

ARE STATEWIDE OR REGIONAL DATABASES *REALLY* POSSIBLE?

Barb Horn, Co-Chair Colorado Water Quality Monitoring Council
Colorado Division of Wildlife
151 East 16th Avenue
Durango, Colorado 81301

ABSTRACT

Have you heard the suggestion “what we need is a statewide/regional clearing house or database?” Have you wondered why every state doesn’t already have one? How could something that makes so much sense not exist? We don’t mean a database provided by state agencies with only their data, but a database that serves multiple data generators and users. One that has EPA STORET data and MORE! What we want is an information system, the ability to share information so effectively we are operating and maximum efficiency and make the best possible water management decisions. A database is a crucial piece of the pie of an information system but not the pie itself. Hurdles must be too high if such a common need exists and is not in place in most states.

In Colorado we found two things to be true. First, indeed resource and capacity hurdles are too high for any individual group to create and support a statewide database and yet; many groups were reinventing the wheel at smaller scale and devoting large amounts of resources to creating a unified database. The need is there and it involves more than a database. The Colorado Water Quality Monitoring Council (CWQMC) is successfully making the case that resources could be reallocated to support a true statewide information system for less than groups are expending on smaller efforts, liberating resources to be used for more data analyses, collection or other priorities, and in the process provide more data for each individual group.

The CWQMC is in the second year of a three year grant to implement the Colorado Data Sharing Network (CDSN). The project has three components each with a different target audience. The database itself uses STORET format and templates for uploading, has additional features and can handle chemical, biological and physical data from rivers, lakes and groundwater. Next is the web-based ArchIMS map, which allows anyone to tell everyone what they are doing, when, where and how regardless of where the data actually lives. Direct links to other databases are used where possible. Finally, the most important component is our watershed SWAPS or placed based exchanges of monitoring priorities, needs and concerns. This third component includes outreach and sustainability strategies.

Learn the pro’s and con’s of this approach, the resources it has taken us to start and sustain this effort, and tips on how you can make this happen in your state, region or watershed.

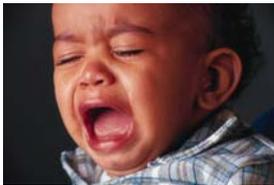
KEYWORDS

Monitoring, network, outreach, statewide database, regional database, statewide monitoring councils, STORET, ArchIMS, web-based maps, clearing house, data management, sustainability, monitoring priorities, data storage, and data retrieval.

INTRODUCTION

Have you ever heard the suggestion “what we need is a statewide/regional clearing house or database”? Have you wondered why every state doesn’t already have one? It makes so much sense, one place everyone could go to get data. We don’t mean a database that is provided by a state agency with their data only, but a database that serves multiple, even all, data generators and users. We mean a true statewide information system where data and relevant monitoring and assessment activities, priorities, needs and concerns can be exchanged in a timely and meaningful manner at local and state levels. There must be high hurdles if such a broad based common need exists and is not yet in place in most if not all states.

Figure 1 - Typical reactions to the term “EPA STORET” or when asked “would you develop a statewide database?”



In Colorado we found two things to be true. First, indeed resource and capacity hurdles are too high for any individual group alone to jump and yet; many groups were reinventing the wheel at a smaller scale and devoting a large amount of resources (money and time) to creating a unified database. This demonstrates the need. The Colorado Water Quality Monitoring Council (CWQMC) is successfully making the case that those resources could be reallocated to support a statewide database and information system for less than groups are expending on smaller efforts, allowing liberated resources to be used for more data analyses, collection or other priorities, and in the process provide more available data for each individual group. We had to find the selling points from their point of view for actual participation, even though everyone wants the desired the end point. Once a critical mass is using the system, the rest join from peer pressure or because this is now the best feasible option. The CWQMC is finishing the second year of a three year grant to implement the Colorado Data Sharing Network (CDSN) and has begun to concurrently implement their sustainability plan to carry this project beyond the initial grant.

Who is the CWQMC and how did we morph this support?

This piece of history is important to how we got where we are today and hopefully will save you time and energy. The CWQMC informally formed, under a different name, twenty years ago when the era of decreasing budgets, increasing Clean Water Act responsibilities and complexity, long drought periods and rapid growth in population and resource extraction industry all began. In other words the pressure to collaborate and be more efficient and effective hit a threshold across the state.

Since the beginning the CWQMC has been and continues to be a diverse group of volunteers, all volunteers. Participation includes Forest Service, Bureau of Land Management, Colorado Division of Wildlife, Colorado Public Health, Waste Water Utility Council, Water Supply Utility Council, municipalities, counties, special districts for water, irrigation and soil, multiple non-

profits, industry representatives and consultants. Faces of the Council have changed over the past 20 years but the group has had the common goals to share data within and across watershed and political boundaries and increase on the ground monitoring collaboration.

Efforts evolved through three eras in order to arrive at the current project. In the 1980's the focus was on standardization of field and laboratory and methods. This failed because no entity was willing to change their methods because they had 20 years of data at risk or their needs varied too much. This evolved into the 1990's focus to just exchange data and monitoring locations. This too failed for a variety of reasons some of which included:

- Data was in a variety of formats, structures and nomenclatures, sometimes not even electronic
- Entities were not gathering or recording sufficient information about the data (meta-data)
- Trust, not everyone wanted everyone else to have their data
- Entities did not want to have to manage their data in two locations and two different ways if we required a certain format, it would not be compatible with their current management scheme
- Time, people wanted others data but didn't want to take the time to organize their own to share

Our initial failure moved the group to try just exchanging meta-data, information about the data such as what they monitored, where, when, why, how and providing contact information to access the data. We called these Watershed SWAPS. We piloted a SWAP on a 14 digit hydrologic code watershed size, Clear Creek that follows I-70 from Loveland Pass into Denver. Then we tried a larger 10-12 digit hydrologic code, the South Platte River that runs through Denver. Both of these worked well if conducted on a watershed scale versus statewide. During these pilot swaps we also shared watershed monitoring priorities, needs and concerns and were also able to develop on the ground monitoring collaboration efforts. The Council took this concept and developed it into the current Data Sharing Network (CDSN).

At this point funded a facilitator via donations and had a \$15,000 EPA Grant. We were going to develop a static map that would illustrate what is happening, where, when, why, how and provide contact information for that amount. That is when we began to leverage the need for our non point source program in Colorado, as well as all other states, to get legacy data into EPA STORET. Colorado had over 100 projects that needed to get their data into EPA STORET since 1990.

METHODOLOGY

Concurrent to the need and existing efforts was the need for funding. Up to this point the Council was basically volunteer and had solicited enough funds to hire a facilitator that helped us get to this common point as well as conduct initial SWAPS. Our funding helped us define what we wanted into what we could actually achieve.

Funding

We started with a \$15,000 from EPA Region 8 Geographic Initiative and then creatively found 319 Non Point Source Funds of \$100,000. Our timing was impeccable. In the late 1990's the

USEPA began a concerted effort to get states to comply with importing all NPS project data into the National EPA STORET (STOR=Storage and RET=Retrieve) warehouse database. The Council convinced Colorado's NPS Program Leadership to partner and create a system that would benefit all data generators, not just NPS Project data. Thus, the target audience for CDSN became

- 1) Legacy and future NPS project sponsors and
- 2) Any other entity collecting watershed data that was either not organized or not being shared beyond the initial project.

An example of an already organized database would be the USGS National Water Information System (NWIS) or a group who already submitted their data to EPA STORET, such as our state health department. These databases are already organized and available. By the way the monitoring and assessment units of Colorado Public Health and Environment ***DID NOT*** support our project. Their fear was they would have to take responsibility for the end product and service and their own data was still not organized and in STORET.

Once the funding was in place, the Council developed four major components to the Data Sharing Network:

1. To establish a water quality data management system that meets the needs of Colorado's Non Point Source Program and local data providers of data sets not already organized and widely accessible. *Water quality data includes chemical, physical habitat and biological data for rivers, lakes, reservoirs, wetlands and ground water.* The hardware and software system:
 - a. needs to be able to have simple uploads, meaningful downloads, ability to conduct simple summary statistics
 - b. must use existing available technology at low or no cost, and be available via internet
 - c. employ a minimum set of data elements (specific fields) for each area
 - d. place data accuracy and quality responsibility on the generator not system operators
2. To establish a web-based map, ArchIMS that would serve as comprehensive directory of who is doing what, where, why, how and contact information. The actual data would be in our system if possible, if not a direct link to the data would be provided or contact information is provided.
3. To actively put data into the system via strategic training and outreach for three years
4. To develop and implement a long term support strategy for the system, training and outreach

This set the stage for our work plan and we set out in three years to simultaneously develop the web-based water quality directory map, a database, outreach and conduct SWAPS statewide (about 16 events) and populate the system, and our sustainability strategy to be implemented before the end of the three year grant cycle.

We also needed some guiding principles to make decisions by and those came from our previous failures and successes. We could only accomplish this by using existing technology. This type of system would cost too much to build from scratch and we would have to make all the mistakes someone has already made. So we collaborated with EPA Region 8 STORET staff to use their existing hardware, software, documentation, tools and expertise. We knew for us that the system

hardware and software must include:

1. EPA STORET database structure – STORET has 20 plus years of developing a relational database structure, standardized formats and upload tools to deal with many types of data in many formats. Any data STORET can manage we can manage and we didn't have to expend resources to get a functional data storage tool. STORET stores data well, will be around along time, makes our database compatible with our state health department, and is where NPS data needs to end up, so why not use it? This strategy also has the benefit to allow our system to contain all STORET legacy data *in addition* to other data and thus our system is the primary Colorado data source, larger than even STORET is for Colorado.
2. A mechanism to include more function than just store data, so we developed a simple set of templates to upload data, and user friendly functional data downloads. We will add the ability to conduct simple statistics and graphics. This aspect includes developing an initial minimum set of data elements (fields) for water chemistry, macroinvertebrate/fish data and physical habitat data.
3. A mechanism for people to get legacy data into the system as well as their new data. People will typically not enter data into two systems, so we will help them do this while building our critical participation mass.
4. Addressing security concerns, so we have three levels of security, password protected, all controlled by the data generator. Data can be available to all or not, the user can decide if the data will be uploaded to EPA STORET or not (unless the data is a NPS project).
5. A way to assess the appropriate use of others data. We don't get involved in judging the data quality or purpose, similar to STORET, but require that entities provide that information so other users can adequately determine the use.
6. A visual map to determine what data exists where they might have data needs. We are using an ArchIMS map developed in EPA Region 10 and replacing the underlying layers with Colorado information. This map will serve as a directory for all activity in the state, regardless of where the data actually lives, displaying what data is available, where, when and how collected. If appropriate the data will be in our system, if not there will be a direct link or information on where to get the data. This also allows people to "privately share" if trust is their sharing issue. They can tell people about their data but select who they give it too.

We had a work plan and an outreach plan and began developing and implementing simultaneously all components of the project. This was done by a technical advisory committee.

RESULTS

How have we achieved these four components? Synergistic collaboration by a core group that held the work together long enough to get all four areas to critical juncture where others can support the effort. Just starting our second year of the three year grant, we have:

1. Conducted 16 SWAPS statewide, documenting these via basin fact sheets that summarize the activities in the basin, list priority, needs and concerns of all entities. These are posted on our website providing a voice across watersheds and from the local to state levels.

Participants from every SWAP all stated they learned something and that these events need to continue.

The target audience for these events is anyone involved in water management. We believe we have reached the target audience that has any capacity to utilize our product and services. We estimate this is about 80% of that total audience.

2. Developed an interactive web-based map that now is a state-wide water quality directory of who is doing what, where, when and how. When possible we direct link to data not in our system. The target audience for this component is any data generator of water information. That includes chemical, physical, biological information on rivers/streams, lakes/reservoirs, wetlands and ground water. We believe we have reached 90% of all data generators on average in each basin.

This component has become more valuable than the database because it can be a one-stop-shop where as a database will never succeed at that task. Existing organized and available databases such as the USGS NWIS, or EPA STORET or state health department data can be accessed via the map with out dealing with duplication in our database and their existing residence. The disadvantage of this is that the task of pulling data from multiple sources and combining them into one dataset still exists, but you can go to one site to find all existing data. This component also allows for entities to “privately” share data if they do not want to have data in our database.

It is the vision of DSN to generate funds to employ USGS or other entities to conduct annual “data analyses” on a rotating basin. This task would involve combining data sets, and multiple stations into one, then providing an overall gap analyses for all users. The map is functional but not quite where we need it to be. Plans are to continue to upgrade this aspect.

3. Developed a database with input and output functions to serve as a data manager for small unmanaged, unshared data sets or entities that have a need to get their data into STORET. We believe we have identified all existing datasets and are working to get them into the system.

This brings up the issue of legacy data. A group wants all their data in one place, so in order to get entities to put new data into a system we help them get all their data in as well. The idea is to not have a group manage their data in two places or change the way they manage their data. We instead, create the up front templates to take their management system and translate to ours. Once this is done uploads are easy.

This system puts the responsibility of data accuracy and quality on the data generator. As with STORET, DSN does not judge the data but requires minimum data elements so any user can determine the appropriate use of the data. DSN minimum data elements list is much larger than STORET and exists for all media and water body types. The system also has three levels of security where the generator determines when each level is enacted. Data in DSN can be uploaded to National STORET but is not required to be.

DSN had to manage the STORET transition from WebSIM to WQX as well as part of this project. DSN has a seat at EPA Region VIII equal to a state agency. We are going to help the regional effort build components that the EPA cannot assist with this transition. This aligns with our leveraging existing resources and technology principal.

4. We have a sustainability plan we are implementing. This plan re-organized the Council structure and function to support DSN long term. We have a Leadership Team that is composed of a representative from every major basin in the state. For the first time we really are a state-wide council. We also developed a sustainability plan to transition DSN from the initial grant to beyond. This plan assessed costs and based on feedback a funding strategy that is inclusive, utilizes a fiscal agent versus becoming a 501(c)3 and a sponsorship / grant strategy. An annual budget and work plan is developed by the Leadership Team and subcommittees that determines a dollar goal that is to be raised during the current year for the next year operations.

The Colorado Department of Public Health and Environment provided the initial \$100,000 for this project, the EPA RGI Grants about \$15,000. In the future our estimated annual budget at this time is about \$135,000. Of that total the Colorado Department of Public Health and Environment will be supporting 20%, in-kind match is about 40% and cash from other sources is about 40%.

Hard cost allocation starts with providing a server to host the ArchIMS map for one year and cost \$20,000. Administration, fiscal agent fee (reduced) and providing training, database management, some outreach and support and system operation and maintenance are the other hard costs. We propose conducting two SWAPS (8 events) annually, every fifth year a workshop. Most of the Council in-kind match helps here and is about 30% of total costs.

DISCUSSION

A great database system is useless unless there is a way to get people to use it (outreach), competently and consistently (training and user support), foster a forum for groups to talk about their monitoring priorities and needs (sharing) and trust that the system will be here for a while (sustainability strategy). DSN recognized this as a key reason that STORET was not used more widely in Colorado, along with the technical and fiscal barriers to sharing data. To address these needs we:

1. Have provided user documentation on our website and developed an outreach training schedule that will provide a week of training in each of the major basins within two years. Each training session is only one day. We plan to continue these trainings on a rotation schedule beyond the grant cycle. During this funding cycle we have the benefit of literally formatting templates for data sets for groups while teaching them how to do it. This work occurs in between training sessions. We also provide annual "data calls" to remind people to enter their data into the system if they are not using it as their routine data manager. One paid staff (during the grant and most likely post grant) and 2-3

volunteers do this work, while a team of 6-10 of us conducts the outreach and marketing for each basin sessions.

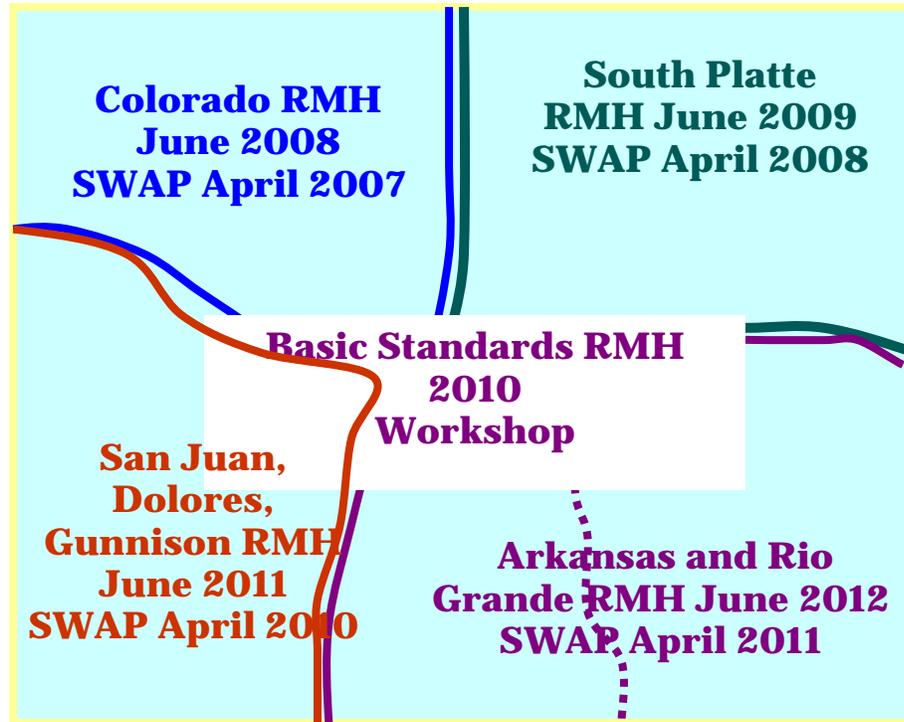
2. Created Watershed SWAPS. The morning before each CDSN System training we host a SWAP where any entity involved watershed management swaps their monitoring priorities, issues and concerns. After each SWAP we produce a series of fact sheets documenting and summarizing the discussions. We started in the Colorado, are currently doing the South Platte. Next fall will be the San Juan and Dolores Basins followed by the Arkansas and Rio Grande in the spring. We will continue these SWAPS and trainings after the grant is over, clearly the demand is there. You can download our outreach strategy documentation from the website. We are in the middle of outreach, training and developing the system so you may not find everything on our website, but it changes frequently.

SWAPS are part of our sustainability strategy also. This forum is providing a local voice among local entities that don't always communicate. In addition SWAPS are providing a voice from the locals to state entities who do not reach that scale AND a mechanism for state entities to reach the local scale. As state, local and NPS funding continues to shrink in comparison to the amount needed to eliminate pollution from non point sources, as states implement drinking water protection plans, watershed plans and weather patterns create water crises, collaboration is a critical strategy.

Timing is part of successful participation. We planned the training sessions and SWAPS around the Colorado Department of Public Health and Environments monitoring schedule for their Clean Water Act Major Basin Rule Making Hearings. In these hearings they determine appropriate classifications and standards for each basin, rotating around the state every five years.

This strategy has garnered long term support from Colorado Department of Public Health and Environment because it aligns with their data calls and decision making processes, saving them resources.

Figure 2. Example of CWQMC Data Sharing Network SWAP Exchange with the Clean Water Act Basin Rule Making Hearings in Colorado.



3. Made developing and implementing a sustainability strategy a primary task of the current grant. We are gathering the financial information to sustain the hardware, software and technical support while we increase system participation. Concurrently, the Council is developing a fiscal capacity and “membership or user” fee based strategy to fund this in the future. We do not want fees to be a reason a group cannot participate. We are very optimistic we can raise enough funds through the demand for the product and service that will support groups have financial challenges. The Council raised \$20,000 or more dollars when we had no product or service and we already have key data generating groups stating their financial support, documenting the demand. At this time it appears there may be a fee for data generators versus users, but the price will not be prohibitive to participation. We also plan to continue using existing resources to help support the system for example finding a server host.

Our prediction is that the Colorado’s Department of Public Health and Environment will be using *OUR* system to upload their data to STORET and require others to use it for data they will use in their Clean Water Act decision processes. This is a way that more data will be available to make better decisions for everyone. However, if the data is not sufficient quality now for their decision process, this data base will not help that issue.

CONCLUSIONS

This model may not work for all states, but may for some or some larger watersheds.

Tips from DSN:

- It is all about diverse partnership, collaboration that sticks together long enough to get it going
- Leadership, a core group of people will need to carry this through to the next level, sharing the load
- Develop and document your plan that will get a critical mass to participate so you can demonstrate success and sell it, *messaging* is a key aspect to success
- Don't reinvent the wheel – for the system, outreach or training, just modify and tweak to meet your needs
- Begin to quantify what people spend (or should spend) on data collection and management so you can show them how they are saving money. Many groups hire this task and thus we are targeting consultants as well because they should be telling their clients to use this system so they can do more data and gap analyses, etc. Dollars speak to decision makers.
- Figure out the right scale and scope to start. We are starting with water chemistry, macroinvertebrates and physical habitat so that we appeal to a variety of entities; they are the most common data areas and have the most data sets in Colorado. We plan to add, periphyton, fish tissue, ground water and other areas as we get priorities from conducting basin SWAPS. We are also outreaching and training one basin at a time to complete the state in a three year period.
- Include a sustainability strategy in the effort. If you cannot demonstrate you will be around awhile people will not invest time or data. We have groups like the Waste Water Utility Council offering \$20,000 to support this effort. The need is there if you can get beyond the differences.
- Identify the common set of needs, prioritize and communicate what you will be addressing
- The system will require a server to host the ArchIMS map and data sets; our strategy is to use an entity that already has that capacity but does not have fire walls that are prohibitive like most federal agencies and many state entities.
- Have the system and process hold the data generator responsible for the upload and quality of data – it becomes an unmanageable task when responsibility is with someone who knows nothing about the data
- People don't like to change, even if what you have to offer is better, figure out what their needs are and sell this to them from that perspective, for us that was they were spending the money for the same thing on a smaller scale and struggling. Our offer was to join us first for free see how much more you will get then help us support it.
- Will this cost you \$115,000? It depends on your scope, what you re-invent or use and how much collaboration exists, this is not a high price tag for the product and service

ACKNOWLEDGMENTS

Gracious acknowledgement to the core team of dedicated individuals and organizations that have persisted in manifesting the Colorado Data Sharing Network. They include Laurie Fisher, Colorado Department of Public Health and Environment, Vic Lucero - City of Thornton, James Dorsch – Metro Waste Water District, Sarah Sauter and Jacob Bornstein – Colorado Watershed Network, Marty McComb and Karl Hermann – USEPA, Dave Litke – USGS, Esther Vincent – Northern Water District, Dave Kanzer – Colorado River District, Sarah Reeves – South Platte CURE. Special thanks to the Colorado Division of Wildlife to support this vision through my time.

REFERENCES

1. For more information the website www.coloradowatershed.org/CWQMC has the:
 - CWQMC Charter
 - Outreach plan
 - Sustainability plan
 - SWAP Fact Sheets (as they are completed)
 - User documentation
2. To visit our interactive web-based map and/or database go to: www.codsstoret.com
3. For more detail on project components, grant proposals, process, work plan, costs, updates or other question contact:

Barb Horn, Colorado Division of Wildlife at barb.horn@state.co.us