5826.*

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

MURPHY: Software for Data Sharing

by SarahRose Werner

Since early 1992, a coalition of state agencies and volunteer monitoring groups in Maine has been working to develop one good database program that every volunteer monitoring group in the state can and will use. The potential benefits are clear: If the state's 40-plus monitoring groups all use the same software, it will be easy for groups to share data and, ultimately, for data from all groups to be merged into one statewide database. The challenge has been to create a program that is friendly to nontechnical users and capable of running on the older computers that many volunteer groups use, yet at the same time compatible with existing, more complex State of Maine data management and GIS systems.

Frustrating Early Efforts

The coalition (the Maine Department of Environmental Protection, the Shore Stewards Program at the State Planning Office, the Casco Bay Estuary Project, the University of Maine Cooperative Extension Service, and Friends of Casco Bay) first tried hiring a programming consultant to create a program from scratch using the database software FoxPro, which runs on both MS-DOS and Macintosh computers. This plan fell through, largely due to communication problems with the programmer, who was not always familiar with specific monitoring procedures and who did not work closely with the people managing the data on a day-to-day basis. Over the next two years, the program went through three successive versions - each incompatible with the last. It also became clear that the program had not been structured to allow data to be transferred to and from floppy disks, an important feature for sharing data between groups. In the meantime, data sheets were piling up. Friends of Casco Bay alone filled two file drawers with data sheets. People who asked to use our data were understandably put off by the idea of digging through the drawers. When the data were needed for a study on fish habitats, we used up two reams of paper copying the accumulated data sheets. Other monitoring groups kept their data on paper or developed computerized systems using a variety of spreadsheet, database, word processing, and graphing software programs. None of these systems were designed to be compatible with each other.

Cit/Mon*MAN to the Rescue

In early 1994, Friends of Casco Bay began investigating a program called Cit/Mon*MAN, originally developed to manage volunteer monitoring databases for the Alliance for the Chesapeake Bay and later adopted by groups in Rhode Island. Cit/Mon*MAN has many attractive features: It's easy to learn and use. It protects the quality of the database by catching many kinds of data entry errors. It produces a variety of reports and graphs. It allows the user to transfer data to and from floppy disks. It's been well tested by a variety of users.

Cit/Mon*MAN is based on an older version of Paradox (3.5 for DOS), a useful feature for volunteer groups that often depend on donated equipment and don't have the latest in computer technology. The hardware requirements for Cit/Mon*MAN are relatively modest: an IBM compatible computer (PC, XT, or higher) with a hard drive, one floppy drive, 512K RAM, and 4M of space free on the hard drive. One limitation is that Paradox does not have a Macintosh version (however, it does export to a variety of spreadsheet applications which could provide a link to a Macintosh database).

The coalition was impressed with Cit/Mon*MAN and decided that with some fine-tuning it would nicely meet our needs. With the permission of the original programmer, Frederick Hoffman, Friends of Casco Bay staff spent the summer and fall of 1994 working on modifications and writing a 45-page manual. New features written into the program included:

- 60°/60° F hydrometer conversion tables for calculating salinity data (rather than the old 15°/4° C tables), and the ability to accept salinity data measured using a refractometer or salinity-conductivity-temperature meter.

- Corrections for the salt effect on pH values measured using cresol red.

- Statistical modifications such as calculation of sample standard deviations rather than population standard deviations, and calculation of average pH as the negative log of the average H+ activity.

MURPHY is Released

The original Murphy is the office mascot for Friends of Casco Bay. Both Murphy and MURPHY are extremely user-friendly and run well.

In December 1994, the new program was released with the temporary working name MURPHY. Copies of MURPHY have been distributed to monitoring groups throughout Maine and in New Hampshire, Massachusetts, and New Brunswick.

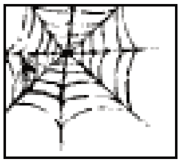
A second round of modifications is planned for 1995. Among other things, MURPHY needs to be able to handle water column profile data. Each Friends of Casco Bay profile data sheet contains a single set of general conditions (e.g., date and time, weather) along with several sets of readings (water temperature, salinity, and dissolved oxygen) taken at successively lower depths in the water column. MURPHY needs to be able to store the general conditions in one file and relate them to the depth-specific readings stored in a second file. Reports and graphs need to draw on data from both files.

MURPHY will also be modified to provide an import/export utility for moving data to and from other software applications. One important goal is to enable MURPHY to communicate with Alice, a software program being used to link math and science classes throughout Maine. (For more on the Alice network, see "Using Data in the Classroom.")

Developing MURPHY ourselves has certainly required a sizable commitment of staff time and resources. On the other hand, it has allowed us to get to know the program structure well, to respond to the specific needs of Maine monitoring groups, and to realistically envision possibilities for future modification and expansion.

*For more information on MURPHY, contact SarahRose Werner, Friends of Casco Bay, 2 Fort Road, South Portland, ME 04106; 207/799-8574. For more information on Cit/Mon*MAN, contact Frederick Hoffman, Hoffman Ecological Consulting, 2604 Darnell Road, Richmond, VA 23294; 804/527-0610.*

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995

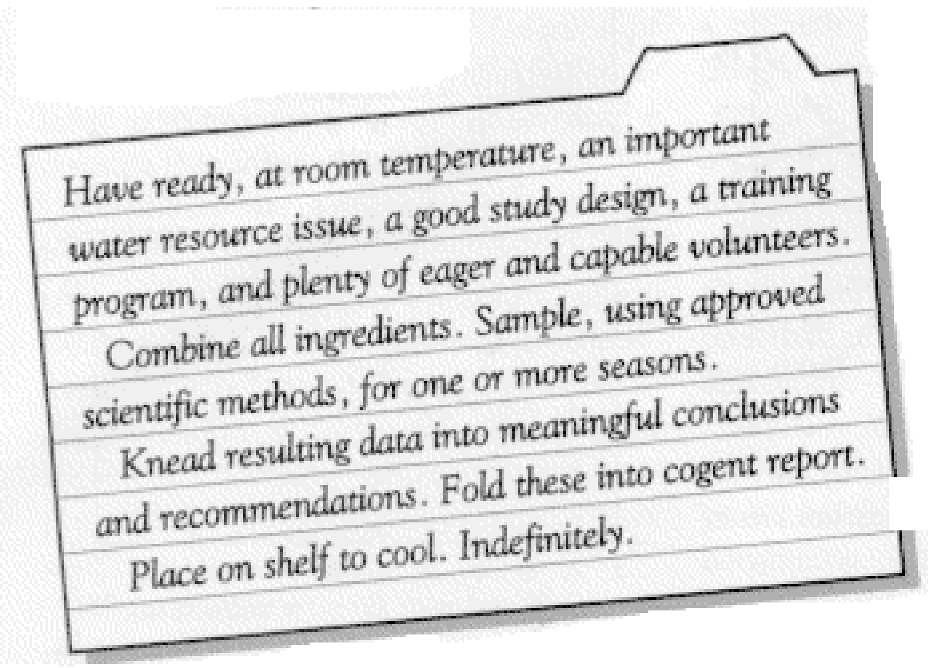


Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Beyond Reports: Packaging Data Creatively

by Jerry Schoen and Marie-Francoise Walk

Here's a recipe for a failed volunteer monitoring program:



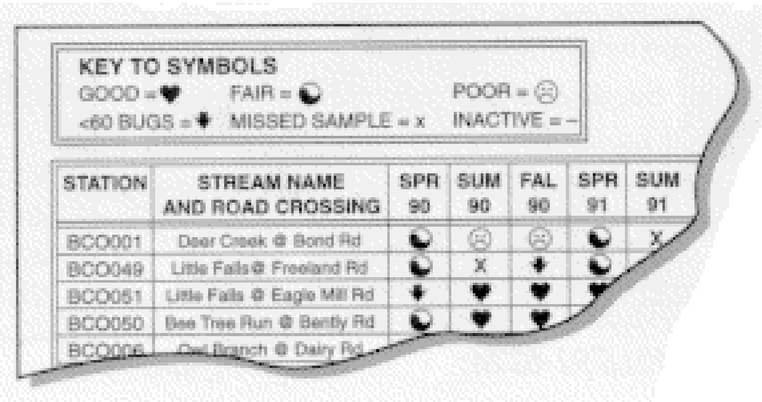
In our experience, most groups pay far more attention to getting good data than to using their data to enhance their environmental programs. This oversight shortchanges not only the volunteers who produce the information but also the people who have a use for it. And that's more people than you might think. Everyone today--backhoe operators, bureaucrats, and biologists--makes decisions that affect the environment. If you have information that can help them make better decisions, don't hoard it. Share it with them.

But don't wait for the stampede to grab your product off the shelf. Get out there and market it. How do you sell data? First, decide who the customer is--the backhoe operator, the biologist, or both? Then consider their tastes. Envision your intended audience staring at a copy of your 3-inch-thick report. Are they smiling? If not, maybe you need to garnish up the data with an appealing presentation to whet their appetite. Let's take a look at some of the components of a good data presentation strategy.

1. Assemble the Ingredients

The basic ingredients you'll use are tables, graphs, maps, and photos.

Tables aren't known for pizzazz--but they're an indispensable step in organizing your data. And they can be dressed up. Maryland Save Our Streams makes clever use of symbols to compress data into a visually appealing table (see example below).



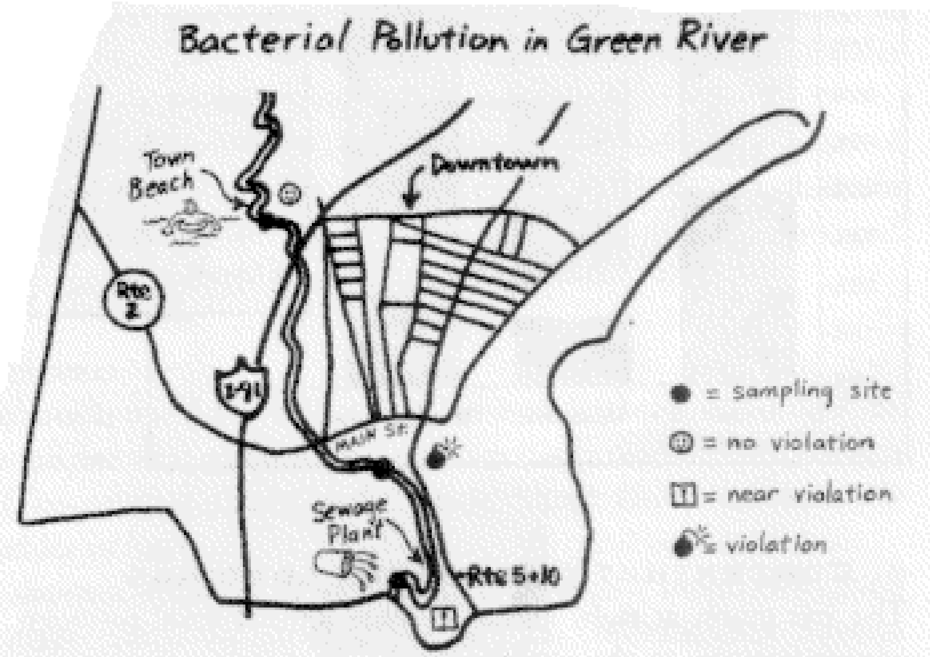
KEY TO SYMBOLS
GOOD = ♥ FAIR = ◐ POOR = ☹
<60 BUGS = ♦ MISSED SAMPLE = x INACTIVE = –

STATION	STREAM NAME AND ROAD CROSSING	SPR 90	SUM 90	FAL 90	SPR 91	SUM 91
BCO001	Deer Creek @ Bond Rd	◐	☹	☹	◐	x
BCO049	Little Falls @ Freeland Rd	◐	x	♦	◐	
BCO051	Little Falls @ Eagle Mill Rd	♦	♥	♥	♥	
BCO050	Bee Tree Run @ Bently Rd	◐	♥	♥		
BCO006	Owl Branch @ Dairy Rd					

Charts and graphs are excellent for "boiling down" your data and turning it into a visual image. See "Using Graphs to Tell Your Story" for tips and ideas.

Maps should accompany every report or presentation. Show your monitoring locations in relation to familiar landmarks to help the audience feel connected to your study. Include watershed boundaries and shoreline features to highlight the relationship between the water body and surrounding land uses. Do you want to make people think hard about where those high bacteria levels are coming from? Portray your coliform data on a map that also shows locations of dairy farms.

Color maps make a particularly strong impact. For example, document thermal impacts of deforestation with different stream reaches colored blue, green, yellow, and red. For newspapers that require black and white, use simple patterns or shades of gray. Symbols are especially effective (see illustration below).



High-tech computer-generated maps are nice, but a hand-drawn one can be just as effective.

Photos: We can't understand why more monitoring groups don't use photography in their data presentations. Enlist the amateur photographers among your volunteers to photograph your program in action. Put together a collection of black-and-white prints for newspapers, color prints to hand to your legislator or use in your poster display, and color slides for the Lions Club talk.

Photos are great "message enhancers." That map showing fecal coliform data in relation to dairy farms will pack an even bigger punch if you add a photo of a cow making a bank deposit.

Sometimes photos themselves provide all the data you need. For example, in an upcoming Massachusetts Water Watch Partnership (MassWWP) study, photos will serve as a visual measurement of percent exposed stream channel - an important indicator of fish habitat quality.

2. Create the Products

Now you can take the basic ingredients and creatively package them to suit the tastes of most any audience.

Reports are the staple of your data presentation strategy. A technical report aimed at a government agency, court, or other "official" audience should establish your credibility by containing the traditional elements of a scientific paper - methods, results, conclusions, etc. However, technical data reports have probably cured more cases of insomnia than Sominex. For a lay audience, substitute a friendlier version that emphasizes data summaries, graphs, maps, and photos.

Put your results on a T-shirt and they'll travel everywhere.

Posters let you combine text, maps, charts, and photos into portable, eye-catching displays. Since they're easy to update, posters keep your message fresh. Using Velcro and foam-board backing, you can rearrange your material to target different audiences, feature a "parameter of the week," or add the latest Secchi disk reading.

A good basic display includes one panel with background information, another showing volunteers gathering data, and a third displaying results and conclusions.

Slide shows, the old standby, are still hard to top. Many groups already have a "stock" show that advertises their organization. All it takes to turn that into a data presentation are a few extra slides focusing on your results.

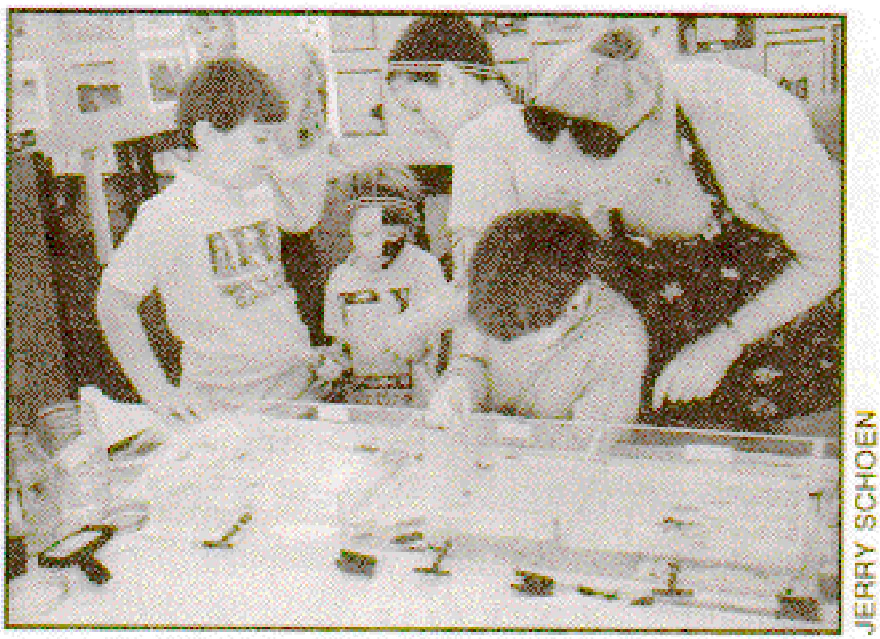
When creating slides, respect the nature of the medium. Minimize text. Give your audience a visual treat with colorful graphs, maps, and charts. Intersperse data slides with photos of the resource. Make the show dynamic - for instance, sequential shots of a map of your lake showing increasing milfoil coverage will demonstrate how the problem has spread out from boat launch areas.

Videos can really pack a wallop. Incorporating sound and motion, they're the closest thing to being there. When using video to show your data, avoid "dead shots" (i.e., static footage of a data table or chart). Liven things up by filming one of your volunteers explaining the data in the style of a TV weather reporter (hopefully with better jokes).

A good resource for video production is your community's public access cable TV station. You'll get equipment and training for free or a small fee. Ask for a list of people who produce local shows - they may be able to work with you to create a program, again for little or no expense.

Once you complete a video, it can be shown again and again. Donate copies to your public and school libraries, and circulate others to environmental and civic groups.

Physical props can bring all the senses into play as audiences heft, sniff, and scrutinize plants, bugs, rocks and other specimens. To demonstrate the connection between substrate composition and aquatic insect populations, the Hoosic River Watershed Association brings samples of cobbles, gravel, and sand from various sampling locations to its Riverfest. MassWWP lends out a "Bugquarium" (see photo below) that monitoring groups can use to display live macroinvertebrates collected from three sampling sites (clean, polluted, recovering).



Fairgoers check out macroinvertebrates in the "Bugquarium."

3. Deliver the message

Remember what your data have in common with pizza and the daily newspaper: most people want them delivered. Where should you serve your various offerings? That's something to think about before you start cooking, because no one product is appropriate for every situation. Here are a few suggestions:

News media. When you prepare materials for the media, assume the audience is environmentally ignorant. Avoid jargon, and keep everything as simple as you can. Usually smaller local outlets are more interested in your information than the large outfits. The local newspaper might welcome a regular column reporting your data.

Forums bring you face to face with your audience. This category includes:

- *Press conference:* Call one when you have big news, such as the illegal dump you've discovered that's leaking PCBs in your stream. Give a short presentation with photos and maybe a good graph. Have your raw data ready for scrutiny.

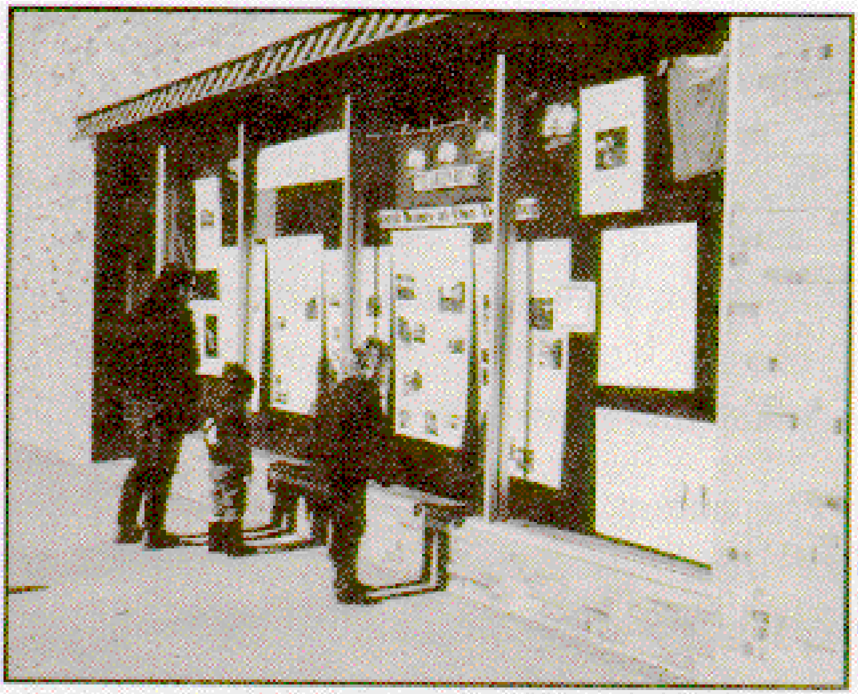
- *Testimony at a hearing:* Here's where your data convinces County Commissioners that the proposed development on your lake's shore will have to hook up to the sewer system. At hearings, you must follow their rules, so check ahead to see what visuals you can use and how much time you'll have. With the right props, you could be the star who wakes people up after a parade of droning speakers.

- *Your own public forum:* For complete freedom of expression, rent a hall and run your own show.

- *Dinner circuit:* Civic groups (Lions, Rotary, Garden Club) are always looking for speakers for dinner meetings. They expect to be entertained, so slides are good. Keep your talk short and lively, and tailor it to the group's specific interest - whether it's boating, birding, or keeping their taxes low.

Fairs and festivals. When you bring your exhibit to a river festival, lake fair, or Earth Day event, you'll be pitching to a sauntering crowd. Make your display bright and colorful, or draw people in with something interactive or unusual, like the Bugquarium.

Once the event is over, don't mothball your display. It can be re-used year-round in schools, libraries, museums, and even empty storefronts (see photo below).



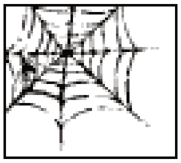
MARIE-FRANCOISE WALK

By "recycling" their poster display into an empty storefront window, the Green River Watershed Alliance reached a whole new audience.

No action will result from your hard work if nobody hears about it. So put on your apron and get creative. And if you've got a data presentation trick you're especially proud of, let us know. MassWWP is compiling a handbook of successful techniques, and we're always looking for new ideas.

Jerry Schoen is Statewide Coordinator and Marie-Françoise Walk is Special Projects Coordinator for Massachusetts Water Watch Partnership. Both may be reached at MassWWP, Blaisdell House, University of Massachusetts, Amherst, MA 01003; 413/545-2842.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Interpreting Your Data

by Geoff Dates

Eyes glazed, you're staring at pages full of numbers, tables, and graphs and wondering: "What does it all mean?" Unfortunately, there's no "interpretation" key on any keyboard I've ever seen that will spit out a cogent, clear, concise analysis of monitoring data. That's still up to us humans.

The purpose of this article is to get you pointed in the right direction, so that you can begin to solve your own data interpretation mystery. Though statistical analysis is one useful tool, the emphasis here is on a common sense, yet systematic, "low tech" approach.

Let's illustrate this approach by considering a hypothetical example. Suppose you are monitoring a river and your study is designed to answer the following basic question:

> *Where does the water meet the numerical criteria in the state water quality standards for bacteria, temperature, dissolved oxygen, turbidity, pH, and benthic macroinvertebrates?*

(The numerical criteria in the water quality standards define acceptable levels of these water quality indicators that will support "designated uses," such swimming or aquatic life.) Note that the study question implies a related question: *If the water does not meet the standards, why doesn't it?*

For simplicity, let's assume that you've collected data for at least one season and that you looked for the right things, in the right way, in the right places, and at the right times. You used appropriate analytical methods. Your monitoring sites bracketed all the known sources of pollution. You monitored during both wet and dry conditions. Your quality assurance/quality control results met your data quality objectives. Finally, let's assume that you've summarized your data in tables and charts so you can view the entire data set. Now you're ready to take a closer look at your data and start figuring out what it means.

Think of data interpretation as a process in which you ask a series of questions that lead you to findings and conclusions. This article divides the questions you should ask into two basic categories:

1. Questions that lead to findings. (For the purposes of this article, findings are defined as objective

observations about your data.) These questions prompt you to look at your data systematically.

2. Questions that lead to conclusions. Conclusions are your explanations of why the data look the way they do. These questions lead you to examine your data critically and help you determine whether and how you might answer your study question.

Questions that lead to findings

Given the basic study question posed above, your findings will be a series of observations about how your results compare with the numerical criteria in the water quality standards. Here are some questions that will help you arrive at findings:

- Which sites consistently did not meet the water quality criteria? By how much? ("Consistently" might mean more than 50 percent of the time.)
- Were there sampling dates on which most or all sites did not meet the criteria? What were the flow and rainfall like on those dates?
- Do levels increase or decrease in a consistent manner upstream to downstream?
- If you are monitoring the impact of a pollution source, are the results different above and below the impact?
- Do changes in one indicator coincide with changes in another? For example, there is frequently an inverse relationship between water temperature and dissolved oxygen, since warm water can hold less oxygen than cold water. Your results might show that when temperature was high, DO was low.

Questions that lead to conclusions

Now that the data are organized into findings, it's time to see if you can answer your study question. First you want to determine whether your monitoring sites appear to support their designated uses. Consider questions like:

- Do bacteria results show safe levels for swimming at town beaches?
- Do temperature and dissolved oxygen levels support cold water fish species at nursery and spawning areas?
- Do benthic macroinvertebrate results show impairment in the aquatic ecosystem at riffle habitats?

After answering questions such as these, you may conclude that all the criteria in the standards are being met and all the designated uses are supported at all sites on all sampling dates. If so, your work is done. You've answered your basic study question.

At the opposite extreme, perhaps you conclude that all sites on all dates did not support their designated uses. More likely, your conclusion will be more ambiguous: some criteria are not met at some sites on some dates. In either case, you'll want to dig deeper. You'll want to ask, What could explain the failure to meet the criteria? Now you've come to the hard part. This is where your judgments and opinions are brought to bear.

Basically, the possible explanations fall into three categories: natural conditions, human alterations of the aquatic system, and errors in sampling and/or analysis.

Let's start by looking at some questions that would help you decide whether human alterations or natural conditions might explain your results.

- Might natural upstream-to-downstream changes in the river account for your results? Your benthic macroinvertebrate results might be explained by natural shifts in the macroinvertebrate community composition from headwaters to mouth. The increased turbidity levels observed at your downstream sites might be due to a change in river bank soil types and natural erosion, rather than human activities.

- Does weather appear to influence your results? For example, do problem levels coincide with intense rainstorms? Might elevated temperature levels be caused by unusually hot weather?

- Do problem levels coincide with rising flow? Consider the impacts of dam releases and other flow management activities, as well as rainfall.

- Does the presence of specific sources explain your results? For example, can you attribute increased bacteria levels to a wastewater treatment plant or a failing septic system?

- Do changes in one of your indicators appear to explain changes in another? For example, could low dissolved oxygen be explained by high temperatures (which might have been caused by a thermal discharge or a heat wave)?

- Do your visual observations explain any of your results? Did your samplers report any strange pipes, eroding banks, or dry weather seeps from storm drain pipes?

- For multiple years of data, are there overall trends? For example, did the benthic macroinvertebrate community improve or deteriorate over time? The former could be explained by improved pollution control; the latter, by new pollution sources.

- If you are monitoring the impact of a pollution source, are there other upstream impacts that might be influencing and confusing your results? For example, if there's a dairy farm immediately upstream from the wastewater treatment plant you're monitoring, it might be difficult to figure out which source is causing any impairments you detect.

Frequently, your results can be explained by a combination of natural and human-caused phenomena. For example, rainfall causes surface runoff, which may carry human-caused contaminants to surface water.

Don't overlook the possibility that the explanation of your results lies in the way you collected and analyzed samples, rather than in the river conditions themselves. Consider the following questions:

- Could flaws in your field and/or laboratory techniques explain your results? Were problem levels due to sample contamination or sampling error?

- For episodic discharges, did your sampling coincide with the discharge? For example, did you catch the storm-related polluted runoff you were trying to analyze? Double-check to confirm that the flow was rising when you collected your samples.

- Was your analytical method sensitive enough to detect levels of concern? Is the medium you sampled the only place in the aquatic ecosystem where the indicator is likely to occur? For example, if you sampled the water column and found low levels of phosphorus, you shouldn't necessarily conclude that there isn't a problem. There might be high levels of phosphorus in the bottom sediments, or in use by aquatic plants. To investigate this possibility, check your samplers' field notes to see if they recorded evidence of nutrient enrichment, such as algae blooms.

- Did the time of day you sampled affect your results? For example, dissolved oxygen is typically lowest in the morning and highest in the late afternoon. If you were looking for worst-case conditions (low DO) and samples were collected in the afternoon, you could have missed the problem.

From the facts to the story

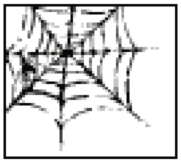
What if you go through the whole data interpretation process, and at the end you haven't solved the whole mystery? What if you haven't solved any of it? Don't worry; that's not unusual. Professional scientists have the same problem. Your data may be inconclusive, especially if you've only been monitoring for a single season. Don't go out on a limb if you're uncertain. It's perfectly valid to conclude that you need additional information in order to understand a system as complicated as a river. Maybe you need to conduct a shoreline survey or habitat assessment, or monitor additional indicators or sites.

Remember, too, that interpreting your data is not a process that you need to (or should) go through by yourself. Make it a collaborative effort. This is a great opportunity to involve your monitors, data users, resource people, stakeholders, and the general public in your program.

In the data interpretation process you move from "just the facts" to the story you think these facts tell. While your conclusions should be supported by your data, they are subjective--they are your judgments. Others may disagree with your interpretation, but as long as your data support it, and you've followed a logical process in collaboration with knowledgeable resource people, you should be able to defend it.

Geoff Dates is Science Coordinator for River Watch Network, 153 State St., Montpelier, VT 05602; 802/223-3840.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Using Data in the Classroom

by Joyce Tugel

As volunteer monitors and teachers know, science is much more than a body of isolated facts. The education restructuring movement is focusing on science as a learning process that incorporates creative inquiry and critical thinking. Goals for students include such skills as data analysis, problem solving, communication, and collaboration.

A water monitoring project can be an ideal way to meet these goals. Yet as I read the Spring 1993 issue of *The Volunteer Monitor* (on "School-Based Monitoring"), I felt a little overwhelmed. Many of the teachers profiled in that issue apparently had been able to develop a whole course structured around a water quality monitoring project. I couldn't help wondering, "Where do they find the time?"

Classroom Reality

As a teacher of high school chemistry, I am pulled in many directions. While most school districts support curriculum reform, they are adamant that students receive traditional content as well. With the science information base exploding into unwieldy proportions, I never seem to have enough time to do all the things I need and want to do.

If you are a teacher struggling with similar problems, take heart. Even if you can't regularly take your students out to a monitoring site, there are other ways to bring the advantages of monitoring into your classroom.

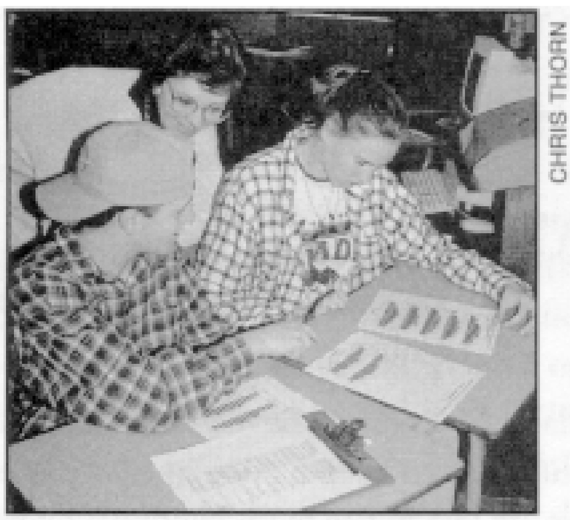
One alternative is to perform water quality testing as an extracurricular activity with a small group of interested students. Each year for the past three years, I have trained about 20 of my students to monitor a site on a local river. Twice a month, on our own time, we test for pH, temperature, salinity, dissolved oxygen, turbidity, and fecal coliform bacteria. The whole class benefits: The monitors demonstrate their techniques to their classroom peers, and I incorporate the science behind the analyses into classroom lessons.

An even easier option is to ask local monitoring groups to provide you with a copy of their databases.

Chances are they'll be delighted to oblige, and your class will gain access to real data that has meaning to their community.

Using Monitoring Databases to Teach Math and Science

Many students enter the classroom unable to identify dependent variables, independent variables, or constants, and needing practice constructing tables and graphs. Monitoring data sets provide the perfect tool for teaching these concepts in a meaningful way. What's more, working with monitoring data encourages students to learn the skills of true investigation. Interpreting the data leads to questions: Why is the salinity different at different sites? Why is the pH constant? Detective work may be needed to find the answers, and thus science is learned on a need-to-know basis. While memorized definitions may be soon forgotten, the outcomes of observing, forming hypothesizing about, and testing a real database will be remembered long after.



Ryan Donn, Joyce Tugel, and Michele Henningsen compare temperature graphs for different monitoring sites.

Computers and Monitoring: Beyond Word Processing

The technology revolution has touched our nation's schools at all grade levels. Many schools are establishing "computer labs." Too often, though, computer technology is thought of as a subject in itself, whose goal is to teach students word processing and perhaps a graphics or spreadsheet program. Students may create documents for fictitious businesses and produce professional-looking reports - but they don't understand the ways computers can be used as analytical tools in many disciplines.

Working with monitoring data introduces students to the real-life applications of computers. After I taught my students how to use spreadsheet and graphing programs on the computer in my classroom, they entered data from our water monitoring project, then created graphs of their own choosing. I heard laughter and excitement as the students found patterns and inconsistencies in their data. Before graphing, the columns of numbers had had little meaning, but now students were questioning values that didn't fit trends and looking for possible sources of error. One salinity value looked suspiciously low until they checked the field data sheet and found that it had rained the night before sampling. The computer became a tool for students to ask and answer their own questions.

My students and I will soon be using computers to examine databases from other volunteer monitoring groups. I was recently accepted into the Telecommunications in Education to Reform Mathematics and Science (TERMS) project, sponsored by the Maine Mathematics and Science Alliance and the Wells National Estuarine Research Reserve. Through the TERMS hub at the Wells Reserve, my class will exchange water monitoring data with schools across the state, using Alice network software developed by TERC. (The name 'Alice' is a reference to Lewis Carroll's character: We're not sure where we're headed on the information superhighway, but it will certainly be an adventure!) We will be able to compare test results from coastal and freshwater sites, from north to south, and from different years, using the graphing and telecommunications capabilities of the Alice software.

As teachers with lots of ambition but little time, it's important to remember that we can't reform our entire curriculum at once. But we can use water monitoring databases to enrich our courses and bring science to life.

Joyce Tugel *teaches chemistry at Marshwood High School, 204 Dow Highway, Eliot, ME 03903; 207/439-5600.*

For more information about the TERMS project, contact Barbara Williams at the Wells National Estuarine Reserve, RR #2, Box 806, Wells, ME 04909; 207/646-1555.

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995

Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

Using Graphs to Tell Your Story

by Meg Kerr

Somewhere in that pile of computer printouts covered with columns of numbers, a story is buried. Your challenge is to transform those numbers into a message that inspires people to action. Graphs are one of the most valuable tools for telling your story, particularly if you're trying to convey trends and relationships. Because graphs present data in a "coded" format, a graph can compress reams of information into a simple visual image. But there's a danger here - if people can't decipher the code easily and correctly, they won't get your message.

What makes a good graph? The same qualities that make a good story. It should be easy to follow. It shouldn't be cluttered up with extraneous information. Most important, it should have a point. Don't get carried away with all the fancy options in your computer's graphing program. The best graphs are simple in design and have a limited number of elements.

It's worth taking some time to make graphs pretty. Graphing software defaults may be all right when you're making quick preliminary graphs for your own use, but for a poster or presentation strive for something a little more artistic.

As a monitoring group, you also want people to believe that your story is true. A sloppy or confusing graph will make people wonder if you know what you're talking about. And a misleading one - for example, one that unrealistically exaggerates small differences - makes you look untrustworthy.

If you're just starting out, remember that learning how to use graphing software programs for the first time can be very time-consuming. Also, not every software program is capable of producing the finished graph you want.

What Type of Graph?

Line graphs, bar graphs, and pie charts are the three main types of graphs a monitoring group will use.

Line graphs are good for emphasizing the *relationships* between data points - for example, changes in conditions over time or space - and can often illuminate trends in data. They usually display time or space along the x- axis (horizontal) and water quality parameters along the y-axis (vertical).

Bar graphs put more emphasis on the individual points. They are useful for comparing the level of a pollutant at one station over time or at several stations at one time, and for displaying summarized data.

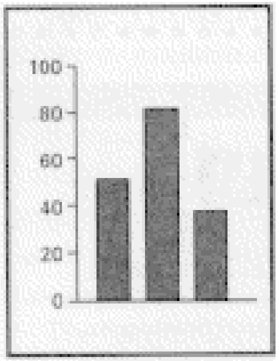
Pie charts use a segmented circle or "sliced pie" to display the relative abundance of various components of the whole. They're easy for the general public to understand, but can only be used for data that can be expressed in terms of proportions, or percentages, of a whole. Some types of data that work well in pie charts are land use (acres of forest, wetland, etc. in a watershed), populations (numbers of trout, carp, etc. at a station), and pollutant loadings (see example 4).

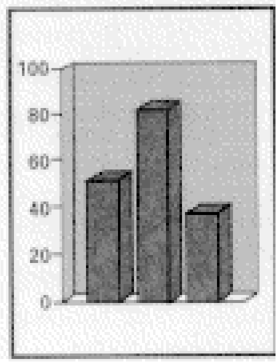
Stacked bar graphs provide another way to show data as proportions of a whole. They're especially useful when you want to show comparisons between several similar "stacks" (see example 3).

The following examples show some of the different types of graphs, and illustrate some important graph-making do's and don'ts.

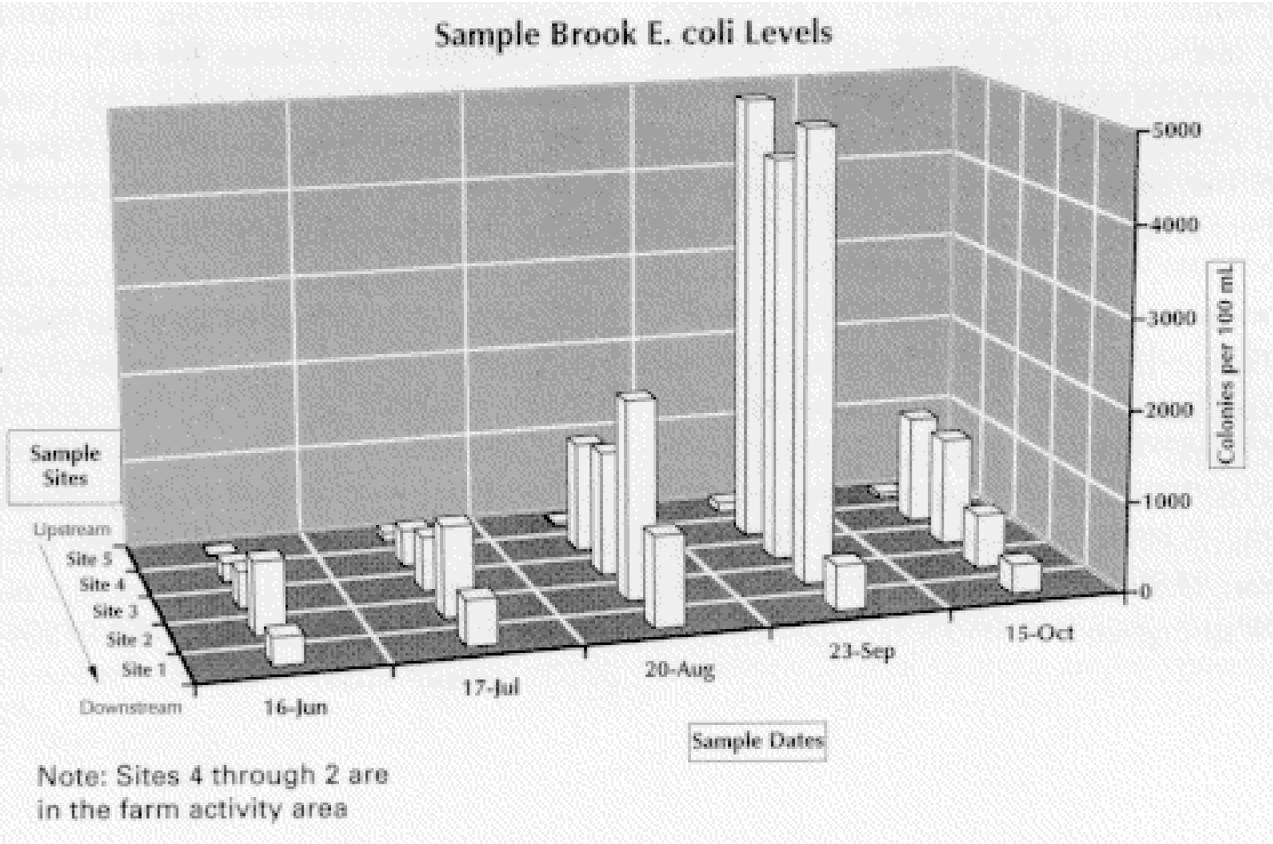
1. When to Use 3-D

The two charts below show the same data. Although the 3-D version is more eye-catching and might be preferable for a poster or slide show, the flat version is better if you want people to actually read the percentages off the graph. It's hard to line up the tops of the 3-D bars with the scale on the y-axis. Moral: Don't automatically make every bar graph 3-D just because it's easy and looks fancy.



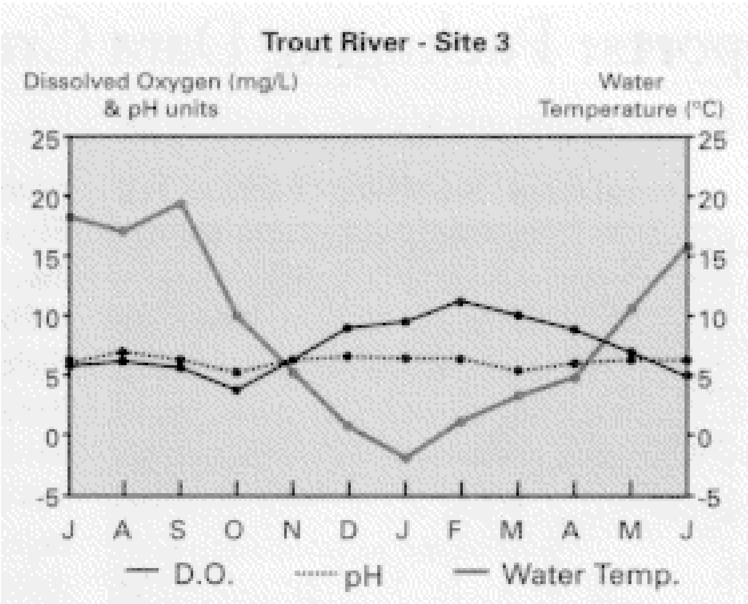


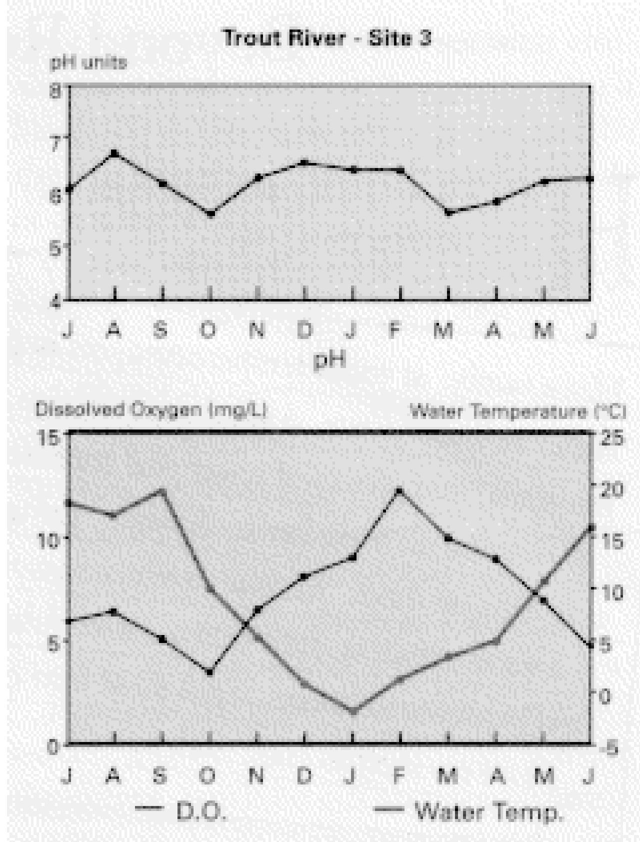
The chart below, on the other hand, shows a situation in which 3-D is genuinely useful in helping you visualize the data set as a whole. This one graph shows an entire season of fecal coliform data. Several separate 2-dimensional graphs would be required to display the same information. Although you have to study this graph for a few minutes to really "get it," your efforts will be rewarded because the graph clearly conveys an important message - namely, that the farm has a clear and consistent impact on bacterial levels. (You can see other patterns too - for instance, counts are highest in September. To help interpret this pattern you would need more information, such as weather and flow data.)



2. Use the Appropriate Scale

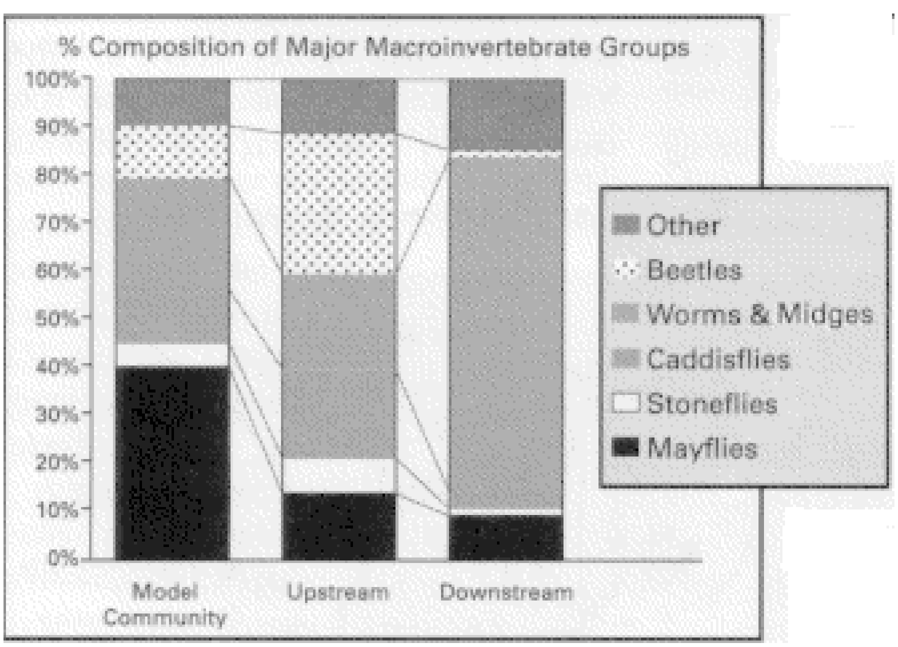
It's tempting, and sometimes useful, to plot more than one parameter on the same graph for easy comparison - but be sure the scale on the y-axis makes sense for all the parameters. The graph at the near right (based on an actual graph from a volunteer monitoring group) has some problems. Look at the line for pH - what does it tell you? Not much. Changes in pH have been obscured by plotting pH on a scale that runs from -5 to 25. The line for DO is also somewhat flattened. The solution (right-side figure below) is to use two graphs stacked one on top of the other. Now each parameter has a scale that makes sense.





3. Stacked Bar Graphs

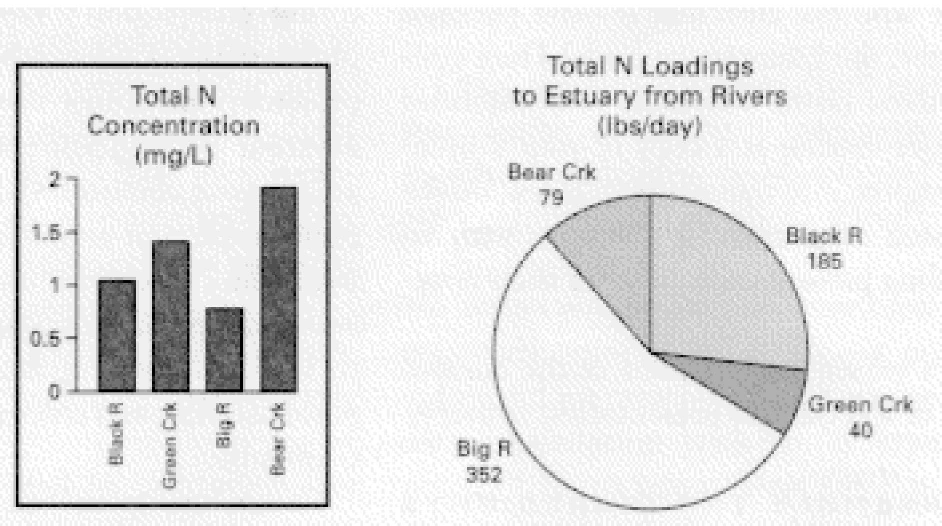
A stacked bar graph can be used to compare percent compositions at several sites - in this case, macroinvertebrate counts at a reference or control site, a site upstream from a pollution source, and a site downstream from the source. A series of pie charts would not have done the job quite as well, because it's harder for viewers to directly compare areas of pie slices than heights of stacked bars.



This graph requires some sophistication to interpret. The viewer needs to know, or be told in a caption, that most mayflies, stoneflies, and caddisflies are pollution sensitive, most beetles are moderately sensitive, and most worms and midges are pollution tolerant. With this information, the story becomes clear. Because the downstream site contains relatively more pollution-tolerant organisms and relatively fewer pollution-sensitive organisms, we can conclude that the pollution source is having an impact on the stream.

4. Concentrations Versus Loadings

The bar graph below (in first figure) shows average annual total nitrogen concentrations at the mouths of the four rivers that empty into an estuary, and the pie chart (in first figure) shows the average daily load of nitrogen contributed to the bay by each river. Loading is calculated by multiplying instream total nitrogen by stream flow. Concentrations can't be shown in a pie chart (because they are not additive), but loadings can. A table showing the loading calculations helps readers interpret the graphs.



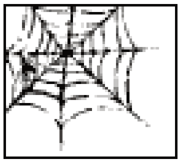
Each of these graphs tells a different piece of the story. The bar chart tells you about conditions in the rivers themselves, while the pie chart tells you which rivers have the greatest impact on the estuary. Taken together, they make an interesting point: Big River is contributing the most nitrogen to the estuary even though it has the lowest nitrogen concentration. So if you're interested in lowering nitrogen concentration in the estuary, you'd probably want to target your efforts at Big River. But if you're more concerned about levels in the rivers themselves, you would target Bear Creek.

	Black River	Green Creek	Big River	Bear Creek
N conc (mg/L)	1.0	1.4	0.75	1.9
Flow (cfs)*	14.3	2.2	36.3	3.2
Loading (lbs/day)	185	40	352	79

*cfs = cubic ft/sec

Meg Kerr *is a Marine Research Associate at Coastal Resources Center, University of Rhode Island, Narragansett, RI 02882-1197; 401/792-6224.*

The Volunteer Monitor, Vol. 7, No. 1, Spring 1995



Note: This information is provided for reference purposes only. Although the information provided here was accurate and current when first created, it is now outdated.

"Variability Happens:" Basic Descriptive Statistics for Volunteer Programs

by Julie Rector

Despite the best of intentions and meticulous care in monitoring, all monitoring data will inevitably have a level of variability and a level of uncertainty. Some variability will result from field or analytical procedures (this is the "error" we can introduce during any phase of sample handling or analysis), but to a large degree, the natural systems we are monitoring are inherently variable.

Compounding this variability is uncertainty. Uncertainty arises because (1) there is no such thing as a truly exact measurement, and (2) we can't collect samples continuously and forever. Instead, we periodically collect samples to represent an environment that is continuously changing over time and space. And we analyze these periodic samples using methods that have limits in resolution, precision, and accuracy. Dealing with variability and uncertainty is one of the biggest challenges in monitoring design and data analysis.

Statistics has been described as the science of making decisions in the face of uncertainty. We can't get away from uncertainty and variability, but we can use statistics to estimate their contribution to our overall results. With this knowledge, we can make a reasonable decision based on the data.

Why Use Statistics?

Volunteer monitoring programs generally use statistics for three main purposes: (1) to summarize and report monitoring results; (2) to evaluate quality assurance/quality control (QA/QC) data; and (3) to help interpret data and draw conclusions about a monitored environment. "Descriptive statistics" are used for the first two purposes, while "inferential statistics" are used for interpretation.

This article will focus on the uses of basic descriptive statistics. It will not attempt to cover inferential statistics, such as tests of significance and methods for trend detection. However, understanding descriptive statistics is the first step in exploring inferential statistics, since many descriptive statistics are used in equations for calculating inferential statistics. Hopefully some commonly used inferential statistics will be covered in a future article in *The Volunteer Monitor*. (*Editor's note:* Anyone interested in contributing such an article is invited to contact the editor.)

Using Statistics in Reporting Data

Obviously, presenting your entire data set is an ineffective way to communicate your findings. Descriptive statistics let you summarize your data in a useful shorthand way - they "describe" a data set in terms of (1) its central tendency and (2) its distribution. Statistics that describe central tendency are called "measures of central tendency" or "statistics of location," and those that describe distribution are called "measures of variation" or "statistics of dispersion."

Measures of Central Tendency

The central tendency of a data set is described by a value that is "typical" or "representative" of all sample observations. Most often this is the mean or median.

- **Arithmetic mean.**

The arithmetic mean, denoted by the symbol x-bar, is the most common statistic used in reporting data results. (Although many people use "mean" and "average" interchangeably, "average" refers only to the arithmetic mean. The geometric mean, discussed below, is not the same.) To calculate the arithmetic mean, or average, simply add up all the values (results), then divide by the number of values. Mathematically:

x-bar = (sum of values) / (n)

where n = number of values
Since each result is given equal weight, the arithmetic mean is affected by outliers (unusually high or low results).

What is the minimum number of values you need to calculate a useful arithmetic mean? As a general rule, when results are expected to be similar (for example, results from lab check standards), a minimum of three results should be used. When results are expected to have a wider range of values, or to vary over time, more values are needed to calculate a mean that will be representative.

Averaging pH values. Because a one-unit difference in pH actually reflects a tenfold change in hydrogen ion concentration, calculating the mean of pH values requires additional steps.

pH is defined as the negative logarithm of the hydrogen ion concentration *(see Note 1 below). To average pH values, first convert each pH to its hydrogen ion concentration. For example, for a pH of 6.5, enter negative 6.5 into a hand calculator and press 10x. The resulting value, 0.000000316, is the hydrogen ion concentration. After you've converted all the pH values, average the hydrogen ion concentrations, then convert this average back to a pH value by taking the log and removing the negative sign. (You'll find that this method of averaging pH values makes the most difference when pH is measured to the nearest whole number. If you measure pH to the nearest 0.01 unit, simply averaging the pH values will give almost the same result as averaging the hydrogen ion concentrations.)

- **Geometric mean.**

State water quality standards for bacteria are often expressed in terms of the geometric mean (usually, the geometric mean of a specified number of samples collected within a specified period of time). Therefore, if you perform bacterial testing you may need to calculate the geometric mean for your bacteria counts.

There are two ways to calculate the geometric mean, but the most practical is to (1) take the logarithm of each result; (2) average the logs; and (3) take the antilog of this average. Mathematically:

geometric mean (GM) = antilog (sum of log Y) / (n)

where Y = each result, or count
n = number of results
For example, consider the following set of bacteria counts from a river site:

```
Count (Y)          log Y

    8            0.9031

    2            0.3010

    5            0.6989

  650            0.8129

    3            0.4771

---------------------------------

arithmetic     sum of log Y = 5.193

          mean = 134


GM = antilog (5.193) / (5) = 11
```

By transforming the data to log form, the geometric mean reduces the effect of a few very high values. As a result, the geometric mean will always be lower than the arithmetic mean. (For the above data set, the geometric mean is 11 and the arithmetic mean is 134.) When high values occur rarely, as is typical of bacterial results, the arithmetic mean is not really representative of all the samples.

- **Median.**

The median is the value that divides the distribution into two halves. To find the median, order the results from lowest to highest. The median is the value that has an equal number of values on each side. (One disadvantage to using the median is that ordering the values can be cumbersome if you have a large data set and no computer.)

The numbers below represent the "index of species richness" calculated from macroinvertebrate counts at each of nine sites:

22, 25, 26, 26, 27, 31, 34, 35, 45

The median value is 27. When you have an even number of results, the median is the average of the two central observations. For example, if the data set above had also included a tenth station with a species richness index of 47, the median would be 29 (average of 27 and 31).

The median is not affected by outliers. If we changed the highest value in the above data set from 45 to 60, the median would still be 27. The mean, of course, would change.

For macroinvertebrate data results, the median is commonly reported along with the mean. In the above example for the nine stations, we would report that the mean species richness was 30 and the median was 27. This information alone gives the reader a clue about how the values are distributed--since the mean is greater than the median, there must be at least one higher value that affected the mean (see Figure 1 below).

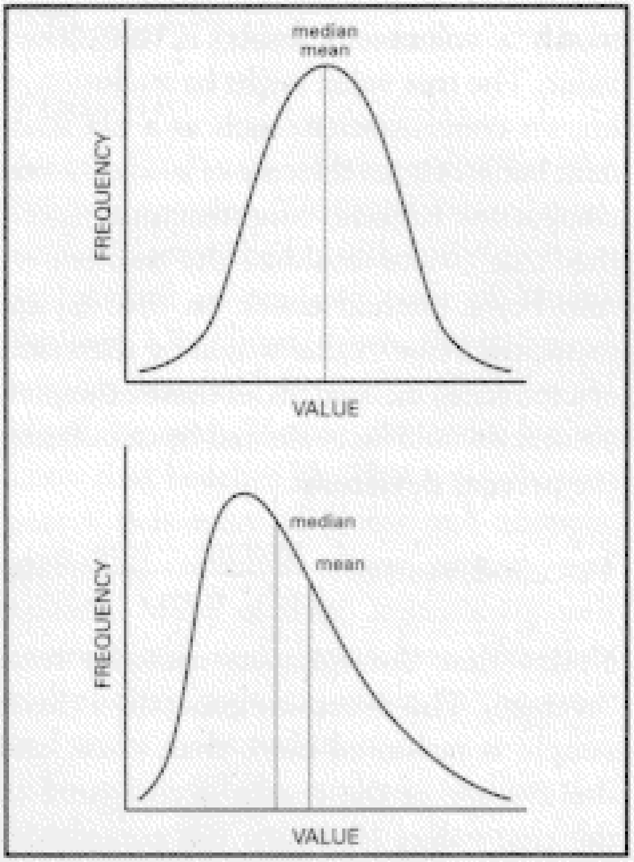


Figure 1. Frequency distribution curves, obtained by plotting the values (x-axis) against the number of times each value occurs (the frequency). Top: Normal distribution curve. When values follow a normal distribution, the mean equals the median, and both are located at the peak of the curve. Bottom: Skewed distribution curve.

Which measure of central tendency should you use?

Deciding which measure of central tendency to report will depend on the type of data collected and the distribution of the results. In general, the arithmetic mean is appropriate when values follow a more-or-less normal distribution with no outliers, and the median is better if the data set is skewed and/or if there are outliers. The only time volunteer monitoring groups need to consider using the geometric mean is when they are doing bacterial testing.

Measures of Variation

Measures of variation, or statistics of dispersion, are used to describe how data are spread out. These statistics are most useful when reported in conjunction with a mean or median.

- **Range.** The range, which statisticians define as the difference between the highest and lowest values, is a relatively crude measure of variation. A small range indicates low variability and, therefore, greater likelihood that the arithmetic mean value is representative of the population. (Note that in data reporting you might say that "results for fecal coliform bacteria ranged from 3 to 65 colonies/100 ml." This is a slightly different use of the word *range*. A statistician would report a range of 62 in this case.)

- **Standard deviation.** The standard deviation (symbolized *s*) describes a population's deviation from the mean. It provides information about the amount of the variability in the results. As illustrated in Figure 2 below, a small standard deviation indicates little variability in the data and a large standard deviation indicates that the samples are more widely spaced.

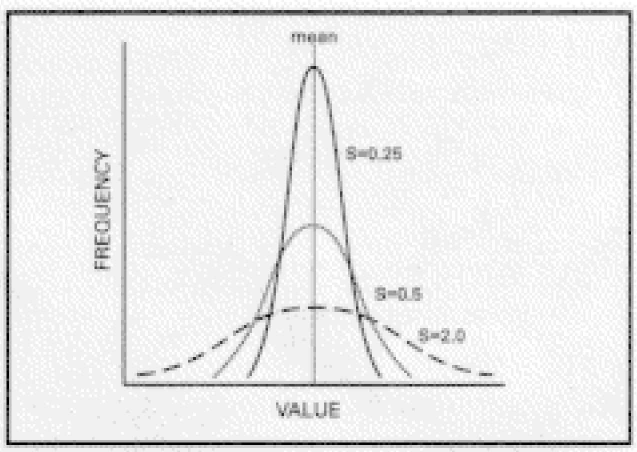


Figure 2. A small standard deviation corresponds to a "peaked" frequency distribution, and a large standard deviation corresponds to a "flat" one.

Most scientific hand calculators can easily calculate standard deviation; you only have to enter your data results, push a button, and *s* appears. Several computer software packages can also calculate standard deviation, but be sure to read the software manual to see whether the program is set to calculate "sample standard deviation" (the statistic you want) or "population standard deviation" (not the one you want) *(see Note 2 below).

Even if you never need to calculate a sample standard deviation by hand, it's useful to take a look at the equation:

s = square root of (sum of (x - x-bar)^2) / (n - 1)

where x= each result
n = number of results
The expression (x - x-bar), called the *deviate*, is simply the difference between each individual result and the mean of the results. Looking at the equation, we can see that standard deviation is strongly affected by variability in the sample results; that is, *s* will be larger if the deviates are large. We can also see that the standard deviation gets smaller if you have more samples (because *n* increases).

When reporting results from a large data set, it's very helpful to report the standard deviation along with the mean and median in order to tell the reader about the variability of the data. For example, if the mean surface nitrate concentration from 32 Puget Sound stations was 0.22 mg/L with a standard deviation of 0.3 mg/L, this tells us, simply and elegantly, that surface nitrate concentrations did not vary much between stations.

Using Statistics in QA/QC

Volunteer monitoring programs also use statistics to evaluate data quality. These QA/QC results are commonly reported along with monitoring results, to show how much variability and uncertainty are associated with the data. This section shows how basic descriptive statistics are used in evaluations of accuracy and precision.

The question of how to determine acceptable limits for accuracy and precision is an article in itself. However, regardless of whether or not you have a formal QA/QC plan, you should evaluate the quality of your data, using the following methods.

Accuracy. Accuracy (sometimes called "bias") describes how close a result is to the true value. It is usually checked by means of various check standards. (*Note:* For more on the types of check samples used in QA/QC, see "The Basics of Quality Control" in the Fall 1992 issue of *The Volunteer Monitor*: softcopy not available)

In evaluating accuracy, you want to show how close a measured value (most commonly a volunteer's result) is to a "true" value. The true value might be a solution of known concentration, such as a pH standard buffer. Or, in the case of a side-by-side comparison between volunteer and trainer, the "true" value would be the trainer's result. For a method check on field equipment, the "true" value would be the result of a reference method. In all these examples, accuracy would be evaluated by calculating the **percent deviation**:

$$\text{Percent deviation} = ((\text{average} - \text{true}) / (\text{true})) * 100$$

Notice that the equation uses the term "average." This is because generally a check sample is measured more than once, and the average of the results is compared to the true value. However, the percent deviation can also be calculated for a single result (just substitute the individual result for "average" in the equation). For instance, let's say your accuracy limit for Secchi readings is plus or minus 6 percent. If a volunteer measures a Secchi depth of 10.3 feet and the trainer measures a depth of 10.9 feet at the same location, the percent deviation is negative 5.5%. The volunteer's result is acceptable. (The negative value shows that the volunteer's result was less than the trainer's.)

Precision. Precision describes how close values are to each other. In QA/QC, precision results can be used to estimate the combined effects of natural variability and introduced error from various stages of sample handling (collection, processing, and analysis). *Total precision* is checked by field duplicates (two separate samples collected using the same methods) and reflects the combined effect from all these sources of variability. *Lab precision* is calculated from the results of lab duplicates (a sample split in two at the lab) and reflects lab error only.

The statistic most commonly used for estimating precision is a measure of variation called the **coefficient of variation (CV)**, sometimes also called the relative standard deviation. The CV is simply the standard deviation expressed as a percentage of the mean:

$$CV = (s / \text{x-bar}) * 100$$

The smaller the CV, the better the precision. Let's say your lake monitoring volunteers test dissolved oxygen at several depths, and on each sampling date they collect and analyze one duplicate sample from a depth randomly selected by you. Your job is to evaluate total precision for each volunteer for each date, and your acceptable precision limit is a CV of 5%. Here are the results for two volunteers on one sampling date:

	Sample result (mg/L)	Field duplicate (mg/L)	x-bar	*s*	CV (%)
Vol. 1	8.6	8.4	8.5	0.14	1.6
Vol. 2	5.6	7.2	6.4	1.13	17.7

Volunteer 2 exceeded the precision limit for the day, so all dissolved oxygen data from this volunteer on this day should not be used.

When samples are analyzed in a lab, it's important to know both total and lab precision. For example, if the CV for lab precision is nearly as high as the CV for total precision, almost all the error is coming from the lab. While some lab error is to be expected, CV values that are close to the limits specified in your QA plan may warrant investigation. On the other hand, if lab precision is good but total variability is high, either natural variability is high or the sampling method is introducing error (or both). Additional training or comparisons with a second sampler may be used to evaluate error due to sampling method. But if you are dealing with a highly variable natural environment, you may need to collect more samples for your program to be able to characterize the environment.

Following Through

What should you do if precision or accuracy limits are exceeded? It's important to have a QA plan that includes guidelines on how to handle monitoring data when precision or accuracy are poor. Depending on how the data are to be used, the guidelines should specify how to flag the data in the database and what restrictions should be placed on the use of flagged data.

More important, you should evaluate the sources of error and make corrections before more data are collected. Accuracy problems can often be corrected by improving technique. Are volunteers measuring sample volumes correctly? Do they know how to calibrate field equipment accurately? Is there any contamination of glassware, equipment, or reagents? Precision problems are harder to investigate. It may be difficult to determine whether the variability is due to field or analytical error, or whether it is simply natural variability.

Finally, keep in mind that statistical evaluation of QA/QC results is not simply a last step that helps you decide when to keep data and when to throw it out. Rather, it should be an ongoing process whose goal is the early detection, and correction, of field or analytical problems.

For More Information

Green, Linda. 1994. "The PARCC' Parameters." In *Proceedings: Fourth National Citizens Volunteer Monitoring Conference*. U.S. EPA. Free. Available from Alice Mayio, Volunteer Monitoring Coordinator, U.S. EPA, 4503F, 401 M St. SW, Washington, DC 20460; 202/260-7018.

Hach Company. "Quality Corners" feature in Hach's newsletter, *News and Notes for the Analyst.* Hach Co., P.O. Box 369, Loveland, CO 80539; 800/227-4224. Free. Concise, basic explanations.

Kelley, Thomas et al. 1992. *Basic Statistics for Laboratories: A Primer for Laboratory Workers.* Van Nostrand Reinhold, New York. ISBN #0-442-00456-7. 175- page book; lots of useful and practical information.

Mattson, Mark. 1992. "The Basics of Quality Control." In *The Volunteer Monitor*, Vol. 4, No. 2 (Fall 1992).

Note 1: For those whose math is rusty: the logarithm, or "log," is the power to which 10 must be raised to produce a given number. The log of 100 is 2. The antilog of 2 is 100.

Note 2: For example, Lotus 1-2-3 will calculate standard deviation using the command @STD(cell range), but what it calculates is the population standard deviation. (The formula for population standard deviation uses n in the denominator instead of n - 1.) To calculate a sample standard deviation in a Lotus spreadsheet, enter the following calculation into the field:

```
(@SQRT(@COUNT(A1..Az) / (@COUNT(A1..Az) - 1)))*@STD(A1..Az)
```

where A1..Az is the range of cells for which you want to calculate a standard deviation. (Note that the variance, s^2, also needs this correction in Lotus. Substitute @VAR for @STD in the equation above.)

Julie Rector *is the Coordinator for Washington's Citizen Lake Monitoring Program, Washington State Department of Ecology, P.O. Box 47710, Olympia, WA 98504-7710; 360/407-6680.*