

## **ELECTRONIC DATA MANAGEMENT SYSTEM FOR WATERSHED ASSESSMENT AND RESTORATION PLANNING<sup>1</sup>**

A. Edward Sciulli, and David G. Minnear<sup>2</sup>

**Abstract.** Watershed Restoration Plans (WRP) and subsequent restoration activities are generally only as good as the data and data analyses on which they are based. Until recently, watershed assessments in Pennsylvania have generally not incorporated a watershed-wide approach. Portions of a given watershed may have been assessed several times over the years by various groups, organizations and agencies, each collecting and analyzing their own data without coordinating efforts, techniques, or interpretations. This usually resulted in an incomplete assessment of the watershed as a whole and subsequent allocation of funding to projects that may not efficiently lead to watershed restoration. This paper presents the development and use of an Electronic Data Management System (EDMS) designed specifically for watershed groups to manage and evaluate water quality and other physical data for the preparation of a WRP.

The EDMS consists of two separate but linked components. The first is a relational database designed for both functionality and ease of use. Key to the database structure is the ability to conduct comparative analyses across several discharges, as well as comparative analyses within individual watersheds and between watersheds. More importantly, the EDMS provides a means for quick and efficient input of water quality and other physical data as well as providing a final repository for the data. The second component is a geographic information system (GIS) which provides the spatial relationship lacking in the database. Key to the GIS structure is the development of routines to display and analyze the spatial distribution of discharge loadings and levels of specific contaminants.

The EDMS has been found to significantly increase the ability of local watershed groups to conduct watershed-wide assessments and data analysis. The system allows less technical users to conduct more thorough evaluations of data and data trends and assists in the spatial analysis of contaminant loadings. Incorporation of widely available GIS data themes and layers provides additional data tools to evaluate local land use, topography, and land ownership, if available.

Additional Key Words: database, GIS, restoration plan, prioritization.

---

<sup>1</sup> Paper was presented at the 2004 National Meeting of the American Society of Mining and Reclamation and The 25<sup>th</sup> West Virginia Surface Mine Drainage Task Force, April 18-24, 2004. Published by ASMR, 3134 Montavesta Rd., Lexington, KY 40502.

<sup>2</sup> A. Edward Sciulli, PG, is a Project Manager and David G. Minnear, PE, is a Senior Technical Leader with L. Robert Kimball & Associates, Inc., Ebensburg, PA 15931.

Proceedings America Society of Mining and Reclamation, 2004 pp 1627-1642

DOI: 10.21000/JASMR04011627

<https://doi.org/10.21000/JASMR04011627>

### **Introduction**

One of the key elements to watershed-wide restoration planning is the collection and evaluation of existing environmental data. Most volunteers are aware that regional GIS data, including stream networks, topography, and demographics are widely available for many states, counties and communities. However, most are unaware of the large amount of water quality, biological and other watershed specific environmental data that may be available for inclusion in the restoration planning process. Furthermore, groups that do obtain the existing data are then faced with the problem of data management and evaluation. Due to the expense and a lack of experience with alternate methods, watershed groups generally resort to manual manipulation of hard copy data and multiple spread sheets. These methods tend to generate incomplete evaluations since the volume of data analyzed is limited by the method. A truly efficient and easy to use means of data management and analysis is needed to help volunteer groups develop restoration planning strategies, analyze water quality trends, incorporate new data, track progress, and effectively share results and ideas.

This paper presents a potential solution to the above problem through the development of an Electronic Data Management System (EDMS) designed specifically for watershed groups, to manage and evaluate water quality and other physical data for the preparation of watershed-wide restoration plans and prioritization. The EDMS consists of two separate but linked components. The first is a relational database produced using Microsoft Access<sup>®</sup>. Several custom user interfaces for efficient data input and evaluation of water quality results and trends are included in the system. The key to the database structure is the ability to conduct comparative analyses across several discharges or sample locations, comparative analyses within individual watersheds, and comparative analyses between watersheds. The EDMS also provides a means for quick and efficient input of water quality and other physical data while providing a final repository for the data. The second component is a GIS developed using Environmental Systems Research Institute, Inc.'s (ESRI) ArcView GIS<sup>®</sup>. The GIS provides the spatial component lacking in the database analysis described above. Similar to the database component, the GIS contains several user interfaces and custom scripts to perform specific data and spatial analyses. The key to the GIS structure is the use of built-in routines to display and analyze the spatial distribution of discharge loadings and levels of specific contaminants.

### **Relational Database Development**

Microsoft Access<sup>®</sup> was chosen as the platform for the relational database mainly for its ease of use, compatibility with other applications, and ability to provide a customized user interface. The two main objectives for the database design were ease of use through the development of standard user interfaces, and flexibility to perform a wide variety of functions. Specific design, data management, and analysis objectives were as follows:

- Data relationships to reduce database size and simplify querying;
- Efficient and reliable entry of field and hardcopy water quality data;
- Automatic import of available digital data (laboratory results, positional data, etc.);
- Automated quality assurance and integrity checks;
- Access to sample locations (discharge information, samples and water quality);
- Water quality trend analysis by individual location, watershed or sampling event;
- Comparison of results to watershed-specific or other criteria (by matrix); and
- Production of report tables (water quality, loading, total concentrations, etc.).

#### Data Relationships

Fig. 1 presents a graphical representation of the database designed to meet the above objectives, including a presentation of the various data tables and relationships. Relationships between data tables were established to define data uniqueness and efficient querying. One-to-many relationships, where one data element such as a sample may have many related data elements such as water quality results, were incorporated between data tables to limit data duplication and reduce the size of the resultant database. From this basic design, user interfaces were developed for manual data input, digital data importing, accessing built in specific data queries, and the input of user defined queries. Interfaces were designed to allow maximum flexibility while entering and accessing the data. The design allows end users to perform powerful data management tasks with little or no knowledge of how data tables were related, the specific formats of the underlying data, or Structured Query Language (SQL).

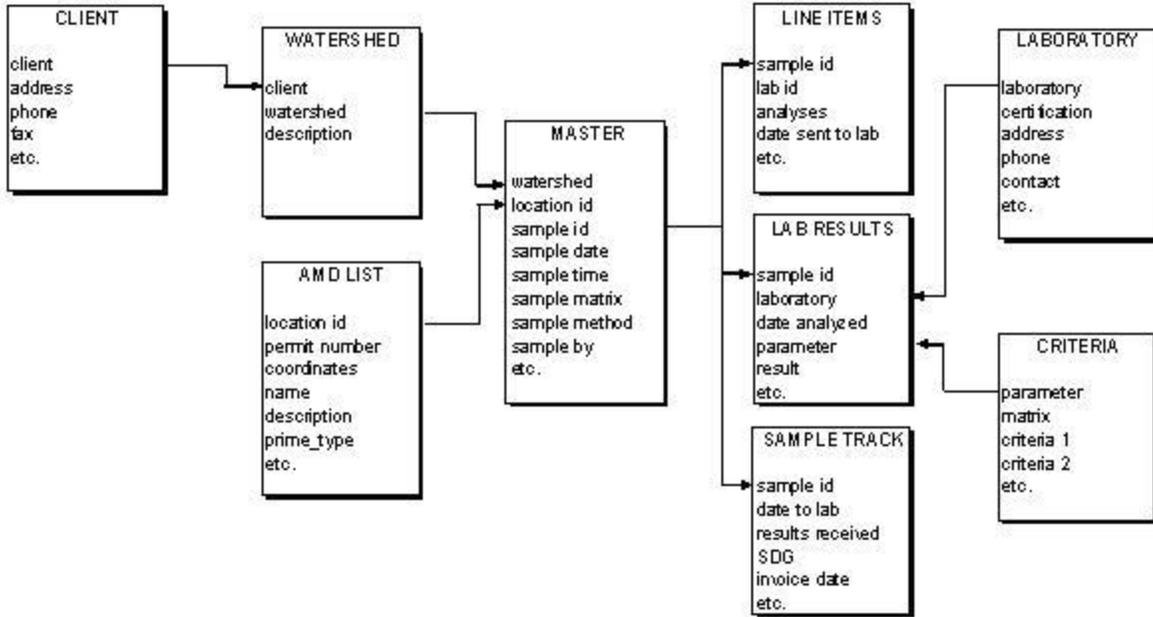


Figure 1 – General database design showing the data tables housed within the database, types of data in each table and the relationships between tables.

### Data input

Data input is facilitated through user interfaces and automated input routines. Data that can be received in digital form (water quality data, sample location coordinates, etc.) are requested in specific formats for automatic upload to the system. For manual data input (discharge locations, sample locations, field measurements, etc.) the end user is provided with the option of either entering data to a prescribed spreadsheet format ready for eventual automatic upload, or manually through the provided user interface. Because manually acquired data are eventually input to digital form for later production of report tables and figures, input directly to the system was found to take less effort; in part due to the structured nature of the supplied input forms.

### Data Integrity

Quality assurance and maintenance of database integrity are of the utmost importance in any database design. Integrity is controlled through a series of checks and balances. Designation of maximum data lengths, specification for required input, the setting of specific data types and data checking routines to evaluate relationships are a few of the methods used to maintain integrity.

Imported digital data are checked for completeness, data types, and lengths with regard to the destination tables and fields. In addition, both manually entered and imported data are screened to avoid mishaps such as duplicate entry. In all cases, the end user is provided with prompts, warnings and guidance throughout the data input process.

Viewing Discharge/Sample Locations

The system provides several options for data viewing and analysis. Figure 2 presents an example user interface developed to display information associated with an individual discharge or sample location by watershed. The end user selects the desired watershed and discharge or sample location and all available data are displayed. In addition to physical and identification attributes such as the discharge name, description, coordinates, and source, the user is supplied a tabular display of available water quality data associated with acid mine drainage (AMD) and a chart of the available data showing the concentrations over time.

Simple navigation buttons are provided for the user to enter a new discharge or sample location, edit existing data, enter/edit samples collected and manually view/input/edit water quality data for a selected sample. Once activated, these navigation buttons launch similar user interfaces designed to lead the user through the desired step. As new locations and/or data are entered, the underlying database is immediately updated to display the most recent information.

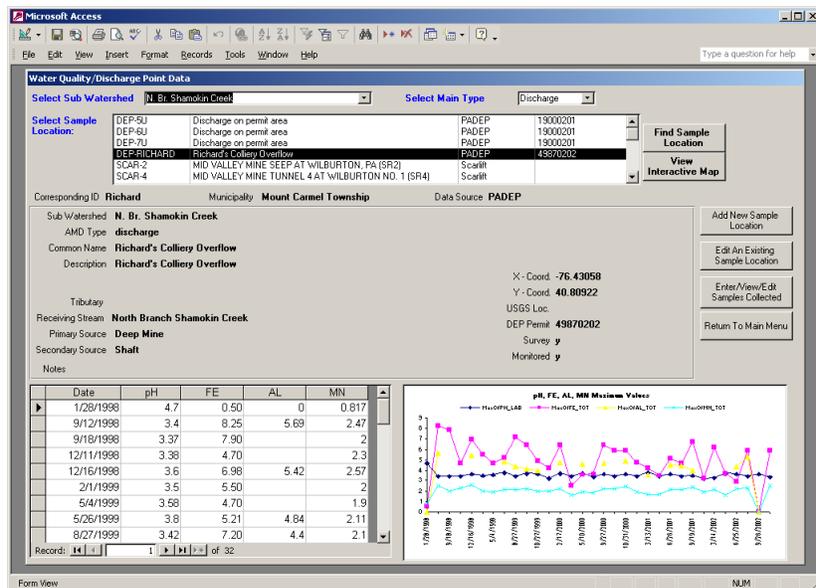


Figure 2 – Typical user interface showing user selections, physical description, tabular data and data chart for a given discharge.

Water Quality Analysis

Another set of user interfaces were developed to access water quality or other data associated with individual sampling locations, groups of sample locations, sample locations by watershed, or specific sampling events. These user interfaces provide an easy-to-follow system of entering the specific criteria on which to query the underlying data. Fig. 3 presents an example interface developed to query water quality results for up to three sample or discharge locations. Using the interface, the end user is prompted to input or select from pull down menus the watershed, type of locations to query, up to three sample or discharge locations, the range of dates between which samples were collected, and the water quality parameters. Based on the input criteria, analytical results are displayed in both tabular and graphical form showing the relationship of selected parameters for each sample location over time. Data can then be printed or exported to a variety of formats for later use in other applications or for presentation.

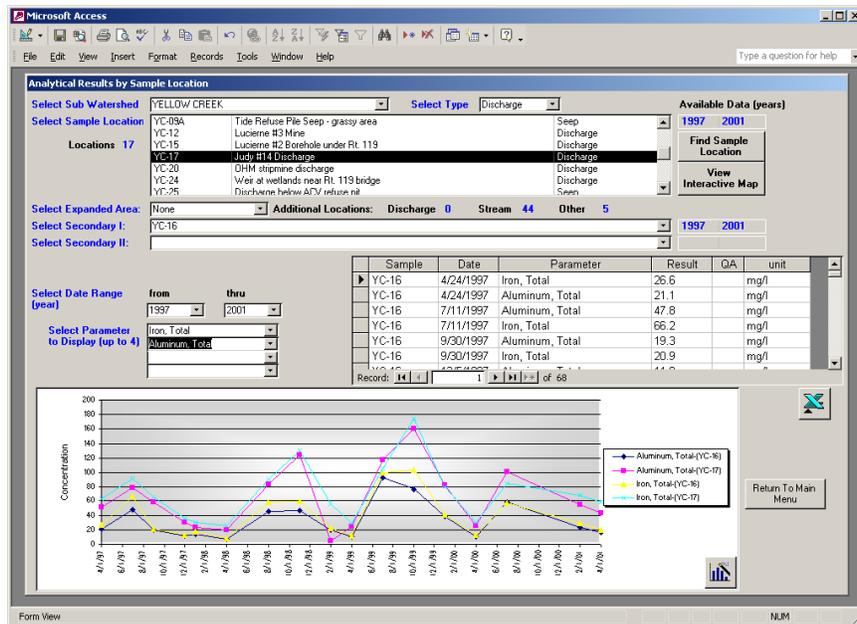


Figure 3 – Example interface comparing total iron and aluminum concentrations between 1997 and 2001 at two discharge locations.

Similar user interfaces were developed to display query results for a selected watershed by sampling event or by all sampling events over a given time span. Fig. 4 presents an example interface developed to query water quality results for a selected watershed. The end user is prompted to input or select from pull down menus the type of location (discharge, stream

sample, or other) to query, the year samples were collected, the quarter and month samples were collected, and the water quality parameters. Based on the input criteria, water quality results are displayed in both tabular and graphical form showing the relationship of selected water quality parameters between sample locations. Data can then be printed or exported to a variety of formats for later use in other applications or for presentation.

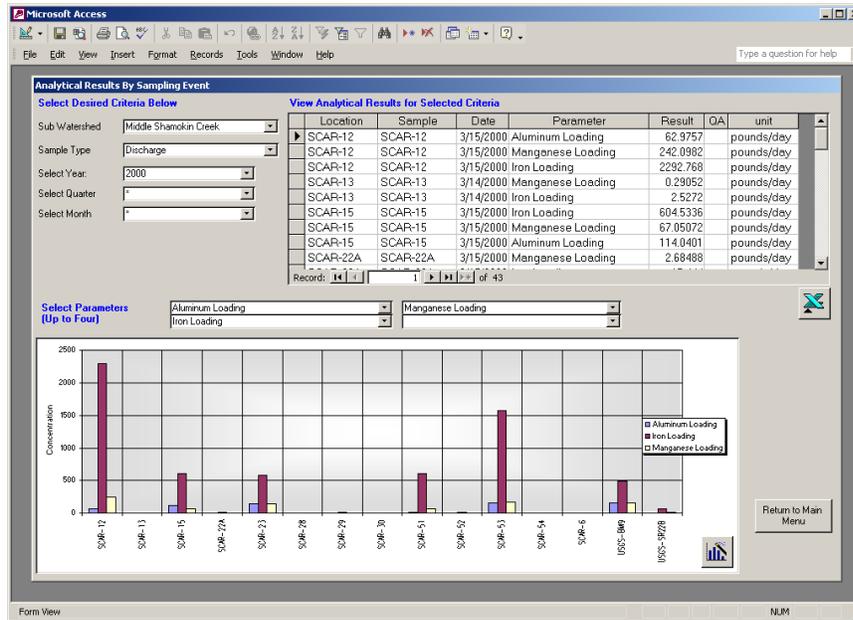


Figure 4 - Example interface showing a comparison of the aluminum, iron and manganese loading at all discharges within the user selected watershed.

Simple statistical analyses and calculations of water quality data built into the database design include output of minimum and/or maximum values displayed as total concentrations or loadings, average values, arithmetic means, geometric means, frequency analysis (frequency of occurrence), and calculation of minimum wetland size for passive treatment by dividing maximum loading by a factor of wetland capacity (Hedin et al., 1994). In each instance, data are selected through the use of a custom interface with selectable options including the watershed name, type of sample location, individual water quality parameter, and sample date range. For example, the end user is able to evaluate average iron loading during low flow events as opposed to high flow events across sampling events conducted in 1998 and 2000.

One other data analysis option includes comparison of water quality data to watershed specific or other criteria. The current database design incorporates effluent limitations for active surface or underground coal mining operations in Pennsylvania (Commonwealth of Pennsylvania, 2000, 2001), calculated stream quality criteria for pollutant loading or total maximum daily loads (TMDLs), if available and continuous exposure limits tolerated by aquatic organisms without unacceptable effect (U.S. Environmental Protection Agency, 2002). The interface allows the end user to select a particular dataset by watershed, sample date range and water quality parameter and compare that dataset to one or more of the available criteria.

#### Report Tables and Other Output

Routines built into the relational database allow the end user to output queried data to hardcopy charts, spreadsheets, or pre-defined report formats. Specific to the system design was the production of a pre-defined report output which resembles standard reporting forms used by the Pennsylvania Department of Environmental Protection (PADEP) in discharge and stream monitoring. Watershed personnel can use these forms to document water quality when submitting for restoration grants and funding.

### **GIS Development**

The GIS component of the EDMS was designed following recommendations contained in the manual entitled *“Building a Geographic Information System for Acid Mine Drainage Remediation Planning”* (Benhart, 1999) to provide the spatial analysis component lacking in the relational database. Environmental Systems Research Institute, Inc.'s (ESRI) ArcView GIS<sup>®</sup> was selected as the platform for the GIS design. ArcView GIS<sup>®</sup> is easy to use, compatible with other applications, and flexible with regard to customization. With the relational database as the digital data repository and primary filtering/querying tool of the water quality data, the design objective was to create a GIS for the evaluation of the spatial distribution of said data, and for site prioritization. Specific GIS development objectives consisted of the following:

- Display topographic, geographic and demographic features of the watershed;
- Display the spatial distribution of discharges and sample locations;
- Provide easy access to the physical attribute information for each location; and
- Analyze the spatial distribution of water quality by location type and parameter.

### Watershed Features and Demographics

Several base data layers were selected for the design of the system. In selecting the base digital data layers, two important factors were evaluated. First, the selected data layers should provide the end user with the basic information needed for data analysis and decision making in regard to watershed planning and restoration without making the management too cumbersome. Base information was selected to include hydrology and geology for analysis and display of surface water systems, watershed boundaries, topography, soil types and rock strata. Road coverages and other infrastructure layers, political boundaries and topographic mapping were selected to provide points of reference for data analysis and later presentation of data. Second, data layers should have similar characteristics. In general, data layers with no map projection and with their map coordinate system in latitude/longitude (decimal degrees) were selected since no additional projection or processing would be required. Using data layers with these characteristics also simplifies the plotting of point data (discharges, sample points, etc.) with coordinates maintained in decimal degrees. In accordance with the above, the base design included the following geographic and topographic coverages.

- Rivers, streams and other surface water features
- Major watershed boundaries
- Sub-watershed boundaries
- Geology (soil and bedrock)
- Land use
- Roads and highways
- County boundaries
- Municipalities
- Urban areas
- Digital topographic mapping (raster)
- Digital ortho-photography (raster)

From this basic setup, additional data layers can be added depending on availability, depth of analysis desired and end user abilities. Some additional data layers found to be useful in the watershed assessment and prioritization included mined areas, mine shaft locations, permitted areas and zoning, property boundaries and ownership, state game lands, parks and forests.

Spatial Display of Discharge and Other Locations

Critical to the design of the GIS is the ability to plot and display AMD discharge and sample locations on the base data layers. To accomplish this goal, the relational database was directly linked to the GIS. Using location coordinates stored in the database, discharges, sample points, and other features are assigned symbols and dynamically plotted in the GIS as new point based data layers. Fig. 5 shows a typical GIS data display of base data layers and the point layer based on database information. In this example, discharge locations and sample locations are plotted based on coordinates and attributes housed in the database. Tabular data at the bottom of the screen displays attributes for each point as stored in the database.

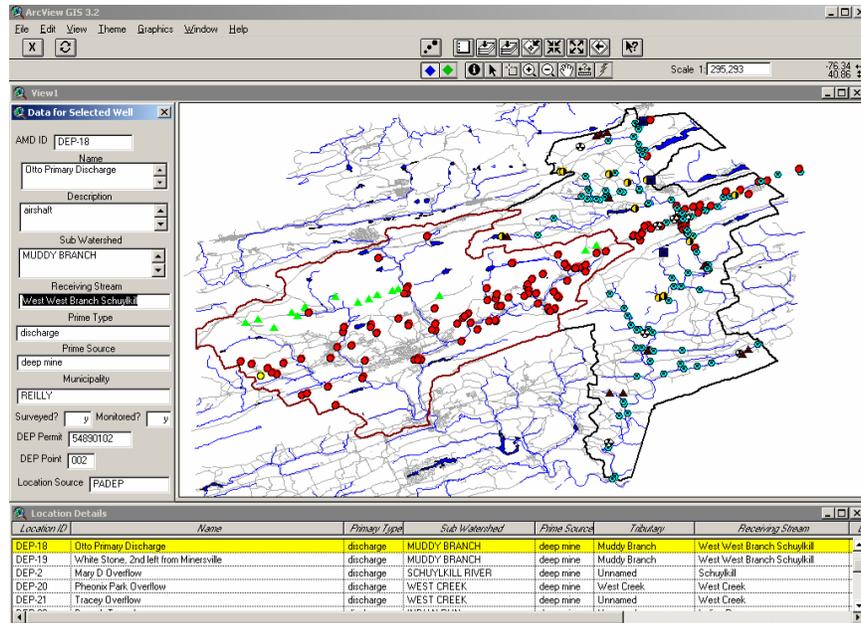


Figure 5 - Typical GIS data display showing a point layer (discharge and sample locations) from the database plotted on base GIS layers.

Two very important advantages were achieved by following this design path. First, all tabular attribute data associated with the new point layers remains housed in the database. The end user is not required to manage multiple attribute tables within the GIS display. All descriptions, labels and water quality data displayed in the GIS View are generated from the database. Secondly, data associated with the point layers can be edited, deleted or added using

the database interfaces described above. When data are manipulated within the database, the GIS is automatically updated with the most recent edits.

### Spatial Analysis of Water Quality

The biggest advantage to utilizing the GIS for watershed restoration planning is the ease in which data can be analyzed spatially. The ability to interactively select, query, and display water quality data in terms of concentration or loading and sample location allows the end user to make informed decisions on prioritization. In order to take advantage of the spatial analysis capabilities of the GIS, the design incorporated a user interface similar to the ones developed for the relational database previously described (Fig. 6).

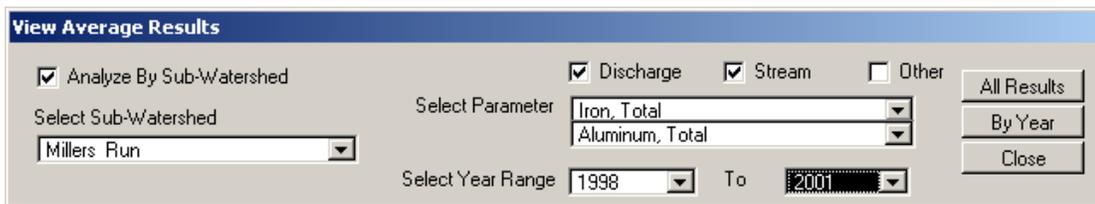


Figure 6 – Example GIS user interface for the selection of parameters to query and display water quality data.

Using the interface depicted in Figure 6, the end user selects whether to display water quality data by a selected watershed or by the extent of the current GIS View window. The user then selects whether they would like to view water quality data from discharge samples, stream samples, other discharges or any combination of the three. Pull-down menus are provided for the user to then select up to two analytical parameters and the date range over which samples were collected. Navigation buttons are provided for the user to choose whether they would like to display average results by year over the selected date range or average results for the entire data set regardless of the selected date range.

Once data have been entered in the user interface and the type of data display is selected, custom scripts are used to query the linked database and provide an interactive display of the spatial distribution of average water quality. Query results are provided in tabular form and as graduated symbols depicting the spatial range of concentrations for up to two water quality parameters. Fig. 7 presents an example output showing the distribution of water quality by

concentration. With this view, the end user is able to quickly identify the worst discharges within the watershed and their relative spatial relation.

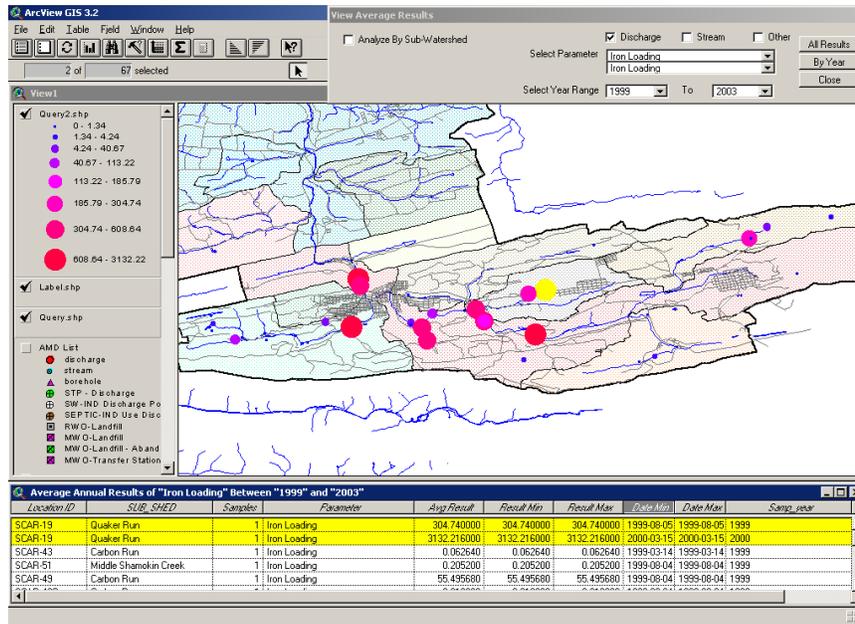


Figure 7 – Example GIS display showing the range of iron loading for discharges within a given watershed. Graduated symbols and colors were used to depict varying loading values.

Tabular data provided at the bottom of the view provides the user with important statistical information including the number of samples used to derive the resultant average results, the date range those samples span, and the minimum and maximum results obtained. With this information, the user can make timely decisions regarding the relative validity of the calculated average concentrations. For example, the user may evaluate the data to decide how best to compare a location with only two samples as opposed to one with ten or more over the same time span.

Using other data layers within the GIS, the user can identify which streams may be most severely impacted, identify upstream priorities as opposed to downstream locations, identify general topography near high loading discharges and make informed decisions regarding the availability of space for passive treatment. If additional property ownership data are available, the user may also identify property ownership at and surrounding the discharge point, gaining a jump start on access permission or land acquisition (Fig. 8). From these displays, the end user is

able to generate scaled, hardcopy maps and figures for sharing information, presentation and grant applications.

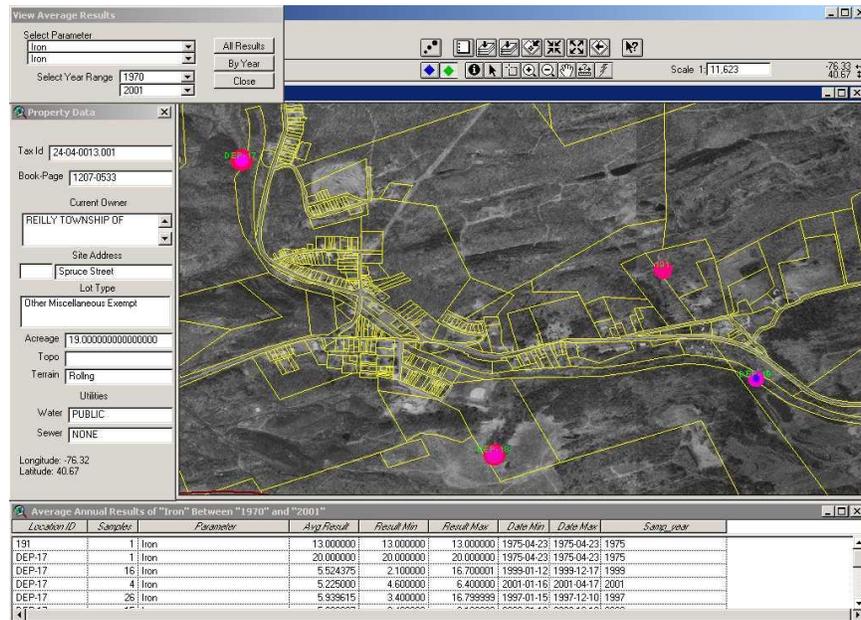


Figure 8 – Example GIS view showing property information and the distribution of water quality.

### Advantages and Limitations

#### Advantages

The EDMS has proven to be a very effective data management and analysis tool for watershed groups. The initial design has been implemented in the assessment of two watersheds in eastern Pennsylvania, encompassing approximately 260 square miles of drainage area. Benefits of using the EDMS to conduct the assessments included the following:

1. The system provides the end users with access to detailed information for over 250 sites consisting of AMD discharge locations, stream sample locations, industrial discharge locations, waste water treatment discharges and other environmentally sensitive discharges and facilities. For these sites, the users are able to access over 3,500 individual water quality samples and over 27,000 analytical results.

2. Data gaps within the watershed were quickly identified, consisting of areas requiring additional sample locations or areas where the current sampling locations were sufficient but the water quality parameter list was inadequate.
3. Initial site prioritization was accomplished on a watershed wide basis incorporating data from several community groups, local agencies, and state and federal government agencies.
4. With very limited training, data input and maintenance of the system were completed by watershed volunteers.
5. Detailed maps, tabular reports and graphs were produced quickly and efficiently for use in prioritization, restoration, as back up for grant applications and for sharing information with the community and local and state agencies.
6. Communication with local community members has increased and educational benefits were realized by using the system to give clear and easy-to-understand presentations of watershed group activities and plans.
7. The system serves as a living document that, with routine updates, will provide the ability to continually check and update the restoration plans and will provide a means to monitor restoration activities.

### Limitations

The design elements of the EDMS and the software platforms chosen provide the opportunity to continually update and customize user interfaces, queries, display features, input routines, and output formats. Taking advantage of the design, new and improved elements of the system are continuously added in order to reduce the limitations of the system. However, no system is completely fool-proof and certain limitations do exist. Some of the more prevalent limitations identified in the initial use of the system included the following:

1. End users must own a version of both Microsoft Access<sup>®</sup> and ArcView GIS<sup>®</sup> to utilize the full function and customization abilities of these two platforms.
2. Sharing of analytical data between the relational database and the GIS resulted in problems with significant figures for water quality results (additional significant figures

added). Although this limitation was not of great concern at the site prioritization stage, data needed for later treatment system design must be more precise. It should be noted that this limitation only existed for tabular data in the GIS portion of the system.

3. The GIS portion of the EDMS required customization on a per watershed basis. This limitation was actually a data availability issue depending on the location of the watershed, the data layers available and the characteristics of the data layers. For example, if only projected data layers were available, some pre-processing of the data would be needed for inclusion in the system. In addition, scripts and queries within the GIS may need modification to update data layer names. The pre-processing and modification would require a deeper understanding of ArcView GIS<sup>®</sup>.

### **Conclusions**

The development of the EDMS for the watershed restoration planning has been successfully applied to two watersheds in Pennsylvania and is being used in two additional assessments. Several advantages were realized over traditional spreadsheets and manual interpretation and evaluation of water quality data. Once data is input to the system, real-time interpretations and evaluations can be made. Data integrity and reliability are greatly increased due to quality assurance checks and balances built into the system. Unlike managing spreadsheets and manually drafted products, which may be manipulated and transcribed by several personnel over the course of an assessment/restoration, the raw data integrity is maintained during the evaluation procedures. In addition, the Microsoft Access<sup>®</sup> database is contained within a single file representing a final repository for all digital data. Similarly, all views and data layers in the ArcView GIS<sup>®</sup> are contained in a single project. Only limited management of shapefiles, which make up the basic geographic data layers, is required.

Finally, the most advantageous characteristic of the system is that it puts real data analysis into the hands of watershed volunteers and community groups. The user interfaces require limited training to master. Initial site prioritization can be conducted quickly and efficiently throughout the life of the watershed restoration. New data can be entered quickly and easily to the relational database with automated updates to the GIS. Once prioritization is completed,

restoration projects can be entered to the database and restoration activities monitored in a similar fashion.

### **Literature Cited**

- Benhart, J. Jr. 1999. Building a geographic information system for acid mine drainage remediation planning: A manual for nonprofits. p. 28. The Center for Rural Pennsylvania. Harrisburg, PA.
- Benhart, J. Jr. and T. Simmons. 1999. Developing a geographic information system (GIS) methodology for acid mine drainage remediation planning. p. 12-17. *In: Acid mine drainage: studies in remediation.* The Center for Rural Pennsylvania. Harrisburg, PA.
- Commonwealth of Pennsylvania. 2000. Chapter 89. Underground mining of coal and coal preparation facilities, Section 52. Water quality standards, effluent limitations and best management practices. Pennsylvania Code, Title 25. Environmental Protection.
- Commonwealth of Pennsylvania. 2001. Chapter 87. Surface mining of coal, Section 102. Hydrologic balance, effluent standards. Pennsylvania Code, Title 25. Environmental Protection.
- Commonwealth of Pennsylvania. 2003. Chapter 93. Water quality standards. Pennsylvania Code, Title 25. Environmental Protection.
- Hedin, R.S., R.W. Nairn and R.L.P. Kleinmann. 1994. Passive treatment of coal mine drainage. U.S. Bureau of Mines Information Circular 9389, Department of the Interior, U.S. Bureau of Mines, Washington, D.C.
- U.S. Environmental Protection Agency. 2002. National recommended water quality criteria, 2002. U.S. Environmental Protection Agency EPA 822-R-02-047.