



DATA MANAGEMENT IN ACTION 2: WEB-BASED SYSTEMS AND EDAS DEMOS

SESSION INFORMATION:

Moderator:

Ken Cooke, Kentucky Waterwatch

Presenters:

Bryan Parker, Missouri Stream Teams
Web-Based Systems

Mellini Sloan, Tetra Tech, Inc.
EDAS: Facilitating the Analysis and Archiving of Ecological Data
(abstract only)



DATA MANAGEMENT IN ACTION 2: WEB-BASED SYSTEMS AND EDAS DEMOS

Web Based Systems

The first half of this presentation looked at the various ways water quality monitoring groups are using the Internet (World Wide Web) to report and collect volunteer data. The second half discussed the Missouri Stream Team and Volunteer Monitoring programs use of a relational database, Microsoft Access, to store, compile, and report volunteer data.

On-Line Data Retrieval

Volunteer monitoring data is available in many forms on the web. The methods used to present the data seem to fall into three general categories. The first is advertisement, 'contact us and we'll send you the data you request.' The second is data that is "hard coded" into web pages (this data can be presented as text, tables, or graphic images). The last is "database" driven. The database driven reporting systems usually require some sort of programming to make it all work.

- *Advertisement.* The data is available, and information about how to obtain the data is on the web.
- *Hard Coded.* The data is assembled and converted into html for use in the web site. Charts and graphs are created and treated as image files. If you don't have a lot of data, or a programmer, this is a quick way to get your data on-line and available.
- *Database Driven.* The data is stored in a database file, and the html page you see is generated by a program. The data is usually searchable. The sites range from fairly simple to very complex systems that generate maps and data.
 - Missouri Stream Team <http://www.mostreamteam.org/vmsearch.html>
- *Metadata.* Metadata is the description of each data item, the data format, what each item means, and the item's limitations. For example, using a series of litmus paper tests to determine pH does not produce the same quality of results as a \$100 pH pen. Methods, monitor qualifications, and QA/QC should be noted.

On-Line Data Submission

If you keep the data in a database, why not let the user type it in instead of you or your staff? Hopefully, the "ease" of online submission will get those data reports in a little sooner too.

To understand on-line data submission, you need at least a partial understanding of internet forms. Forms are all those blanks and buttons you fill out to send data over the internet. All data submission starts with a form. The simplest form is an "email" form. You fill out the data, press "Submit," and the browser sends the data to the email address specified in the form. The email results are coded with the form item name and the resulting user data. This makes it a little hard to read. Here's an example of data taken directly from an email form:

&Name=Bryan+Parker&Address=20+Richard+Drive

- To really be able to use this data, you need to parse it. Parsing programs are usually custom written in PERL or some other programming language.

More sophisticated data submission sites use multiple pages to get the data where it needs to go. Here are some examples:

- **The Vermont Rivers Project** http://dauntless.smcvt.edu/vt_rivers/db_password.html. This online database requires the user to log in before submitting data. If data is sent directly to an on-line database, this form of data protection may be necessary.

CONTACT INFORMATION

Bryan Parker
Missouri Stream Team Volunteer
20 Richard Drive, Rolla MO 65401
phone: (h) 573/341-2226 or (w) 573/364-6362
fax: 573/364-4782
email: baparker@rollanet.org

For more information:

Call the Missouri Stream Team program office at 800/781-1989 or visit the Web site at <http://www.mostreamteam.org/>

The presentation outline and examples from the Missouri Stream Team and Volunteer Monitoring Program databases are available online at this URL: <http://www.mostreamteam.org/conference/vmdb/index.htm>.



- **Missouri Stream Team On-Line Data Submission** <http://www.rollanet.org/~streams/vmdata.html>. These forms check the users input data, and return the form to the user to verify out of range data. For example, if a user input pH=27, the form would be returned with a question about the pH data. The pull down boxes also help limit the input in several forms. These forms email the results to the Stream Team office for verification and entry into the master database. If the user supplies an email address, the program e-mails a copy of the data to the user, as well as to the office.
- **The Rivers Curriculum Project On-Line Data Entry Form** <http://www.siu.edu/OSME/river/wq.html>. The form in this last example is created by a commercially available program. The program can also be run on the computer receiving the email, and will parse the email submitted by the form directly to a database. The program is called WebForms™, by Q&D Software Development <http://www.q-d.com/>. The program sales for about \$40.

Here are other resources to help you get your on-line data submission up and running.

- There is a free, email parsing perl scripts available at Matt's Script Archive <http://worldwidemart.com/scripts/formmail.shtml>. Programming experience is helpful.
- WebForms™, by Q&D Software Development <http://www.q-d.com/>.

The complete outline, links, and examples are available on the Internet. The URL for this page is: <http://www.mostreamteam.org/conference/online.html>.

Missouri Volunteer Monitoring and Missouri Stream Team– Relational Databases for Data Management

A database is an organized collection of information. In relational databases, groups of data reside in separate tables, linked by one or more fields. The goal is to minimize or eliminate duplication of data. For example, volunteer name, address, etc. is one table. This table is linked by a volunteer ID to an activity, to monitoring results, and team databases.

The Missouri Volunteer Monitoring Program (V.M.) began in 1993. The program staff recognized the need to use a relational database instead of multiple lists. The original database was started in FoxPro, and later switched to Microsoft Access. Originally the V.M. database evolved independently from the “master” Stream Team (ST) database. Now ST and V.M. databases are linked together.

The V.M. database consists of five primary tables. The “Site” table contains the location of the each site. Each site has a unique identifier or primary key, “GISID”. The data tables “Invert”, “Visual,” and “Waterchem” contain data for each monitoring event, and are linked to the Site table. Each record is linked to a site by the field GISID. The “Members” table contains the name, contact, and training information for each volunteer monitor.

The Site database, which is the list of each monitored location, has 989 individual sites monitored since 1993. The program is currently monitoring 330 sites, and has trained more than 2045 people. The Water chemistry database has 3,537 records, each record is a monitoring event. The Macroinvertebrate database has 1,665 records. The Visual database has 1,387 records. The combined ST and V.M. databases fill 15 megabytes of disk space.

Lessons Learned

The Missouri ST and V.M. staff has done a tremendous job of coordinating a vast amount of information for the various beneficiaries of the program. Here are some of the lessons learned along the way:

- Maintain uniform field names and field types across tables. Tables with fields for County, Co, or Cnty can't be linked until the field names are changed. It is easy to change the field names, but renaming fields will mess up existing queries and reports.
- Spelling, naming conventions, and abbreviations are important. Searching for “Stone Cr.” won't find “Stone Creek” or “Stone Branch”. Do you use periods or not? For example “Cr.” (period inside of quotation marks) is not the same as “Cr” (no period inside of quotation marks).
- Add comments to all your data, fields, queries, etc. How will you remember what the query “getall” does a year from now?



- Sanity check all data, use QA/QC info for each record. QA/QC levels change as volunteers get more experience and training.
- Repeat, or check, data entry. Better yet, do both.
- Keep all original paper records. If something dies, you can go back and recreate from the original records.
- Make field names as explicit as possible.
- The site location is the most difficult item from which to get correct data from volunteers. Volunteers have a hard time obtaining GIS compatible location information. The program staff resorted to asking for a map with an “X” on it. The staff determines legal description and latitude/longitude or UTM coordinates. The correct location information is important to create an accurate Geographic Information System (GIS) map.
- Need a database to track people, and their level of training.
- Don’t delete names or records from the database. In Missouri, “Sunshine Law” (public access to state records, meetings) requires this. Mark as inactive.
- Make one person, or a committee, responsible for the database or anarchy will quickly take over. Old, one time use queries and reports fill up space.

Many state, county, and local agencies, as well as private organizations and individual citizens, benefit from the data that the ST and V.M. programs have collected. The site geographic information is important to generate reports by county, state representative district, watershed, etc. Many volunteers want to know who is sampling in their area. The volunteer monitoring data is used to supplement enforcement data, and create baselines to measure the change in streams over time. Quarterly activity and water quality reports are sent to regional state environmental offices.

The amount of volunteer time spent monitoring, and number of monitoring events, can be sorted and reported in various ways. One way is the ST program annual report, which estimates the value of the volunteer's time. The ST database also generates the program mailing list, and tracks the distribution of incentives and rewards.

Any mention of the Internet brings to mind security and privacy issues. In Missouri, all monitoring data is available online. The names of the program volunteer submitting monitoring data are not available online, and are not linked to the volunteers’ data except through the GISID field. The program limits the distribution of volunteer names to protect the volunteer's privacy.



DATA MANAGEMENT IN ACTION 2: WEB-BASED SYSTEMS AND EDAS DEMOS

EDAS: Facilitating the Analysis and Archiving of Ecological Data

CONTACT INFORMATION

Mellini Sloan, Environmental Scientist
Tetra Tech, Inc.
10045 Red Run Boulevard, Suite 110
Owings Mills, MD 21117
phone: 410/356-8993, fax: 410/356-9005
email: Mellini.Sloan@tetrattech.com

Ecological data derived from assessment and monitoring programs include biological, physical, chemical, and geographical data. In an effort to facilitate the analysis of this data, we have developed a tool to calculate various biological and habitat metrics used to assess the health of waterbodies. The *Ecological Data Application System* (EDAS) is a relational database application for Microsoft Access®, designed to enable users to easily manage, aggregate, integrate, and analyze ecological data.

Often, data are collected and individual sites are described on field sheets, but this data lingers on paper and no integrated or comprehensive analyses are performed. Data management systems, like the USEPA's STORET, which are designed to organize this data, do not incorporate components to analyze, interpret and incorporate the data they store into management decisions.

EDAS provides flexibility in data manipulation through a collection of tables that are linked by one or more fields. Edits, additions, or other changes in the data can be easily accommodated by modifying tables without affecting the overall database design. EDAS can be linked to various other applications such as ArcView®, ArcInfo®, USEPA's STORET system, image files, etc. Access® supports Open Database Connectivity (ODBC) and Structured Query Language (SQL), allowing it to interact with other ODBC-compliant databases.

EDAS draws on data stored within its own data tables to perform queries for data reduction and data analysis to accelerate biological assessment of water resource quality. EDAS utilizes sequential and "spreadsheet" type forms for data entry to minimize the burden of learning new software, and creates batch files for upload into STORET. Through its assemblage of abilities, EDAS eases the transition of state data from spreadsheets to a centralized database; increases state's willingness to archive data in STORET, and facilitates data analysis and the development of biological criteria to be integrated into management efforts.