Considerations for Planning Your Program's Data Management System

University of Rhode Island   **University of Wisconsin**

Kris Stepenuck, Elizabeth Herron and Linda Green

Managing Your Data Effectively

Having a data storage, management, and retrieval system is essential for every monitoring program. Without such a system (called a data management system), information that is collected cannot be tracked in an organized fashion. There are multiple data management system options that you can choose from when planning a monitoring program. Deciding which is best for your program's needs is the first step. To help you in the planning process, we surveyed a number of volunteer monitoring programs that have developed online databases as their data management system so that we could incorporate their experiences - both good and bad - into this fact sheet. Our goal is to provide guidance to help you develop your own data management system.

We will provide answers to common questions, tips, anecdotes, and resources to give you a knowledge base about data management, and database planning and development. The questions you ask when developing a data management system will ultimately lead to other questions. The goal is to ask *enough* questions about what you want to do with the data for the system to be a success. Keep in mind that available resources may require you to make compromises in your planning and that developing technology may open new opportunities for your system. Although the process may be challenging, it will ultimately provide your program with a solid foundation for organizing, analyzing, and sharing information in an effective and efficient manner.

Data Management System Options

Each data management system has its own strengths and weaknesses (Table 1) for you to reflect upon as you choose and develop one for your program. Storing paper data sheets is the simplest form of data management. However, data aren't organized for easy analysis with a paper system, and if multiple users are interested in obtaining the data, sharing data with them might be problematic. Computerizing data allows them to be stored logically, and shared and analyzed easily. There are a variety of computerized data management options available, and understanding the benefits and drawbacks of each is important for determining which may meet your program's needs.

Considerations for Planning Your Program's Database

Our familiarity with spreadsheets leads many of us to use them for data storage. In spreadsheets, data are stored in tables that are made up of rows and columns. Data can be analyzed and graphed with ease, but data integrity can be compromised, data sharing can be difficult, and data duplication may be prevalent. Have you ever sorted data in Excel and failed to select all the data in the worksheet only to end up with data in a scrambled mess? Or do you have several people from one organization on your mailing list that is stored in a spreadsheet? What if that organization's office location moved? You would have to manually change the address for each individual. You can minimize such errors and frustrations in spreadsheets through programming, but it's generally easier to use a database, which is designed for storing data.

Table 1. Issues to consider when choosing and developing a data management system.
   (Adapted from "Data to Information: A Guidebook for Coastal Volunteer Water Quality Monitoring Groups in New Hampshire and Maine" by Geoff Dates and Jeff Schloss)

	Paper files	Spreadsheets (e.g., Excel, Lotus)	Desktop Databases (e.g., Access, Filemaker Pro)	Online Databases (e.g., Access, MySQL, Oracle)
Set-up and Costs	•Very inexpensive •Good to keep even with electronic data management system	•Inexpensive •Familiar to many and easy for most to use	•Inexpensive ($100-$200) •Fairly easy to use •Good for small datasets	•Costs vary- free to tens of thousands of dollars •Need a programmer to set up •Likely need to train users
Maintenance, Security, Backup, and Data Quality	•Lowest percent chance of error in data reported on data sheets •Save copies of field and lab datasheets •Maintenance, security and file back up dependent on amount of data collected and room to store it •Depending on amount of data collected, may require a large amount of storage space	•Electronic file back up simple •Can be programmed to meet individual program needs, but easier to use a database with built-in backup capabilities	•Electronic file back up simple •Can be programmed to meet individual program needs (e.g., allow for data entry templates, recognize data entry mistakes, provide template reports, create special data dumps for data sharing with other databases and STO-RET)	•Need a programmer to maintain database code over time •Electronic file back up simple •Can be programmed to meet individual program needs (e.g., allow for data entry templates, recognize data entry mistakes, provide template reports, create special data dumps for data sharing with other databases and STO-RET)

Table 1 (continued). Issues to consider when choosing and developing a data management system.

	Paper files	Spreadsheets (e.g., Excel, Lotus)	Desktop Databases (e.g., Access, Filemaker Pro)	Online Databases (e.g., Access, MySQL, Oracle)
Entering and Storing Data	•Similar amount of time as with spreadsheets, but depends on amount of data being stored	•Only one user can enter/modify data at a time •Good for storing small datasets	•Generally only one user can enter/modify data at a time (unless used in networked mode, during which data integrity may be jeopardized if more than one person uploads data at once, thus overwriting the others' data) •Data entry from a single computer •Can store large volumes of data •Training needed for data entry at a single location, and often for only a few individuals	•Multiple users can modify data at one time •Data entry can be done from any computer with Internet access (if access is granted) •Can store large volumes of data •Training required for many users; likely needed at multiple locations
Searching and Sharing Data	•Depends on amount of data to be shared and how well organized your files are	•Not as query friendly as a relational database •Good at statistical and mathematical analyses •Graphing friendly (although sometimes not as easy as other graphing software)	•Good for searching and retrieving data (query-friendly) •Data on web usually static (e.g., a person has to upload information to the web manually, it is not uploaded to the web automatically as a user enters data into the database)	•Good for searching and retrieving data (query-friendly) •Web posting of data is dynamic (e.g., as data is entered into the database, the website is updated automatically) •Training may be required to teach people to use

Considerations for Planning Your Program's Database

Databases are a collection of records stored in a structured way that allow a computer program to easily search for and locate desired information. According to Webopedia (http://www.webopedia.com/TERM/D/database.html), "databases are organized by fields, records, and files. A field is a single piece of information; a record is one complete set of fields; and a file is a collection of records. For example, a telephone book is analogous to a file. It contains a list of records, each of which consists of three fields: name, address, and telephone number." The most common type of database structure used today is a relational database. Relational databases have multiple tables which are related to one another by values that are common between the tables. For example, you might have data about water transparency at a site in one table. In another table you might store data about the latitude and longitude coordinates of the site. You can query a relational database to link the information about the transparency with the latitude and longitude of the site.

If you aren't sure if your program should store data in a spreadsheet or database, "I Get It Development" (http://www.igetit.net/newsletters/Y03_04/ssvdb.aspx) offers some key questions to help you decide. If you answer yes to any of the following, you should consider developing or using a database for your program:

- Do the data need to be stored long-term?
- Do multiple people need access to the data?
- Do I need to safeguard against erroneous entries?
- Do the data need to be protected against inadvertent corruption?
- Is a large part of the information redundant?

Two excellent resources to help you better understand data management systems and the differences between spreadsheets and databases are:
1. "Converting data to information (Section 6 of Volunteer Surface Water Monitoring Guide, 2003 - Minnesota Pollution Control Agency) http://www.pca.state.mn.us/index.php/view-document.html?gid=6860).
2. The Spring 1995 edition of the Volunteer Monitor newsletter (Volume 7, No. 1, URL currently not available but see https://www.epa.gov/aboutepa/about-office-water#wetlands).

If you have decided that a database is the best data management option for your program, the next step is to decide what type of database will best suit your needs and resources. Databases can be made available on just a single computer (we'll call these desktop databases), or they can be part of what's called a distributed data collection system (we'll call these online databases). On-line databases create a

system in which different users are able to upload data to or search for data within the system at the same time from different physical locations. About.com offers a series of articles to help step you through the decision process of choosing if a desktop database or an online database is right for your program needs (http://databases.about.com/od/administration/a/choosing.htm).

We'll focus the remainder of this fact sheet on development of online databases, though much of the following information is relevant no matter what type of data management system you are developing.

Technical components of an online database

There are four key technical components of online databases:

1. Operating system – This is the computer program that operates the hardware and software of a computer. It's essential in order for the computer to function in an orderly way. The operating system responds to input from users and equipment (such as keyboards or CDs), and is responsible for output to screens, printers, speakers, etc. It controls security of the computer as well.
2. Database – As mentioned, this is a collection of records stored in a structured way that allow a computer program to easily search for and locate desired information.
3. Programming (or scripting) language – These languages are the translators that allow us to give instructions to computers.
4. Web server – Servers accept and respond to our web (or http) requests by providing information that is requested (such as displaying a web page).

There are a variety of options to choose from to complete the set of technical components needed to have an online database; examples are listed in Table 2.

Table 2: Examples of open source and for-purchase technical components of an online database.

Components of online data management systems	Open source (i.e., free) examples	For-purchase examples
Operating system	Linux	Microsoft Windows
Database	MySQL	Oracle
Scripting language	PHP, ASP, C++	Cold Fusion (.cfm)
Web server	Apache HTTP Server Project	Microsoft Internet Information Server

Questions to consider when planning your database

Karen Diamond from Great Bay Coast Watch in New Hampshire noted in an EPA volunteer monitoring listserv discussion, "the MOST important thing you need to know before you start developing a database is what you want out of it." To help guide you through this process, Greg Johll, programmer at the University of Wisconsin-Extension shared his knowledge of database development and helped craft some key questions to consider carefully when planning and developing your database.
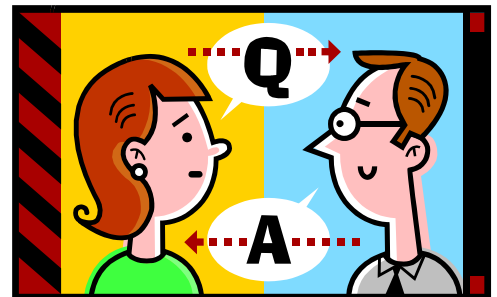
Considerations for Planning Your Program's Database

Questions to consider when planning your database

What resources do you have available to develop your database?
Carefully considering what resources you have available to you from the start will allow you to create a database that will both work to meet your needs and be feasible for your program to develop. Often when people think about resources, money is all that comes to mind. However, resources include much more than that. You should take into account three kinds of resources: human, database software and financial. Programmers may be available within your organization to assist (or lead) you in development of the database, or you may have to hire a consultant. Databases such as Oracle cost agencies thousands of dollars per year to maintain licenses. There are additional costs associated with supporting professional staff to maintain these databases as well. Other databases, such as Access or MySQL may be available at a lesser cost. Remember that not only will you need resources to develop and start up your database, you will just as importantly need resources to maintain and improve it!

Who will use the data?
This key question should be answered early in the planning process. Knowing the answer will allow you to craft the database so that it is user-friendly for most everyone from citizens to local municipal officials to agency staff, provided that is your goal. If you design for the least sophisticated user, you will probably not insult the most sophisticated. If you do the reverse you will likely frustrate your least savvy volunteers and may alienate them from your program. Design your database to enhance your program!

How will the data be used, and what type of output will best serve these purposes?
Think both about specific program uses for the data as well as external uses that you will want to support. You need to ensure that the database will be compatible with all of the uses you have planned.
Specific questions about data use to consider include:

- **What parameters do you want to be able to search by (e.g., waterbody name, county, Hydrologic Unit Code (HUC))?**

- **What type of searching options do you want?** At least three options exist:
    - Simple: A search on a unique identifier such as station name or county. For example, within the Wisconsin Water Action Volunteers' database (http://watermonitoring.uwex.edu/wav/monitoring/database.html) you can search by county, site location, or stream name. Search results are provided for whichever option is chosen and you cannot further narrow your search.
    - Dynamic: A search in which the options change based on the first selection you make. In the Hoosier Riverwatch database (http://www.hoosierriverwatch.com), you initially search by watershed (i.e., HUC), then narrow your search by river, and then by organization, site, or date, and finally by type of data (i.e., flow, chemical, biological, or habitat). In each level of searching, you are only provided valid options within that category. So, for instance, after choosing which HUC you are interested in, you are only able to choose from rivers within that HUC, not all rivers in Indiana.
    - Multivariable (or Advanced): A search in which multiple parameters can be chosen at one time, without narrowing options down. This type of search will be an option in the Wisconsin Department of Natural Resources' new online database for its Citizen Lakes Monitoring Network. A current example of this type of search is available through the USGS Water Quality Data for the Nation (http://waterdata.usgs.gov/nwis/dv/?referred_module=qw).

- **Will you store raw data, a calculated result, or both?**
Hoosier Riverwatch (http://www.hoosierriverwatch.com/) monitors determine a Pollution Tolerance Index Rating (rating) based on the macroinvertebrates found in the stream. Rather than just having citizens enter a rating that they have calculated, the database contains a data table in which citizens record the types and numbers of macroinvertebrates collected in their sample. The database then automatically calculates the rating based on the organisms found in the sample. This allows database users to compare which types of macroinvertebrates were found at a site in a given year. If only the calculated rating was stored in the database, valuable information about which types of macroinvertebrates were present at the sites would be lost. Further, if Hoosier Riverwatch changes how the rating is calculated in the future, scores can be updated.

- **What reports will you want to get out of the system?**
Do you want to display data collected on a given date like West Virginia's Save Our Streams database (http://www.dep.wv.gov/WWE/getinvolved/sos/Pages/VAD.aspx) does? They offer reports about water quality, physical conditions, biological conditions, habitat and several other parameters for database users to view. Or maybe you want to create a report like Wisconsin Citizen Lakes Monitoring Network (http://dnr.wi.gov/lakes/clmn/) that shows such things as average Secchi depth and calculated Trophic State Index value for the lake of interest, and also provides interpretation of the data for that lake?

- **Do you want to graph the data, view it in tabular format, etc?**
Coastal Carolina University Volunteer Water Monitoring Program database (http://bccmws.coastal.edu/volunteermonitoring/index.html) has the ability to plot water temperature, bacteria, stream flow, or a multitude of other data about individual sites. Users click the box to choose sites, which parameters and timeframes they are interested in and data are loaded with plots are automatically created, and tabs for data or additional charts.

- **Will you want to upload the data to a specific organization or another database such as STORET? If so, how would the data need to be compiled in order to do that?**

EPA's STORET database is a national database into which states report water quality data. Each IOWATER site is assigned a STORET number as it is registered in the program's online database (https://programs.iowadnr.gov/iowater//). Data entered for the site are then associated with that STORET number. Then, when the Iowa Department of Natural Resources (IOWATER's sponsor and database host) uploads data from its databases to STORET, the citizen-collected data are uploaded as well. More information about EPA's STORET database is available in the Winter 2005 edition of the *Volunteer Monitor* Newsletter (Volume 17, No.1).

- **Do you want to email the data or alerts about the data to an organization or individuals?**
Maine's Healthy Beaches Program database (http://www.mainecoastdata.org/public/) shares data with beach managers, resource managers, and the public regarding public health. Email alerts are sent to the program coordinator and beach managers (and often other officials). These individuals make daily decisions based on the data, deciding if a beach should be posted and if it should be sampled again or not.
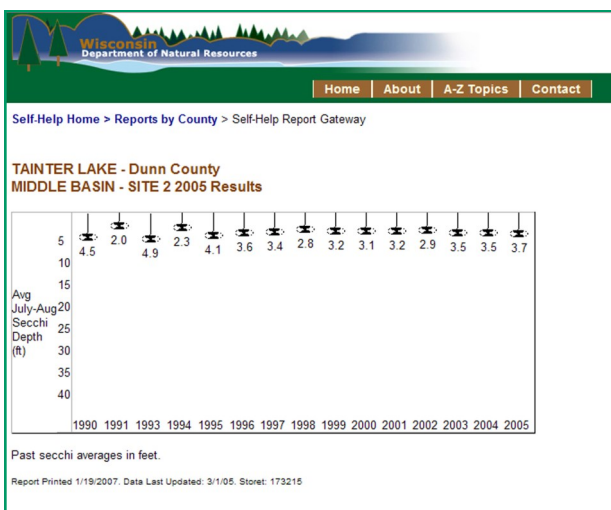
- **How many significant figures should be entered to the database and reported in output for each parameter?**
  This depends on what parameter is being measured and what is being used to make that measurement. For example, did you use pH paper or a laboratory meter? The resolution of these two methods is very different, and while meters can often report to 0.01 or 0.001, the litmus (pH) paper can't. If you are using methods with low resolution, you wouldn't want your database to imply greater resolution by just automatically setting all your significant figures at 0.001. (See *Factsheet VI Building Credibility: Quality Assurance and Quality Control for Volunteer Monitoring Programs* for more information http://www.volunteermonitoring.org/ look for Guide For Growing Programs)

- **How do you want the data to be stored within the database? Do you want to store data as a numeric or a character value?**
  This might seem a simple question, but consider the issue of a parameter that could be greater than or less than some value, such as with parameters that might be below or above the limit of detection. Such a parameter (e.g., < or >) would need to be stored as a character rather than as a numeric value. This will prevent automated calculations, unless a surrogate numeric value, such as half or twice the limit of detection, is assigned to such scores.

- **Do you want the database to provide statistical output? If so, what types of statistics will you want?**
  This will help determine the structural layout of the data (e.g., should your columns be organized by date, site or some other variable?). Most spreadsheets and databases have the ability to perform simple statistics such as standard deviation or averages, but if you want other, more advanced statistics automatically run, you will need to find out if your sponsoring organization has a statistics server or if the database needs to be specially programmed to provide those statistics. If having automated statistics is not a goal of your program, another option is to utilize Excel or a statistics program to analyze data after completing a query of the database and downloading the results to that program for analysis.

- **Do you want the database to have mapping capabilities?**
  The Colorado Water Quality Monitoring Council developed a data sharing network (http://www.coloradowaterquality.org/cdsn) between monitoring groups in the state. A web-based map will be linked to the database, serving as an online directory of monitoring activities. From the map, database users will be able to find data about what, when, where, why, how and contact information for chemical, physical, and biological data for all water body types.

IX - 8

How will you control the integrity of the data?

Controlling the integrity of the data by preventing data entry errors could include setting range restrictions for parameters, providing drop down menus, and programming it to auto-calculate values that are derived from data input to the system. These measures will minimize data entry errors and help ensure only high quality data are housed within the database.

West Virginia's Save Our Streams database (http://www.dep.wv.gov/WWE/getinvolved/sos/Pages/VAD.aspx) has dropdown lists with categorical choices for several physical conditions monitored by citizens. These include water clarity, color, and odor, streambed color, and algal abundance. Similarly, the World Water Monitoring Day database (http://www.monitorwater.org) provides drop down lists for what type of equipment was used to determine results for such parameters as pH, temperature, and dissolved oxygen. Offering these dropdown lists minimizes data entry errors, provides uniformity to the data set, and allows metadata (see sidebar) to be simplified.

Of course there are peripheral ways besides database programming to control data integrity that should be developed as well. For instance, after answering the question *"Who will input data to the system?"* you should develop training and written instructions for those individuals. You should also develop a plan for how you will quality check data once the information has been entered to the system.

Who enters data to the database is not only important to data integrity, but relates to the needs and interests of data users and data collectors. In New Jersey's online database, volunteers are able to enter data if they have internet access and interest in entering their data, otherwise local coordinators enter data from the volunteers. When volunteers do enter data, the information is not visible to the general public until it has been approved by the local data coordinator. Specially trained local coordinators, as well as program staff, are able to run statistics, develop graphs or otherwise view or analyze the data before they are available publicly. This "time lag" to view data in the system was suggested by volunteers who wanted to ensure their data were reviewed and satisfactory before being viewed publicly.

Metadata - "data about data"

They describe who, what, where, why, when, how, and quality of the data collection and analysis. For more information about metadata, see the Winter 2005 edition of the *Volunteer Monitor* newsletter, Volume 17, No.1.

For a detailed summary of meta data see also http://www.language-archives.org/documents/gentle-intro.html.

Considerations for Planning Your Program's Database

What sort of security do you want the database to have? How will you control that security?

Databases have both internal and external security options. Most volunteer monitoring programs control security internally. This level of security will be visible to system users as a request for people to enter usernames and passwords for data entry and editing.

Some databases use "encrypting" as an external security measure. In encrypting, data are protected with a code (or are scrambled) as they travel electronically between a website where data are entered and the server where data are stored. This process helps protect information from hackers during transmission. To encrypt data, a certificate of security must be purchased by the database developer. If you have done any online shopping, you have likely seen this security measure in action. In your web browser (e.g., Internet Explorer), an "s" follows "http" in the web address and a padlock symbol appears at the bottom of your screen when you are entering secure data such as your credit card number.

Who will have administrative privileges within the database?

That is, who will have the right to update and/or edit the database? There may be several levels of administrative privileges within a given database. For example, you may opt to give volunteers the ability to enter data, but not edit them once they are in the database, such as in the IOWATER program. Only IOWATER staff are provided rights to make modifications to data submitted by the volunteers. Once those staff have the appropriate administrative privileges, they can do such things as change incorrect data, and delete duplicate records or records entered by mistake for the wrong site. For quality assurance purposes, added administrative rights may come with a requirement for additional data entry. In Iowa, as changes are made to the IOWATER database, the date and time the changes are made is recorded, and the person making the changes must fill in a field to explain what the change was and why it was made.

A slightly different approach to administrative privileges is followed in New Jersey. The program coordinator assigns a start and end date for each registered monitoring project. Then local data coordinators have rights to register specific monitoring sites for those projects, and volunteers have access to enter data for those sites during that time period.

Communication - the Key to Successful Data Management

Good communication between the programmer/consultant and the program coordinator is an essential element in developing a successful online database. The Volunteer Monitor newsletter highlights the topic of managing and presenting data in its spring 1995 edition (Volume 7, No. 1). One aspect mentioned is that there is a tendency by those involved in volunteer monitoring to think that data management isn't "my" job. Yet, in fact, there is a role for everyone in data management and database development – from programmers and program coordinators to volunteers and program sponsors. Besides the questions already mentioned, you will undoubtedly encounter many more. Discussions between developers and database users should be on-going throughout the planning process.

Online Databases of Volunteer Monitoring Programs

Considerations for Planning Your Program's Database

A number of volunteer monitoring programs have decided that the type of database that would best meet their needs and goals are those which offer online data entry and/or searching capabilities. If, after review of your program's goals and resources you find that an online database is right for your program, it may be helpful to visit these programs' databases to help start planning your database.

- Alabama Water Watch (http://www.alabamawaterwatch.org/water-data/)

- CitSci.org provides tools including: creating new projects, managing project members, building custom data sheets, analyzing collected data, and gathering participant feedback (http://www.citsci.org/cwis438/websites/citsci/Data.php?WebSiteID=7)

- Colorado River Watch (http://cpw.state.co.us/aboutus/Pages/RW-DataSheets.aspx)

- Colorado Data Sharing Network (http://www.coloradowaterquality.org/cdsn)

- Great North American Secchi Dip-in (http://www.secchidipin.org)

- Illinois' Volunteer Lake Monitoring Program (http://www.epa.state.il.us/water/vlmp/index.html)

- Indiana's Hoosier Riverwatch (http://www.HoosierRiverwatch.com)

- IOWATER (http://www.iowater.net/database/viewdata.asp)

- KnowledgeBase Science & information management for the Gulf of Maine & its watershed (http://www.gulfofmaine.org/kb/2.0/search.html)

- Maine Healthy Beaches Program - Maine's Shore Stewards (http://www.mainecoastdata.org/public/)

- Michigan's MICorps Data Exchange Network (http://www.micorps.net/data/view/search/)

- Missouri Stream Teams (http://www.mostreamteam.org/mapwelcome.asp)

- National Estuarine Research Reserve Centralized Data Management (http://cdmo.baruch.sc.edu/)

- New Hampshire's One Stop Database (http://www2.des.state.nh.us/OneStop/Environmental_Monitoring_Menu.aspx)

- New York's Westchester County's Citizen Volunteer Monitoring Program (http://planning.westchestergov.com/volunteer-streammonitoring)

- Oregon's Student Watershed Research Project (http://www.pdx.edu/student-watershed-research-project/data-sets) Go to Data.

- Pathfinder Science's online mapping system (http://www.pathfinderscience.net/?maps=1)

- Utah Water Watch (http://extension.usu.edu/utahwaterwatch/htm/database)

- Washington's Nature Mapping (http://naturemappingfoundation.org/natmap/

- West Virginia Save Our Streams (http://www.dep.wv.gov/WWE/getinvolved/sos/Pages/VAD.aspx)

- Wisconsin's Surface Water Integrated Monitoring System (SWIMS) (http://dnr.wi.gov/topic/surfacewater/swims/helpdocs.html)

- World Water Monitoring Day (http://www.monitorwater.org)

Considerations for Planning Your Program's Database

Interesting statistics about developing volunteer monitoring online databases (based on 19 programs)

- Timeline: Database development and management is an ongoing process. On average, it took programs 2.5 years to develop and implement online databases, though several coordinators indicated that their databases are being continually updated and modified. The shortest time period a database was created and went "live" was 3 months. The longest was 12 years. Thirteen programs had a trial period for the database before all users were allowed access to it.

- Funding: Most programs (10) used state funds (at least in part) to develop their databases. Six used federal funds, five used grants, and four used donations. A few programs used in-kind services to develop and maintain their databases. Most of the programs continue to use the same sources of funds (i.e., federal, state, grants and donations) to maintain their databases today.

- Data Entry: Most programs (11) have trained their volunteers to enter data into the database, though eight also have staff enter data for those who prefer not to or are unable to enter the data themselves. Three programs rely strictly on their volunteers and two programs rely solely on staff for data entry.

- Data Searching: The primary provision for searching in these online databases is by waterbody name (14 programs), with watershed the next most common option (10 programs). Searching by county, site, parameter or HUC is available with about half of the programs. No programs offer the option to search by STORET number.

- Security: Fourteen programs indicated they required usernames and passwords for entering, editing, and/or deleting data. Four of the programs have specific security measures defined for users who wish to search their databases.

Ensuring data quality

An essential component of any data management system is ensuring that the data are of known quality. This influences the usability, reliability and validity of the data. It is important not only to record what information was collected (e.g., what the water temperature was), but how it was measured (e.g., digital thermometer, spirit-filled thermometer), the precision of the method, as well as other metadata. Database users can then access the data along with the metadata and have a complete picture of how the data collection was done.

The Citizen Monitoring Program (Clean Water Team) of the California Water Resources Control Board has a data quality management system in which data are organized along with extensive metadata, allowing the data to "speak for themselves". Better yet, it's available for modification as a Microsoft Excel workbook with template files to be downloaded and modified to meet other monitoring programs' needs (http://www.swrcb.ca.gov/water_issues/programs/swamp/cwt_toolbox.shtml). It offers the ability to enter data about monitoring results (e.g., data range of the equipment used) as well as information about which equipment was used and the quality of the measurement (e.g., precision of the instrument used). Other information can be stored in the workbook as well, including station type and weather conditions during sampling. Additional supporting documentation about data quality management is available in the Clean Water Team's Guidance Compendium for Watershed Monitoring and Assessment (http://www.swrcb.ca.gov/water_issues/programs/swamp/cwt_guidance.shtml).

## Tips from those who have developed online database

Program coordinators across the nation offered suggestions for those who are considering developing an online database. It was striking how similar they were. Here's what they had to say:

- Where possible, share. It is definitely worth exploring where you might be able to use data systems that already exist to store and manage your data.
- Double your anticipated time line, if you think it will be done in a year, make it two years.
- Double your budget too.
- Think about the data you are collecting first. This includes considering:
  o What data will you collect, what parameters will be monitored, what units will you report results in,
  o How you want the data to be retrieved from the system,
  o Where the data are going to reside (what database),
  o How you want to sort the data, and
  o What your reports are going to look like.
- Make sure the volunteers, program coordinators and data users are all involved in the creation of the data system, from start to finish.
- You may need the advice of folks who have particular areas of expertise much different from yours: standards, methods, database management, etc.
- Be extremely clear when you talk to the developers about expectations. Define what you want in the database and how you want it to work. Then check, double check, and triple check that the database developers understand, and know what you expect in the way of appearance, functionality, etc.
- Database developers often speak a different language than program managers. You may think you explained something well enough, and even though it seems intuitive for you, it isn't for them. And once they are done coding, it is REALLY expensive to go back and change what they have created.
- Be in constant communication with the database developer. Fifteen emails and/or phone conversations a day is okay and is well worth it if you get exactly what you want. You can never be too clear on what you want the database to look like and what functions you want it to have.
- Draw pictures of how you expect your screens to look. You can use control-alt-print screen and paste the screen shot into Paint, Word, or another program. Then you can cut and paste to move text around, etc. and save and/or print it out.
- Help whomever is designing your system to understand your business processes and those of your volunteer monitoring community.
- Understand that you will not be able to make everyone happy, do not feel you have to cater to one or two individuals.
- Develop step by step guidance about how to use the database that you can easily share (preferably electronically) with people who have questions. You can spare yourself a lot of time if you create easily accessible, easily understood guidance about using the system.
- Make the application as easy to use as other popular applications, like *iTunes* for example.
- Consider creating an online lookup or dropdown menus for volunteers who have forgotten their volunteer ID or site numbers.
- Allowing volunteers to enter their own data gives them additional ownership of the data. It can be the capstone of the monitoring process, and supports our idea that volunteers should be monitoring for their own interests and use of the data - not just to provide data to the state.
- Hire a data person.
- If staff enter data, some OTHER staff person should be validating that entry.

The best tip? Hone your sense of humor and diplomatic skills. Developing a sophisticated system will be stressful for everyone involved. How you handle it will make a big impact on your project.

IX - 13

Considerations in Planning Your Program's Database

The Environmental Information Exchange Network (http://www.exchangenetwork.net/basics/how_it_works.htm) is an emerging and important development for sharing water quality data across the nation. The EPA, States, Tribes and U.S. territories are developing this Network to allow data to be shared securely through the Internet in an efficient manner. Partners in the exchange set-up servers called nodes that link to the Network, allowing them to make data requests and to enter and share data (see illustration). A programming language called extensible markup language (XML) is used to convert the data to a common format, allowing the information to be shared among Network partners regardless of what type of computer system or platform is used by each to store their own data. It is likely that XML will gain in use and importance.

Environmental Information Exchange Network
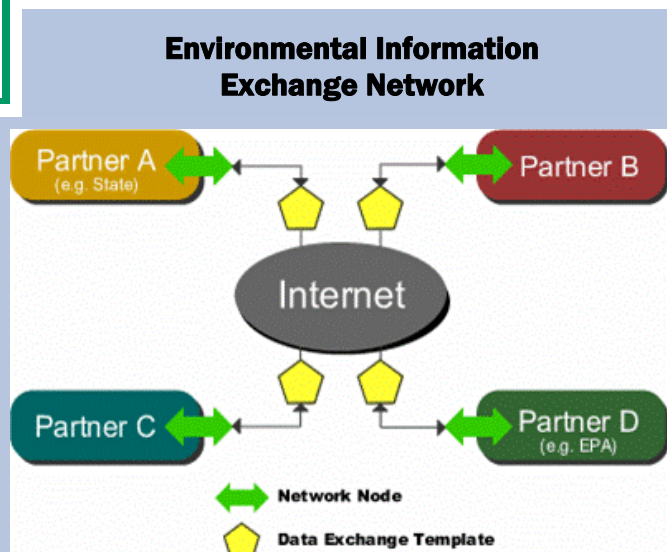
(Illustration courtesy of the Network Operations Board)

Summary

Developing a data management system for a monitoring program can be a long and complicated process. Decide what you want out of the system before you create it. Use available resources to determine the type of data management system that best fits your program's needs. Ask questions throughout the development process to help clarify how it will be set up, what types of output it will produce, and what information should be collected to ensure the data stored in the system are of the quality needed to meet program goals. If an online database is the best fit for your program's goals, look for opportunities to share a system with other programs and agencies to save costs for everyone. If you develop your own system, plan to be in constant communication with programmers, and to spend double the amount of time (and budget) you expect to in the development process. Remember that you are not alone in your endeavors; there are a number of programs that have developed online databases who are willing to offer advice based on their experiences. In addition, take advantage of the opportunity to review existing online databases and note features you like/dislike. You will be rewarded for the efforts you put into the development process.

CONTACTS

Linda Green
    Phone: 401-874-2905, lgreen@uri.edu
Elizabeth Herron
    Phone: 401-874-4552, eherron@uri.edu
Arthur Gold
    Phone:401-874-2903, agold@uri.edu
University of Rhode Island Cooperative Extension
Coastal Institute in Kingston, Rm 105
Kingston, RI 02881

Kris Stepenuck
    Phone: 802-656-8504, kris.stepenuck@uvm.edu
University of Vermont
Extension Leader, Lake Champlain Sea Grant
81 Carrigan Dr. #312F
Burlington, VT 05405