Tips and Tricks for Setting Up STORET Data

STORET

Tips and Tricks for Setting up STORET Data

Introduction

One of the most important features of STORET as a data management system is its ability to maintain data of documented quality. Properly documenting your data by providing rich metadata (i.e., data about data) can ensure that monitoring results will be used appropriately. In STORET, much of this metadata needs to be set up in the system up front, before you begin to enter monitoring results. STORET was designed to be very flexible so that varied ways of documenting and describing data can be accommodated.

Because of the many ways STORET can be set up, the task can seem daunting to first-time users. To provide useful insight to new users, we conducted a survey of current users asking how they implemented their initial setup and entered metadata. Thirteen users from different types of monitoring organizations responded to the survey. The purpose of this document is to provide helpful ideas for setting up various aspects of STORET. This information was compiled from the survey responses and is designed to help you customize STORET for efficient data import, retrieval, and display.

We asked the respondents to comment on various types of STORET data, including

- Water chemistry,
- Sediment chemistry,
- Tissue chemistry,
- Physical measurements,
- Habitat,
- Biological information (both raw and metrics data), and
- Data loggers.

Although the survey respondents enter data in all of these categories, most submitted input on water chemistry, physical measurements, and raw biological data.

Before you enter information into STORET, it is important to take the time to analyze your data and decide how you want to enter it. You will also want to consider what data you want to get out of STORET. Determining how to organize and enter data before you start may save you many hours of reformatting and processing the data. While it is useful to determine how to organize your data before entering them into STORET, it is also important to be flexible, since you may have different types of data to put in and get out of STORET.

In general, it is helpful to structure the data you plan to enter into STORET in a fashion similar to your current data analysis system—this minimizes time to rearrange or process data. You can use the information in this document to help you find a data situation similar to your own and determine how someone else approached the problems and set up their data for STORET. It is up to you to define a system that is useful and efficient for your situation.

This document is divided into sections according to the main information categories in STORET: **projects, stations, station visits, trips, activities, characteristic groups, lab/field analytical**

procedures and equipment, and sample collection procedures. Each of these information categories provide important metadata for results you will enter into STORET. The next sections cover each of these categories, describe options for defining them in STORET based on the experiences of the survey respondents, and list advantages and disadvantages of the various approaches. For more details, see Appendix A, which contains the complete text of all responses to the survey.

Projects

Project descriptions contain essential information about purpose, procedures, standards, and methods related to the project. The advantage of using the projects in STORET is that they allow you to retrieve data sets that are specific to a project. For example, by associating sampling stations with a lake monitoring project, a staff member can go into the database and retrieve all sampling data associated with the specific project for analysis.

A project in STORET must include the following elements at a minimum (note that other metadata can be stored as well):

- Unique Identifier (up to 8 characters long),
- Start Date,
- Duration (enter "ongoing" for continuous monitoring),
- Name (e.g., "Alameda Creek Volunteer Monitoring"), and
- Purpose (e.g., "Monitor effects of urban development on water quality of Alameda Creek").

There are many ways to set up projects in STORET, each with advantages and disadvantages. The survey respondents reported setting up anywhere from 2 to 1,000 projects, based on how their projects are defined. Table 1 shows the different ways the survey respondents have set up projects for STORET.

Table 1. Options for Defining Projects in STORET

Options for Defining Projects	Advantages	Disadvantages
Program Name or Study [1*, 2, 4, 9, 13]	Unambiguous[1] Easy to retrieve data from STORET[1, 13] Less initial formatting[2] Data owners can define data organization[4] Quick to retrieve project-specific data[4]	May limit retrieval options[9]
Year[3]	Matches USGS data "water year" books[3] Easy to retrieve data from STORET[3]	Difficult to deal with multiple years[3]
Program Purpose[5, 8, 10]	Can track separately[5]	Difficult to categorize data that fulfill more than one purpose[5]
Program Name and Year[6]	Easy to retrieve data from STORET[6] Easy to see any changes to stations[6]	
Hydrologic unit code (HUC) or Watershed[3, 7, 12]	Forces a QC check of HUCs[7]	Does not take advantage of STORET metadata capabilities[7] Requires more cooperation between counties[12]
Collection Entity[9]	Works well when multiple entities are involved in one project[9]	
Location and Date[11]	Tracks environmental problems by location and time[11] Data can be sorted different ways[11]	

---

* Numbers in the tables correspond to the number assigned to each survey respondent and identify the respondent(s) associated with an option or response. The full text of the specific comments can be found in related sections of Appendix A.

Stations

Stations, also referred to as sites, identify or describe the physical location at which monitoring occurs. All data collected in the field are linked to a specific location, or site, at which the field work was conducted. Recording station data links water quality measurements to the place they represent.

Precise location definition is very important to environmental analysis, and EPA data standards for locational data are strictly followed in STORET. All applicable federal standards (e.g., Federal Information Processing Standard [FIPS], National Institute of Standards and Technology [NIST], and others) are adhered to wherever possible.

Stations may be part of external reference schemes, and may carry a multitude of identifiers from each of these schemes. For example, a station in STORET might have a National Pollutant Discharge Elimination System (NPDES) number or a state regulatory program code.

Each monitoring station must include the following elements at a minimum (note that other metadata can be stored as well):

- ID Code (up to 15 characters, e.g., for Alameda Creek Volunteer Monitors, stations could be labeled "AC-001, AC-002, etc.),
- Station or Waterbody Type (e.g., Stream, Lake, Well, Estuary),
- Latitude and Longitude (does not have to be accurate to nearest square inch but try to be as precise as possible),
- Geopositioning Method (e.g., GPS, map interpolation),
- Datum (e.g., North American Datum 1983, etc.), and
- State and county.

The number of stations reported by survey respondents ranges from 100 to over 18,000 depending on how the stations are set up. Table 2 shows different ways survey respondents have set up naming conventions for STORET stations and the advantages and disadvantages users described for each option. Please note that, in the examples, site IDs and station numbers are synonymous.

Table 2. Options for Defining Stations in STORET

Options for Defining Stations	Advantages	Disadvantages
Each individual program decides how to input data in STORET [1, 10]		Difficult to locate information [1] No consistency [10] Some state agencies prefer direction [10]
Program and Site ID [2]	Easy to import into STORET [2]	
Reach IDs from Legacy STORET [3]	Consistent with Legacy STORET [3]	Difficult to calculate and retrieve [3]
County and site [4]	If site IDs are set, easy for data collectors [4]	Easy to have duplicate IDs for streams [4]
Ecoregion [5]	Helps with biocriteria development [5]	Difficult to separate combined biological, chemical, and habitat data from chemical-only data [5]
Project, county, sequential number [6]		Easy to have duplicate IDs [6]
HUC, waterbody ID, number [7, 10]	Can track by watershed [7]	Must verify and QC each HUC [7] Easy to have duplicate IDs [7]
Stream names [8]		Too long for STORET [8]
ID from state Laboratory Information Management System (LIMS) [9]	Can easily link to source data [9]	Not useful for querying based on topics [9]
Project ID, waterbody ID, site ID [11]	Easy to understand [11] Can track by waterbody [11]	Difficult to track multiple waterbodies [11]
Sequential numbers [12]	Easy to assign numbers to new stations [12]	Not always possible to assign sequentially, so one waterbody may have multiple numbers [12]
Waterbody ID, site ID [13]	Easy to determine location [13]	

Trips

A field trip is a method of grouping actual visits to monitoring stations. One trip could involve a single visit to a single station or multiple visits to several different stations. Trips also provide a framework for storing "blank" samples and other QC activities.

Trip information includes the following at a minimum (Note that other metadata can be stored as well):

- An ID code for the trip,
- Date the trip began, and
- List of projects supported by the data collected on the trip.

It may be useful to define trips by a combination of geography and time. For example, the Nevada Division of Environmental Protection (NDEP) monitors throughout the State of Nevada. Nevada is divided into seven major river basins. A trip is defined as all visits to the stations located within one basin for an entire year. For example, the trip labeled "Carson1999" included all the monitoring stations visited in the Carson River Basin for the year 1999.

Table 3 shows different ways survey respondents have set up trips for STORET.

Table 3. Options for Defining Trips in STORET

Options for Defining Trips	Advantages	Disadvantages
One year of sampling for a particular program [1, 10, 12]	Easy to understand [1]	
"T" plus the station number [2]		May need to reformat data [2]
Year [3, 4, 7, 8]	Users often want to view one year of data [3] Limits the number of trips for easy data management [3] Helps organize and load data [7] Easy to see if data is already in STORET [7]	
Field crew, year, and week [5]	Easy to find date by year and field office [5]	IDs can be too long [5] Time consuming to look up the week a trip started [5]
Project and date [6]	Easy to find data for corrections [6]	
One trip per day, month, or year [9, 11, 13]	Accurate representation of actual trips [9] Can be auto-generated [11] Easy to find data from a particular time period [13]	

Station Visits

Station visits are the events that occur when a particular site or station is visited to conduct monitoring activities. Any number of activities can be done during a single visit. For example, during a site visit, one field observation activity could include measurements of water temperature, dissolved oxygen, and pH. Another field observation activity could involve measuring vegetation cover as part of a habitat assessment. Sampling activities could include collection of a water sample or the collection of fish for tissue analysis. Station visits in STORET allow the user to track the frequency of visits to sampling stations.

Visits consist of the following information at a minimum (Note that other metadata can be stored as well):

- Date and time of the visit,
- Station being visited, and
- Visit number.

A station can be visited any number of times during a single trip. However, the visit number must be different for each visit to a station.

Table 4 shows different ways survey respondents have set up station visits for STORET and the advantages and disadvantages users mentioned for each choice.

Table 4. Options for Defining Station Visits in STORET

Options for Defining Station Visits	Advantages	Disadvantages
One station visit per day [1, 4, 10, 13]	Easy to understand [1]	
Sequential number [2, 12]	Creates a unique ID when combined with date and time [2]	May be duplicates later [2]
Allow STORET Import Module (SIM) to assign [3, 6, 11]	Can assign by sampling date and time [3]	Resampling will create duplicates [3]
Date and station ID [5]	Easy to understand [5]	
Assign number based on visit spanning several days [7, 8]	Good for tracking multiday visits [7]	Various activities on different days may not appear to match [7]
Station, date, time [9]	Accurate tracking [9]	

Activities

Activities define a task accomplished during a visit to a monitoring station. Activities include collecting samples, taking field measurements (including habitat assessments), and making field observations. Activities also document information about the sampling process, including collection methods, sample preservation procedures, and personnel performing the activities. Activities can be associated with specific monitoring projects.

Activity information should include the following at a minimum (Note that other metadata can be stored as well):

- ID code of the activity (for a sample, it may be helpful to make the ID the same as the sample code),
- Activity type (e.g., sample, field observation/measurement [includes habitat assessment], or automatic data logger results),
- Medium (e.g., air, water, soil, biological),
- Date and time of the activity,
- Activity category (e.g., other sample information, such as routine sample, composite, or replicate),
- Activity location (monitoring station where activities occur), and
- Collection procedure (for samples only).

Table 5 shows different ways survey respondents have set up activities for STORET.

Table 5. Options for Defining Activities in STORET

Options for Defining Activities	Advantages	Disadvantages
Combine field measurements and lab analyses and assign numbers [1]	Efficient for data entry [1]	Difficult to track parameters in STORET [1]
Month, day, and year (field and lab results are separate) along with parameter code [2]	Keeps activities separate and unique Easy to determine which data were analyzed in the field or the lab, and whether a result was created from a sample or a field measurement [2]	
Field and lab results combined for U.S. Geological Survey (USGS) data and separate for other data [3]	Separate results helps track and match field and lab data [3]	Must know if trips and visits are new or existing for entry into STORET [3]
Lakes: Sample location, depth, field or sample (F or S), replicate (R) Streams: Lab sample number and descriptive code [4]	Streams: system works well [4]	Lakes: depth difficult to make unique [4] ID may be too long for the activity field [4]
Biological (B), chemical (C), or field (F) plus a unique number [5, 11]	Clear [5] Easy to group field or chemical data [11]	Field and chemical samples are separate in STORET, but must be considered together during analysis [5]
Three different activity IDs: water conditions, atmospheric conditions, and water samples [6]	Can help find data in STORET [6]	
Unique activity ID: year, trip ID, sequential number, and a suffix representing a medium code or sample type [7]	Works well [7]	
Field measurements separate from lab analyses [8, 9]	Can track field and lab data separately [9]	More complicated to keep separate [9]
One activity for date, time, depth, activity type, and category [10]		
Assign field measurements the lab ID for a station with an "F" [12, 13]	Easy to determine what field conditions existed when samples were taken [13]	

Characteristic Groups

Characteristics are things that are actually measured and analyzed, for example, water temperature, pH, arsenic, lead, DDT, total nitrogen, etc. You can set up characteristic groups to help group together the characteristics you use frequently. Characteristics can be grouped by medium, activity type, or any other useful category. Using characteristic groups allows you to enter data with similar metadata as a group rather than providing the metadata for each piece of information.

Characteristic groups can be most helpful when using the batch entry function of STORET. In addition, you can use characteristic groups to assign different metadata to a characteristic that is collected in two different ways. For example, if dissolved oxygen is analyzed by two different procedures, you can use a characteristic group to assign different metadata to each set of dissolved oxygen results.

The information associated with each characteristic is dependent on the type of characteristic, but generally the following are needed at a minimum to set up a characteristic group (Note that other metadata can be stored as well):

- Group ID,
- Group name,
- Medium (e.g., water, biological, habitat assessment),
- Activity type the characteristic will be associated with (e.g., sample, field observation, automatic data logger),
- Characteristic name (select from STORET),
- Units of measurement for the characteristic (e.g., μg/L, count, percentage), and
- Analytical method used with the characteristic.

One way to determine how characteristic groups may be helpful to you is to list the names of all characteristics you will be analyzing, and then group the names into logical categories. Categories could include pesticides, field measurements, or biological measurements (e.g., taxonomic abundance).

Note that if the same characteristic is analyzed in more than one medium, it is necessary to set up a separate group for each medium. This is common with toxics, which may be measured in water, sediment, and tissue.

In general, it is helpful not to be too restrictive when setting up characteristic groups. For example, if you only measure a few characteristics, you may find it easier to set up a single characteristic group, even though it might combine characteristics from a field observation, (e.g., water temperature) with those from a sample (e.g., total nitrogen in water). STORET will prevent illogical groups, such as combining characteristics for taxonomic abundance with those for automatic data loggers.

Table 6 shows different ways survey respondents have set up characteristic groups for STORET and the advantages and disadvantages of each option. One survey respondent suggests using old STORET parameter codes as row IDs since other systems, like the Permit Compliance System

---

(PCS) and USGS/National Water Information System (NWIS), use them. Some respondents do not recommend using characteristic groups because it complicates getting data back out of STORET.

Table 6. Options for Defining Characteristic Groups in STORET

Options for Defining Characteristic Groups	Advantages	Disadvantages
Activity type [1]	Easy to separate types of measurements [1] Easy to set up database [1]	
Standard Analysis Code (SAC) [2]	If already associated with the data, easy to set up for STORET [2]	SACs may overlap [2]
Biological: by order/family Chemical: by lab (except for USGS data) Fish: all in one group [3]		Creates too many groups [3]
One sample group (use lab codes as row IDs) and a few groups for field measurements [4]	Can associate characteristic groups with specific samples through database links [4]	Characteristic groups are less flexible than attributes [4]
Medium; activity type (biological, chemical or field) [5, 7]	Easily corresponds to monitoring data [5] Can set defaults for methods and units [7]	Using characteristic groups from outside sources can compromise defaults [7]
Medium, activity type, and analysis groups [6]		
Collection, sample preservation/transport, field or lab [8]		
One group for lab results, one group for field analysis [10, 13]	Quick to input data [13]	
Do not use characteristic groups [11]	Can associate metadata with each activity using other methods [11] Do not have to set up groups [11]	

Lab/Field Analytical Procedures and Equipment

Lab/field procedures and equipment provide information on how each piece of data was analyzed or measured. For example, for samples analyzed for total suspended solids (TSS), a lab may use EPA method 160.2, "Non-Filterable Residue – TSS." The lab or labs you use should have this information readily available.

Although it is a good idea to have an analytical procedure for each analyte, it is not required for everything. This is often the case for items measured in the field, such as water temperature, dissolved oxygen, and pH.

Lab and field procedures are connected to results in STORET. Table 7 shows the way survey respondents have set up analytical procedures and equipment for STORET.

Table 7. Options for Defining Lab/Field Analytical Procedures and Equipment in STORET

Options for Defining Analytical Procedures and Equipment	Advantages	Disadvantages
National procedures and some state-developed methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]	Creating state-specific methods allows more lab-specific methods to be defined [1, 3, 4, 13]. Can use SIM translation feature to map analytical procedures [9].	Some missing methods in the national list [2]. Finding the correct national procedure can be time consuming [7].

Sample Collection Procedures

Sampling collection procedures describe how samples were collected, including information on sampling gear, gear configuration, sample preservation, and storage. For measurements made at the station, such as water temperature, this does not apply. You can create as many procedures as necessary.

Collection procedures should include the following information at a minimum (Note that other metadata can be stored as well):

- Sample collection procedure ID (up to 12 characters),
- Name of the procedure, and
- Type of sampling gear used (e.g., water sampler, electroshock, net).

Gear/equipment configurations describe the types of field measurement or sampling gear that are used. Once you select gear ID STORET, you can enter the gear configuration. This can include serial numbers, size, manufacturer, and any other information.

Sample preservation, transport, and storage describe how a sample was preserved and transported to the lab for analysis. This basically consists of descriptions of containers, (e.g., glass bottles), and preservation, if any (e.g., dry ice, SO_4, etc.).

Table 8 shows different ways survey respondents have set up sample collection procedures for STORET.

Table 8. Options for Defining Sample Collection Procedures in STORET

Options for Defining Sample Collection Procedures	Advantages	Disadvantages
For gear configuration, describe tools and methods[1] For sample preservation, describe bottle used[1]	Methods are linked to data[1]	
Based on type of bottle used[2, 8]	Easy to separate data[2]	May be difficult to ascertain what bottle type was used[2]
National procedures and some state-developed procedures[3]	Easy to determine whether sample is a grab or spatial composite[3]	
Five collection procedures and six gear configurations that cover all cases[4]	Captures necessary level of detail[4]	
Biological or grab sampling[5]	Easy to track[5]	
Procedures unique to each project[6]		
Designate collection procedure but not gear[10]		
Do not define sample collection procedures[7]	Quicker to set up STORET[7]	Less descriptive metadata[7]

More Information

If you would like more information about setting up STORET, contact the EPA STORET assistance hotline at 1-800-424-9067 or by e-mail at STORET@epa.gov.

[This page intentionally left blank.]

Appendix A
Tips and Tricks for Setting up STORET Data Survey

This appendix contains the survey questions sent to current STORET users and the full text of their responses. The first section, submitter information, includes the name and agency of each respondent, and some other general information. The remaining responses are organized by STORET data categories (e.g., projects, stations, etc.). In each section, the questions are listed first with an assigned letter. The responses are numbered to correspond to the submitter information in the first section and then further identified by the question letter.

A.   Survey Instructions

The purpose of this survey is to gather helpful ideas for setting up various aspects of STORET. This information will be summarized, compiled and made available to the STORET community to assist others in making decisions on how to customize STORET for efficient data import, retrieval, and display. All setup information is helpful – even if it didn't work well. Setting up STORET can be daunting and a document summarizing how others have done it may save much time and frustration for new users as well as provide great ideas for experienced ones.

Please fill out this survey and send it to: wilson.eric@epamail.epa.gov. Thank you in advance for taking the time to provide this valuable information and help others avoid "reinventing the wheel."

Submitter Information

1. Name: Geoffrey Smith
    Agency: Delaware River Basin Commission
    Email Address:
    What type of data are you putting into STORET? Check all that apply.

✖ Water Chemistry	___ Habitat Data	___ Other
___ Sediment Chemistry	___ Biological Data – Raw	
___ Tissue Chemistry	___ Biological Data – Metrics	
✖ Physical Measurements	___ Data Loggers	

    Approximately how many projects and stations do you have in STORET now?
    2 Projects, ~100 Stations

2. Name: Carrie Wengert
    Agency: Pennsylvania Department of Environmental Protection
    Email Address: cwengert@state.pa.us
    What type of data are you putting into STORET? Check all that apply.

✖ Water Chemistry	___ Habitat Data	___ Other
___ Sediment Chemistry	✖ Biological Data – Raw	
✖ Tissue Chemistry	___ Biological Data – Metrics	
___ Physical Measurements	___ Data Loggers	

Approximately how many projects and stations do you have in STORET now?
3-soon 4 Projects, 570 Stations

3. Name: Paul Morton
    Agency: New Jersey Dept of Environmental Protection
    Email Address: paul.morton@dep.state.nj.us
    What type of data are you putting into STORET? Check all that apply.

___	Water Chemistry	_×_	Habitat Data	___	Other
_×_	Sediment Chemistry	_×_	Biological Data – Raw		
___	Tissue Chemistry	___	Biological Data – Metrics		
_×_	Physical Measurements	___	Data Loggers		

    Approximately how many projects and stations do you have in STORET now?
    37 Projects, 6000 Stations

4. Name: Jim Porter
    Agency: Minnesota Pollution Control Agency
    Email Address: jim.porter@pca.state.mn.us
    What type of data are you putting into STORET? Check all that apply.

_×_	Water Chemistry	___	Habitat Data	___	Other
___	Sediment Chemistry	___	Biological Data – Raw		
___	Tissue Chemistry	___	Biological Data – Metrics		
_×_	Physical Measurements	___	Data Loggers		

    Approximately how many projects and stations do you have in STORET now?
    33 Projects, 4421 Stations (Many stations are transfers from Legacy with no new data. 1836 stations have visits in the new system.)

5. Name: Tavis C. Eddy
    Agency: Wyoming Department of Environmental Quality/Water Quality Division
    Email Address: teddy@state.wy.us
    What type of data are you putting into STORET? Check all that apply.

_×_	Water Chemistry	___	Habitat Data	___	Other
___	Sediment Chemistry	_×_	Biological Data B Raw		
___	Tissue Chemistry	___	Biological Data B Metrics		
___	Physical Measurements	___	Data Loggers		

    Approximately how many projects and stations do you have in STORET now?
    2 Projects, _____ Stations

6. Name: Rick Langel
    Agency: Iowa Geological Survey (Iowa Department of Natural Resources)
    Email Address: rlangel@igsb.uiowa.edu
    What type of data are you putting into STORET? Check all that apply.

_×_	Water Chemistry	___	Habitat Data	___	Other
___	Sediment Chemistry	___	Biological Data – Raw		
___	Tissue Chemistry	___	Biological Data – Metrics		
_×_	Physical Measurements	___	Data Loggers		

Approximately how many projects and stations do you have in STORET now?
45 Projects, 326 Stations

7. Name: Deb Dorland
   Agency: MT Department of Environmental Quality
   Email Address: ddorland@state.mt.us
   What type of data are you putting into STORET? Check all that apply.

✗	Water Chemistry	✗	Habitat Data	___	Other
✗	Sediment Chemistry	✗	Biological Data – Raw		
___	Tissue Chemistry	___	Biological Data – Metrics		
✗	Physical Measurements	✗	Data Loggers		

   Approximately how many projects and stations do you have in STORET now?
   ~200 Projects, 4000 Stations

8. Name: James Adkins
   Agency: Div. Water & Waste Management, WV Dept. Environmental Protection
   Email Address: jradkins@mail.dep.state.wv.us
   What type of data are you putting into STORET? Check all that apply.

P	Water Chemistry	F	Habitat Data	___	Other
P	Sediment Chemistry	F	Biological Data – Raw		
___	Tissue Chemistry	F	Biological Data – Metrics	P = Present	
P	Physical Measurements	F	Data Loggers	F = Future	

   Approximately how many projects and stations do you have in STORET now?
   10 Projects, 3500 Stations

9. Name: Dave Wilcox
   Agency: Gold Systems
   Email Address: Dwilcox@GoldSystems.com
   What type of data are you putting into STORET? Check all that apply.

✗	Water Chemistry	✗	Habitat Data	___	Other
✗	Sediment Chemistry	✗	Biological Data – Raw		
✗	Tissue Chemistry	___	Biological Data – Metrics		
✗	Physical Measurements	___	Data Loggers		

   Approximately how many projects and stations do you have in STORET now?
   _____ Projects, _____ Stations

10. Name: Julia Utter
    Agency: Florida Department Environmental Protection
    Email Address:
    What type of data are you putting into STORET? Check all that apply.

✗	Water Chemistry	___	Habitat Data	___	Other
___	Sediment Chemistry	✗	Biological Data – Raw		
___	Tissue Chemistry	___	Biological Data – Metrics		
✗	Physical Measurements	___	Data Loggers		

    Approximately how many projects and stations do you have in STORET now?
    932 Projects, 18515 Stations

11. Name: Rich Hanson
    Agency: South Dakota Dept. Env. & Nat. Res.
    Email Address: Rich.hanson@state.sd.us
    What type of data are you putting into STORET? Check all that apply.

      ×   Water Chemistry   ___ Habitat Data   ___ Other
      ×   Sediment Chemistry   **×** Biological Data – Raw
      ___   Tissue Chemistry   ___ Biological Data – Metrics
      ×   Physical Measurements   ___ Data Loggers

    Approximately how many projects and stations do you have in STORET now?
    66 Projects, 1105 Stations

12. Name: Joe Gross
    Agency: North Dakota Department of Health
    Email Address:
    What type of data are you putting into STORET? Check all that apply.

      ×   Water Chemistry   **×** Habitat Data   ___ Other
      ___   Sediment Chemistry   **×** Biological Data – Raw
      ___   Tissue Chemistry   ___ Biological Data – Metrics
      ___   Physical Measurements   ___ Data Loggers

    Approximately how many projects and stations do you have in STORET now?
    110 Projects, 1275 Stations

13. Name: [respondent left blank]
    Agency: [respondent left blank]
    Email Address: [respondent left blank]
    What type of data are you putting into STORET? Check all that apply.

      ×   Water Chemistry   ___ Habitat Data   ___ Other
      ___   Sediment Chemistry   ___ Biological Data – Raw
      ___   Tissue Chemistry   ___ Biological Data – Metrics
      ×   Physical Measurements   ___ Data Loggers

    Approximately how many projects and stations do you have in STORET now?
    3 Projects, 383 Stations

Projects

*Questions***:**

A.   **How did you define your projects? (By type (water treatment plants, volunteer river monitoring etc.), by calendar year (all data for a specific type broken out by calendar year), by length of applicable QAPP (data for type is added as long as under the same QAPP), by site or facility (i.e., one project = one facility's data), etc.)**

B.   **What were the advantages and disadvantages to setting up projects this way?**

*Answers***:**

    1A. Project Name was defined by the program that the data is collected under.

      B. Advantage – Unambiguous results, ease of sorting once out of STORET.

    2A. I identified my projects based on the programs already designated. For example, one of our projects is WQN, which is all data associated with Pennsylvania's Fixed Water Quality Network program. Similarly, WQF is associated with our Fish Tissue sampling program and GWN is associated with our Groundwater Network.

      B. The major advantage was that the data for these programs were already divided based on the program itself and that was one less step that I needed to perform when formatting the data. I have not found any disadvantages of having my data broken down into these projects.

    3A. Mix: USGS by "Water Year," multi-year projects by Watershed, Bio data by calendar year.

      B. USGS data matches "Water Year" books published by USGS, bio data matches reports, 303(d) data easy to pull out by trip, but doesn't match QA Plans (which were by year).

    4A. A project may be a defined agency program, a more general ongoing monitoring effort, or a specific data provider, such as a local project that sends us its data. Examples: North Shore Load Project, Lake Trend Monitoring, Pipestone Creek TMDL Project. If staff want to be able to query data for a particular monitoring effort, we suggest they set up a STORET project for it.

      B. Data owners (monitoring and related staff) define how the data are organized, so it seems intuitive to them. It makes it easy to query data for a particular program or data provider.

    5A. Our program is specific enough in its use of STORET; for watershed sampling we have two projects: our REFERENCE Project which entails healthy systems that are used in our biocriteria analyses for determining stream condition in general. Our BURP (Benficial Use Recon. Project) Project is for the ambient monitoring of surface water across the state to determine if designated uses are being met.

      B. Some sites, or sampling events, are used for both projects, and we also have cases where a given sampling event (such as a water quality complaint) does not seem to fit into either category. The BURP project ends up being pretty all encompassing. The advantage is that these projects are also (somewhat) aligned with budget history of the same name.

    6A. Our projects are a combination the water-quality project and water year that the data is collected. For example, data collected as part of the Sny Magill 319 Monitoring

project during Water Year 1999 would be assigned a project of SNY1999. Our ambient data collected in Water Year 2002 would be assigned a project of AMB2002.

B. With our web-based retrieval, we can quickly retrieve project-specific data (for people only interested in a one project) and see any changes to our stations (for example, if an stations have been added or removed).

7A. Historic data migrated into the modernized STORET were migrated using projects categorized by type of project. The current data being generated and input into STORET is largely data collected for the TMDL program. Although this could be considered one project with common DQO and data collection methods, the effort includes roughly 100 HUCS across Montana. For this reason, the projects were broken down by HUC or Watershed. (We wanted "Project" to represent a long-term, ongoing grouping for results).

B. One advantage was that the project is always "known" to data management staff. Project corresponds to HUC, a required field for station establishment. (QC on the site location must be performed up front to determine HUC and Project, so it forces a QC check early on site locations).   A disadvantage is that it does not fully utilize the available metadata associated with the project designation. Note that the use of watershed-based project groupings does not preclude the creation of projects that are more consistent with STORET metadata capabilities, but did serve to simplify the STORET start-up for our TMDL program.

8A. Usually, by purpose. Have Total Maximum Daily Load, Regular Ambient, Intensive Survey, etc, projects. Duration is ongoing for most of our projects.

B. [respondent left blank]

9A. We have managed projects in one of three ways:
1. One project for the entire organization.
2. Some organizations have pre-defined projects (studies, etc). In this case we simply use these projects.
3. For some organizations that have several different entities entering data into one org, we have assigned one or more projects to each entity.

B. 1. The obvious advantage of a single project is simplicity. The disadvantage is that you limit your retrieval options.
2. Using existing projects is probably the best option for most programs as it is easy to manage and provides a valuable way to retrieve their data.
3. This final option works well for programs managing data from many different entities. A good example of this would be a state hosting several volunteer monitoring groups. If it is determined that all of the volunteer monitors will share a single org, then project becomes a good way to segregate their data.

10A. 1. The type of monitoring plan the agency is using often defines project.
2. Abbreviate of the existing project name is often used.

B. No apparent disadvantages.

11A. A project is a specific sampling effort delineated by location(s) and sampling date(s). We created an ACCESS table containing projectIDs as well as project duration, start date and a brief description of the project. This table is linked to other tables containing station IDs, sampling results, etc.

Our project IDs are somewhat descriptive of the project with the last character identifying the type of project. Example The Lake Alvin Assessment Project is ALVINZZ1. The number "1" designates it as an assessment. An implementation project will be designated by a "2" (Example – the Lake Alvin Implementation project ID would be ALVINZZ2. Dredging projects end in "D." All following assessments on the same project end in a different odd numbers all following implementation will end in a different even number. If we have something unique that doesn't fit the three typical projects we plan to add a different letter as the 8th character.

B. Most environmental problems, assessment projects, or control efforts are usually defined by where and when they occur so grouping data by project appears to be the most practical way to group data.

ID Set-up: This way we can sort data different ways. All Alvin data can be found by truncating the last letter off the ID. Or all assessment projects can be found by looking for the odd numbers, all implementation projects can be queried by searching for even numbers.

12A. For the most part they are defined by regionality (i.e., Watersheds, Counties, etc.).

B. A watershed may extend into multiple counties, and if defined by county, the project sponsor (319 sponsors) may not be willing to implement the project across county lines. Takes more cooperation between counties/sponsors.

13A. A project is often a study; so for example, one project may be a group of WQM stations along a certain stream. It should be noted that a station might be in more than one project.

B. Assigning data in this manner can be useful when pulling data from STORET.

Stations

Questions:

A. What type of labeling scheme did you develop for your station IDs?

B.   **What were the advantages and disadvantages to setting up station IDs this way?**

*Answers***:**

1A.   It is dependent on who is program manager. We have everything from full numeric station id's to station id's that have a state abbreviation and streams position in watershed longitudinally to collection organization and river mile of that site.

  B.   Extensive research to determine exactly where site is.

2A.   As stated above, our programs are defined with defined goals, sites, etc. The WQN program has stations set up simply by the letters WQN0### and the three digit number of the WQN site. The fish tissue stations are set up a little differently. They begin with WQF indicating the type of station, followed by the five-digit stream code followed by the river mile where sampling began. So, an example of a fish tissue station would be WQF00002-17.8. This indicates a Fish Tissue station on the Delaware River starting at 17.8 RMI. Finally, the Groundwater Network stations were named in the following manner. They too are begun with a three-letter prefix in this case, GWN. The next identifier is a letter indicating the drainage, D=Delaware, S=Susquehanna, etc. The next grouping of numbers/letters indicates the sub-basin (based off of PA's State Water Plan) where the station is located (i.e., 02C = Lehigh sub-basin). Finally, the last digits indicate a numerical sequence of stations assigned in that drainage. So an example of a GWN station is GWND02C023 which tells us the station is a groundwater station located in the Delaware Drainage Lehigh Subbasin and it is the 23 site identified in this region.

  B.   In all cases, the advantage was that the sites were very similar and easy to format so that they could all be imported into STORET following the same template. Again, I haven't found any disadvantages using this system.

3A.   Used IDs from Legacy STORET, based on old River Reach system.

  B.   Consistent with LDC (for as long as it's up) and stations come out in downsteam order. Disadvantages: Numbers are a pain to calculate, because they depend on miles up a segment and I don't have a route built in my GIS coverage. I can't retrieve data on project specific station IDs (USGS, bio) as we could with old STORET secondary IDs. Project specific Station IDs used in 305(b) and 303(d) reporting, so lists on EPA Web sites don't match what can be searched in STORET.

4A.   For lakes, we use commonly accepted IDs developed by the Minnesota Department of Natural Resources. Each ID consists of a county code and a serial number, e.g. 27-0016. To track specific monitoring sites on lakes, we include separate site codes on activity IDs. The site coordinates are tracked outside of STORET, although we hope to add them to STORET as "actual activity locations" eventually.

For streams, we use a serial number in the form S000-000. We record aliases for stations, including Legacy STORET primary and secondary codes, as external

reference scheme station IDs. The goal is for each monitored point to have a single unique ID with separate aliases used by different data providers.

B. Lakes: Almost all data collectors in the state use the same IDs. Because a station is defined as an area and not a point, it simplifies things and reduces the number of station establishments. Tracking the sites (actual activity locations) separately means maintaining another system and complicating our activity ID assignments. Streams: We do not currently have a good system in place to make sure that a monitored point gets established only once. It can be time-consuming to determine whether stations are co-located, and if so, how close is close enough to call it the same station. The IDs are meaningless, which is good data practice but less intuitive for monitoring staff who use the IDs. We used to use a stream name abbreviation and a mile indicator, such as CD-0.5. Increasingly detailed stream GIS coverages have rendered the mile values inaccurate. Assigning codes for new stations relative to old, inaccurate station codes got complicated.

5A. Ordinal by ecoregion, e.g., MRW26 means the 26th station for the Middle Rockies West ecoregion.

B. This is helpful in that it is easy for use to organize data by ecoregion, which is critical in our biocriteria development. Our stations have traditionally been assigned for those locations where we have collected biological (benthic macroinvertebrates) chemical and habitat data, as a full triad. This means that when we collect only chemical data (which is frequently) we have added the WQ (for water quality) onto the front of it e.g., WQMRW05. This has been cumbersome. Another issue with our station IDs is for those early stations; sometimes a MRW5 and a MRW05 both pertain to the same 5th station established for that ecoregion. This discrepancy has appeared and makes organization difficult.

6A. We use an 8-digit numeric code. The first 2 digits are for project identification. The next 2 digits are the county number. The remaining 4 digits are a sequential number.

B. Since stations can be used on multiple projects, we have to take extra time to double check our current stations to make sure we do not duplicate stations with multiple STORET number.

7A. Our station ID convention is HUC based, again. A three-character code designates the HUC. This is followed by a five-character acronym for waterbody, and then a two-digit numeric for a total of 10 characters.

B. Disadvantage: The Cross-reference to HUC code must be known prior to assigning a station ID, and the HUC must have been verified as correct. Another disadvantage is that the station ID must be verified as unique in STORET prior to using to avoid assigning duplicate identifiers.

Advantage is that all the stations in a single watershed group together in the STORET user interface and Report module, and the stations on a single waterbody group

together within a given HUC if consistent acronyms have been used. Also people recognize their station IDs.

8A.   Most of all our streams have Alpha Numeric labels. Many labels are greater than 15 characters allowed by STORET. So for a site, which have Id as WVABC-1-X-123-P-4-ZZ, Station Id could be ABC-001-0123, so it could fit in Station Id. Box.

B.   Need for STORET allow for more than 15 Characters for Station IDs. Then we can better link STORET to other databases in our agency. Now we only use EPA Key Identifiers for State Labels which can exceeds 16 characters allowed for Station Id for quite a number of our streams and waters.

9A.   In most cases, the stations already have some type of ID from their use in the local LIMS system. As there are many better ways to query a station that by ID, we typically use this existing ID or a simple numbering mechanism.

B.   The advantage is to keep it simple and to be able to tie it back to the source of the data. The disadvantage would be to not have a logical key (i.e., with extra info coded into the ID itself) that could be user for querying. In a robust database like STORET, however, you should simply be able to query a station by any of the many station fields and should not feel compelled to encrypt this information into the ID.

10A. 1. The station ids are based on delineation from a HUC code map used in the 1970s.
      2. Each agency has developed their own system for labeling station ids.

B. 1. The disadvantages are that the map is no longer in use and is not in an electronic format. It's hard to find and if you have gaps in your number set you're stuck having to create a new numbering system that doesn't fit with the old system.
    2. Some agencies within Florida would like FDEP to mandate a consistent method for creating station_ids (i.e., lat/longs as the station_id). This would allow any user to understand the specific labeling system during retrievals no matter which agency they are interested in.

11A. The first seven letters are the first seven letters the Project ID. Subsequent letters or numeric characters (the last 8) refer to type of waterbody and site ID. Example: ALVINZZLA01 might refer to the Lake Alvin Project (ALVIN), a lake sample (LA) and site 01 (a specific sampling location in the lake). ALVINZZLAT01 might refer to a tributary site (T) at a specific sampling location (01) on a tributary.

These sites can then be used for different types of projects for Lake Alvin such as an implementation project can have the same site ID as the assessment project. If more than one project was done at Lake Alvin then a numeric character could be inserted following the project name. So ALVIN1LA01 might refer to an in lake site from a Lake Alvin Assessment Project and an ALVIN2LA01 might refer to the same site used during an implementation project.

B. Simple logic and relatively easy to understand what the station ID refers to. This allows for quick recognition of which waterbody is involved and it is often a quick way to discern which project it is and what kind of project it is. It should be noted, however, that some projects involve more than one waterbody and so a project name may not always refer to a specific waterbody. In those cases, one would need more information (location) to figure out the project name.

12A. A 6-digit number beginning with the number 38. For example, our next station will be designated the number 385275.

B. Advantage: Ease of assigning ID's to newly created stations.
Disadvantage: Sequential numbers are not always possible when ids are assigned at different times. For instance, Lake Isabel may have 2 ids associated with it, 384207 and 384208. If a new id is needed, it would be designated the next available number (i.e. 385275). This is not a huge problem. It would be less difficult for the field personnel sampling to remember the code if they were sequential.

13A. A 6-digit number/letter combination is assigned to each station.

B. The letter portion of an ID is set up as an abbreviation of a specific waterbody or area of the state. This makes it easy to tell at a glance, in which region of the state, a station is located. The number portion is unique for each station.

Station Visits

*Questions***:**

A.   How did you set up or define your station visits?

B.   What are the advantages and disadvantages to setting up the station visits this way?

*Answers***:**

1A. One station visit per day.

B. It tends to idiot proof the system…and you know that we need that.

2A. The main identifier with Station Visits was our sequence number that was associated with the original sample. This along with the date and time creates a unique station visit.

B. The advantages to doing things this way is that you do not need to create a field or re-create existing data. The disadvantage, and I have not run into this yet, is that I am not sure what the program will do if a sequence is ever repeated.

3A. Sequentially or allow SIM to define.

  B. Allowing SIM caused problems if we had to go back to a site a second time for the same "round" of sampling. Allowing SIM to assign keeps us from going crazy trying to keep it straight.

4A. One numeric ID per day per trip, usually assigned automatically by SIM. The only exception is for one large volunteer monitoring project, in which we code visit IDs to keep each volunteer's data together. This project uses three-character alphanumeric codes.

  B. This system seems to work well.

5A. Chronological by Station ID; most other cells left blank within the Station Visit Menu.

  B. To be honest, we use this rarely to find/organize data; it does allow us to know how many times we have visited a site, but this ends being easy to discern in many ways.

6A. We let SIM automatically assign station visits into STORET when we upload data. For most of our projects, we will have only on visit per day. So, SIM is setup to assign station visits by sampling date. But, some projects will have multiple visits on the same day. So, SIM is setup to assign visits on sampling date and time.

  B. We do not use Station Visits with our web-based retrieval program.

7A. For our TMDL program, a station visit may span several days for practical purposes. Trips to distant part of the state may involve many hours of driving time. Monitoring staff may visit a number of sites repeatedly to complete assessment and sample collection tasks. A visit spanning several days allows us to group related activities on a single trip into one visit "event."

  B. One advantage is the ease of tracking the visit, and matching the visit when loading results for various activities for a given station on a given trip. Also the activities are grouped together. One disadvantage is that it is not always intuitive that various activities performed on different days involve the same visit. From the monitor's perspective, there may have been two or three distinct site visits during a trip to a given region of the state.

8A. For a given trip, we may visit many of our sites several times during a "trip."

  B. [respondent left blank]

9A. For most programs, we have created a single station visit per station and date that the sampling occurred. If the program identifies that they may visit a station several times in one day, we will set it up as one station visit per station per date and time.

B. Again, keep it simple if the program is not collecting this data. Of course, as soon as a program starts tracking their actual trips and station visits; these trips and visits should be imported into STORET to provide the most accurate tracking of this information possible.

10A. Station visits are defined using the one visit per day default in SIM.

B. None yet discernable. It is not an often-queried data layer.

11A. Station visits are auto-generated. If we need to query some data, we can query by Project/Station/Date and Time.

B. Advantage – One less number to try to come up with. It might be easy querying with more descriptive station ID however with how we organize and use our data our system works for us.

12A. By the visit number for that year. Station 384321 was visited in May. That would be visit number 1. Then visited again in June, that would be station visit number 2, and so on.

B. Works well and is easy to understand.

13A. A station visit is any sampling performed on one day.

B. It seems the logical way to set up station visits.

Trips

*Questions***:**

A.   How did you set up (define) your trips (i.e. by day, month, year etc.)?

B.   What were the advantages and disadvantages to setting up trips this way?

*Answers***:**

1A. We define trip as one year of sampling under that specific program.

B. Advantage – Easy to parse out each sampling season.
Disadvantage – Is a rather broad "trip" designation.

2A. Currently the trips are identified by T and then the station number. Activities are set up using date, month, year, etc.).

B. Several times, the system rejected a trip ID for some reason unknown to me and I had to reformat the data to "make it work." I haven't found any other advantages or disadvantages, and I'm still confused as to the significance of the trip.

3A. Usually by year.

B. People usually want to see a whole year's worth of data and I don't have a billion trips in my list as I would if I used matrix or season as the grouping item.

4A. By project year. For example, the Milestone project's 2002 water year data is under the trip MILE-2002WY. The 2002 calendar year Lake Trend data is under the trip LAKETRND-2002CY. Project staff determine whether they want the data stored by water year or calendar year.

B. This system seems to work well. The only advantage we see to using trips in the literal sense is to track trip blanks. We decided it was not worth the hassle of creating so many trips for that one function.

5A. We created Trips for the sole purpose of STORET. They are specified by the field crew, year and week; for example: LA980831 pertains to the trip taken by the Lander Crew, in 1998, on the week starting 8/31.

B. Advantage: we can isolate by year and field office.
    Disadvantage: the ids get long, and for previous years it is time consuming to look up the beginning of the week that a trip went out.

6A. Trips are a combination of project and date (dependent on the number of samples received in a month). For projects that we expect few samples each month, the trip would be a combination of the project and the year that the samples were collected. For projects that we expect many samples each month, the trip would be a combination of the project, month, and year.

B. We use trips to find data in STORET that may need corrections. Otherwise, we do not use trips for data retrieval.

7A. Trips were defined by year. Generally we use a project year, except in the case of TMDL program where an entire major basin or region of the state is combined into a single Trip. For monitoring projects conducted external to our agency, a separate trip helps distinguish the data and allows us to designate a "trip leader" without maintaining external personnel in our Organization personnel list.

B. The Trip by Year concept helps greatly in organizing and loading current and historic data, and ensuring completeness without duplication. If data from a given data source belongs in a designated Trip, we know right where to look to see if the data exists in STORET. Related data also tends to be grouped together, rather than a statewide grouping based on all samples collected in June, for example.

8A. If for a specific watershed during a given year, we consider a trip for 12 months for that watershed for one Project, like TMDL. Next year we will be sampling another watershed or body for 12 months for TMDL. For our Ambient Sites, we use quarter of a year in which all our sites are visited just once.

  B. [respondent left blank]

9A. For programs tracking trips, we simply import what they provide. For all others we have created either one trip per day, month, or year.

  B. For legacy data, one trip per month or year works well to organize their data without an enormous number of trips. The more years of data they have, the better the one trip per year option gets. For programs that do monthly monitoring, one trip per month works well as this provides a pretty accurate representation of their actual trips.

10A. Trips are defined by using the one trip per year default in SIM.

  B. The advantages are that the STORET interface can handle browsing the number of trips generated by the one-per-year default. Other than that it is not an often-queried data layer.

11A. Trips are auto-generated by day. One trip per day per project.

  B. Advantage – Easy for us to generate trip ID.
     Disadvantage – None really for us.

12A. By the year.

  B. Have not discovered any major advantages or disadvantages with this method of setup.

13A. Each month's worth of WQM data is considered a trip. All data for all stations is included in that trip.

  B. Assigning data to monthly "trips" makes it easy to find data from a particular period of time.

Activities

Questions:

A. **How do you set up (define) activities (i.e., keep field measurements separate from laboratory analyses or combine together, etc.)?**
B. **What were the advantages and disadvantages to setting up activities this way?**

*Answers***:**

1A. We combine all together (I just give them arbitrary numbers) and the designation lies in the Activity type.

 B. Advantage – More time efficient to number this way for data entry.
 Disadvantage – When open screens in STORET you can't really tell what you are looking at in terms of parameters.

2A. Activities are defined by month, day and year. As far as specific activities, field and lab results are separated using our parameter codes assigned to the sample.

 B. The advantage to this is that it keeps all activities separate and unique and allows an end user to clearly see which data was analyzed in the field and which data was analyzed in a lab.

3A. Mix again. Data from USGS is combined; data we do internally is separate.

 B. Separate forces me to keep track of Activity IDs so the field and lab stuff matches up, and also is unique. Also a pain to keep separate because I need to know if the Trip/Visit is new or existing to tell SIM to create a new one or use an existing one.

4A. Activities were more difficult to define than most other STORET concepts. Because field and lab data had to be separated by type into different activities, we had to devise a system to keep sampling events together. An explanation of our ID scheme follows. It might offer some insight into how we did this.

 Lakes: Activity IDs consist of a three- to five-digit site code indicating the sample location on the lake, plus a hyphen, plus a two-digit depth to the nearest meter, plus F for field msr/obs or S for sample, plus R if a replicate, plus a digit if necessary for uniqueness. For 2-meter-integrated samples, the suffix following the site code is – I2S. Examples: 201-00F1, 102-03S, 102-03SR, 401-I2S.

 Streams: Activity IDs are usually the lab sample number followed by one or more descriptive codes. The field data corresponding to the lab data also uses the same sample number for the core of the ID. If a lab sample number is unavailable, we use a separate serial number in its place. The suffix codes for regular samples are F for field msr/obs, S for sample, plus R if it's a replicate. For QC samples, the suffix code is Q, plus either E for equipment blank or R for reagent blank. Examples: 200215678F, 200215678S, 200215678SR, 200215678QE.

 B. Lakes: We set it up this way to track site codes and to keep profile data organized. There are some significant disadvantages, however, due to the rounded-off depths. Say a person uses a Hydrolab or YSI to collect profile data at site 103, resulting in readings at 3.7m, 4.0m, and 4.4m, among others. Under our scheme, each would have an activity ID of 103-04F. I have to add a digit to each to make it unique: 103-

04F1, 103-04F2, and 103-04F3. The activity ID field is not long enough for us to use more decimal places and still accommodate the rest of our coding scheme.

Streams: This system seems to work well.

5A. Our activities are as follows: we conduct biological, chemical and field data collection.   Field pertains to that which was directly observed in the field, such as pH, Dissolved Oxygen, etc. So an activity ID of B67 refers to a biological activity, and the same station and date would have a C67 for the chemical data.

  B. There is clarity to our approach, although the chronological ID does not aid us in any major way, and by splitting field and chemical samples we treat them separately in database organization, while we need to consider them together in analyses.

6A. We keep activities separated. Currently, we use 3 different activity Ids. Two are for field measurements (one for measurements related to water conditions, the other for atmospheric conditions that are recorded) and the final one for water samples that are collected.

  B. Activities are only used to help us find data in STORET. Like trips, activities are not used in our web-based program for data retrieval.

7A. We ordered sheets of pre-printed, color-coded labels from (Shamrock). Each sheet may have, say 10 identical labels. These labels are used to identify the related activities for a given site visit. To ensure uniqueness for each activity ID, a suffix is added that represents a medium–code or sample type. For example, "M" for macroinvertebrate, "W" for water, "S" for sediment, "F" for Field Msr/Obs, etc. Typically, one label is used in the field book, and/or on the site form. This gives all activities related to a site-visit similar activity ID(s), and all group together. The activity ID has a two-digit prefix for year, an alpha character corresponding to a given trip, a three-digit numeric (sequential), then a hyphen and room for the medium code suffix. Example: 02-L127-M. Our activity ID is typically 9 characters, though the suffix for sample type, or the sequential numeric could be longer if necessary.

  B. This convention has many advantages, and is working well for us.

8A. Field measurements are separated from laboratory analysis. Activities, like Nutrients, are kept separated from Metals, etc.

  B. [respondent left blank]

9A. In almost all cases we have taken the effort to break out the field measurements from the lab analyses.

  B. This takes a considerable amount of work, compared to leaving them as a single activity, but the data is ultimately modeled much better. If this is not completed, then users finds themselves being forced to add a sample collection procedure for an

activity that may have only had field measurements. Keeping field and lab results in a single activity also causes the problem that you lose an easy way to track if the result was created from a sample or a field measurement.

A similar situation exists for medium. Often Air temperature is taken with a water activity. As the characteristic actually includes the word 'Air' in the characteristic description, we will sometimes include this with the water activity just to keep things simple.

10A. Activities are defined as a composite set; one activity for each sample date, time, depth, activity type and category.

  B. It seems that this is the only way to set up an activity short of ignoring the activity type data layer and defaulting everything to Sample or Field Msr/Obs. It would help if the Activity ID field were longer than 12 characters so that it would be easier to construct an ID given that there are so many pieces of data that have to be considered in one activity id.

11A. The activity ID for lab analysis is the same as it is for Measurements/Observations except measurements and observation activities have an "F" after the Activity ID to designate as a field measurement.   We use separate ACCESS tables/forms to group field data, chemical data, biological data, etc. The ID for contaminate data has "C" after the lab ID, Elutriate data has an "E" after the ID, the ID for the Algae data has an "A" after the ID and so forth.

  B. As an advantage we can group our field and chemical data easily by comparing similar activity IDs.   As STORET dictates each kind of data (i.e. chemical, biological, field, etc.) has certain attributes and this uniqueness of each data type led us to group the data in our ACCESS table by sheet. In our ACCESS data entry forms we can then keep unique biological sampling or chemical sampling and link them to the field data sheet in the same form.

12A. The lab assigns a Chem-Log number for every sample (i.e. 03-R0001). This log number is used as the activity ID. To keep the field measurements associated with that sample separated, we place an "F" at the end of the log number to differentiate the field measurement results (i.e., 03-R0001F) from the lab results.

  B. No major advantages or disadvantages.

13A. We assign all field measurements the ID number that the laboratory assigned the laboratory data for the station. We add an "F" to the end of that lab number.

  B. Using the lab ID number for field data makes it easy for us to tell that certain field conditions existed when certain samples for lab analysis were taken. Nevertheless, because field data has an "F" at the end of the number, STORET does not actually combine the lab and field data. Therefore, we hope to eliminate any confusion about what analysis was performed in the field and what analysis was performed in the lab.

Characteristic Groups

*Questions***:**

A.   **How do you set up characteristic groups? Do you group characteristics by medium, activity type, i.e., Field Measurements vs. Sample, or do you find it more convenient to combine different types of characteristics into a few or even one group.**

B.   **What were the advantages and disadvantages to setting up characteristic groups this way?**

*Answers***:**

1A. We tend to group by activity type.

  B. Advantage – It allows us to separate out the different types of measurements quickly, eases database set up when using SIM.
Disadvantage – When in STORET App. it is not possible to see all at once when done this way.

2A. Our characteristic groups are defined by the SAC (standard analysis code) used in analyzing the specific samples at the lab. So, for example, if a sample is sent to the lab labeled SAC 10, SAC 10 may consist of testing for Metals. So, under characteristic group SAC, all of the metals would appear as Characteristics.

  B. As is the case with many of the things I've used, this already existed and was associated with the data so therefore, I did not need to create a new way to group the samples. The disadvantage was that the SAC's overlap as far as characteristics so it was timely in setting up initially.

3A. Started when we were limited to 100 Row IDs. Bio by Order/Family. Chem by lab except for USGS, which is by groupings (field, routine, metals, organics, etc). Field/Lab separate. Fish all in one group (created after the 100 Row ID limit).

  B. Creates too many groups, hard to keep track of. Worried that a few big groups would slow SIM.

4A. We set up one large sample-type characteristic group for our main lab using the lab's analysis codes as row IDs. We also set up a few field-type characteristic groups. I had planned to set up one characteristic group per lab, but lately I have been moving away from characteristic groups. Instead, I created something like a characteristic group table in a separate database, and I import data through SIM using the characteristic-by-attributes method. I find it gives me more flexibility to handle special cases that may not exactly fit STORET's characteristic group templates.

B. Setting up sample-type characteristic groups by lab organizes procedures logically. Characteristic groups in general are somewhat less flexible than storing characteristics by attribute through SIM.

5A. We break them out by medium; biological, chemical and field parameters.

B. This has worked well, and philosophically corresponds to our monitoring agenda. One initial problem, that presently needs to be fixed, is that at the time we developed our CGs only 99 characteristics were allowed in each group. Since our taxa list for invertebrates is over 400 taxa, we had to make 5 different biological CGs. Now that we can have more characteristics per group, I have to decide whether to retrofix the older data, or make a decision to choose a year to move forward with a single new CG to cover all bug data.

6A. We set up or characteristics on a combination of medium, activity type, and analysis groups. For example, we have one category for atmospheric field measurements (i.e., air temperature, cloud cover, etc.). We have a category for water field measurements (i.e., water temperature, pH, etc.). For water samples that we have analyzed in a lab, we have characteristic groups for nutrients, bacteria, pesticides, semi-volatiles, etc.

B. [respondent left blank]

7A. Our characteristics are grouped by medium, then activity type.

B. It is convenient for us to use characteristic groups for work where agency staff and set procedures ensure defaults like units of measure and methods will be consistent. We try to avoid the use of characteristic groups in receiving data from projects or data sources external to our agency, as we are concerned about having a number of defaults as "understood." A change in methodology or units cannot be easily accommodated in these cases.

8A. Group characteristics, by collection, sample preservation/transport, or if they are field, or lab types.

B. [respondent left blank]

9A. We have worked with an even mix of programs that have setup characteristic groups and those that just pass all of the characteristic information directly to STORET. When using characteristic groups and rows, the most important thing is to set them up in a way that can easily be mapped from the source data. You do not want to have to work backward and add the group and row to the source data set.

B. [respondent left blank]

10A. Characteristic groups are set up with activity type as the dependency. All characteristics that are analyzed in the lab are in one group; all that are analyzed in the field are in another.

    B. Aside from not being able to default the characteristic group value when creating a batch file, nothing.

  11A. We do not set up characteristic groups.

       There are also physical limitations to the ACCESS data entry forms that predispose us to split the data into different data types. There is only so much information that can fit in a single data entry form. So our contaminant data while considered chemical data is on a different sheet. Special chemical parameters that are not sampled on a regular basis are on a different sheet. Even thought they are on separate sheets all water chemical samples will have the same ID with just a different letter at the end of the ID.

       In addition, our chemical data typically comes to us from the laboratory separate from the biological data and sometimes separate from all of the field data. So it also makes sense to separate the data types simply because we get the data separately anyway.

       We do not use characteristic groups, we use our own built sample collection procedures and all meta data is added to each activity ID via an ACCESS process.

    B. Advantages – Didn't have to set up Characteristic groups. More flexible.

  12A. [respondent left blank]

    B. [respondent left blank]

  13A. We set up characteristic groups by laboratory (one for each lab) and by field data.

    B. The advantage to doing it this way is that when you enter data through SIM, all data from one lab can go in with one characteristic group. It makes data input go faster.

Lab/Field Analytical Procedures and Equipment

Questions:

A.   **How do you set up lab/field analytical procedures and equipment? Do you rely on the "National Procedures" provided by STORET or do you create your own?**

B.   **What were the advantages and disadvantages to setting up lab/field analytical procedures and equipment groups this way?**

Answers:

  1A. National Procedures where apply. For gear and collection we designate our own.

B. Analytical – very time consuming to search for and designate manually not being familiar with methodologies.
Gear/Collection – designating our own allowed us to get more specific on our actual sample collection procedures.

2A. Our lab has given me a listing of all of the methods they use in analyzing the samples and I've in turn relied mostly on the National Procedures. However, DEP has several methods that they've developed or adopted that were not on these lists so I also created my own.

B. I think the National Procedures are a good idea, but they need to be kept current with new releases of publications. The advantage to using them is that it took a lot less time to adopt and already existing procedure, but the disadvantage was that there were some methods missing.

3A. Use national if they match, otherwise created my own (especially for USGS Denver lab methods). Set up both field and lab methods.

B. Creating my own lets users know where we deviated from national procedures and add procedures not supplied by EPA-HQ. National save me time, especially with Citations.

4A. We started out by creating our own procedures, and we still do for our main lab. Increasingly we rely on the national procedures as we collect data from a myriad of data providers and labs.

B. Using the national procedures is less hassle. Creating your own lets you reference lab-specific methods that may deviate from the national procedures.

5A. Most of this was done in communication with our lab manager. We mainly used the National Procedures.

B. It was very organized, and all the relevant material is in one place.

6A. We rely heavily on the national procedures. However, we will create the method only if the parameter does not have an established national method.

B. Most parameters that we test for have established methods, so it is easiest to use the national procedures that STORET provides.

7A. Whenever possible we try to reference National Procedures in use.

B. Tracking down the correct National Procedure in use can be time consuming and frustrating, especially if STORET makes a distinction between methods that does not seem to correspond to method numbers in a laboratory's method documentation. Nevertheless, we try to strive for a high level of data comparability, and approved national methods are an element of that.

8A. Mostly, we rely on National Procedures. A few we have created for visual inspection, in-stream flow measurement, etc, where we could not find any in National Procedures.

  B. [respondent left blank]

9A. Most programs have some field in their source data that can be mapped to the analytical procedure. We use the SIM translation feature to do this mapping. Some programs have utilized several of the National procedures, but just about everyone we have worked with has had to add some of their own as well. Very few programs have done much with Sample collection procedures or Equipment. In almost all cases, we have created a single sample collection procedure that does not require equipment and loaded all samples with this procedure.

  B. The big thing here is that we simply had to make due with the information available in the source data. By defaulting to a single Sample collection procedure that does not require equipment, we have solved the problem of not having the necessary data. In some cases, the program did provide multiple sample collection procedures and equipment, which is nice if you have the data.

10A. National procedures are used when possible, otherwise they are created as org specific metadata.

  B. None.

11A. We have an Analytical Codes ACCESS table that contains information about a specific parameter, its STORET character, fraction analyzed, units, analytical procedure, and detection limits. The analytical procedures follow national procedures and use the same code numbers. All of this information is listed for each laboratory because different laboratories may analyze parameters differently and even have different detection limits for the same analytical procedure.
We try to use only recognized methods. If we can't find a method that works we develop our own, enter it into STORET and then it gets added through our ACCESS automated process.

  B. Simple logic.

12A. Both. If our lab procedure deviates from the National Procedure in any way, it is noted by defining a new state defined analytical procedure.

  B. [respondent left blank]

13A. We do it both ways.

  B. You can customize.

Sample Collection Procedures

*Questions***:**

A.   How do you set up your sample collection procedures, i.e., sampling gear, gear configuration, sample preservation and storage?

B.   What were the advantages and disadvantages to setting up sample collection procedures this way?

*Answers***:**

1A.   Gear Config. – Just described the tools we used for collection and the method of sampling (Bottles/Bridge).
Sample pres/storage – The bottle each sample was in (acid500ml, iced1L, etc.).

  B.   Advantage – Do not have to go back into STORET to look at our methods; it's ingrained in the data request.

2A.   The sample collection procedures were set up based on what type of bottle is used to collect a certain type of sample. For example, CONT-5 is the bottle and method used to collect all metals, CONT10-Phsphorus, Ammonia and TOC, etc.

  B.   The advantages to this were that everything stayed separate based on the type of sample being collected. The disadvantage was that you needed to know what type of equipment was used when collecting the sample and that info might not be easily obtained.

3A.   Creating my own lets users know where we deviated from national procedures and add procedures not supplied by EPA-HQ. National save me time, especially with Citations.

  B.   People basically want to know if it was a grab or a spatial composite, not as concerned about the exact technique. Since we use a certified lab, they're not as concerned about the preservation, transport and storage.

4A.   We have five collection procedures that seem to cover all cases: composite (flow-weighted with auto-sampler), composite (multiple locations), grab, lake depth point, lake surface 2-meter integrated.

We have six gear configurations. We set up one, catch-all, sample preservation and storage default which we plan to refine further later.

  B.   It captures the level of detail we want without being too onerous.

5A. We broke it out as being either biological sampling, or grab sampling, and the sampling procedures were described in the boxes provided.

B. This works well for us, having a basic list. However, for when there are changes in our approach, we will need to be able to change this section easily.

6A. Sampling collection procedures are unique for each project and are based on the project's QAPP or standard operating procedures.

B. [respondent left blank]

7A. To speed start up, we chose to bypass some of the metadata that was not required here, such as detailed sample preservation descriptions. We did not associate gear with sample types (as for grab samples), and defaulted other gear types according to well-established sampling protocols.

B. This was quicker, and placed less burden on monitors and data providers. Disadvantage is less descriptive metadata.

8A. Most of our samples are grab. Then the waters are put into different containers—one for nutrients, another for total metals, dissolved metals, according to sample preservation/storage and transport.

B. Need to have STORET to indicate date time that samples are delivered to labs.

9A. See above.

B. [respondent left blank]

10A. Usually there is just a simple sample collection procedure with no associated gear. The dependency is activity type.

B. Don't know yet.

11A. We use an ACCESS table containing media, projectID, waterbody type, activity class, sample collection procedure, configuration name, and gear ID. Sample preservation and storage is not addressed, other than it is implied through the AnalyticalCodes table and the analytical procedures code number.

B. [respondent left blank]

12A. [respondent left blank]

B. [respondent left blank]

13A. Very general.

B. [respondent left blank]

One Final Question…

*Questions***:**

Are there any other aspects of Modernized STORET that you set up one way originally but now wish you had done it differently or could start over?

*Answers***:**

1. Yes and no. I have started out setting up my data entry in a SIM ready format which conflicts with data analysis format so I have to crosstab in access for analysis. What I've done lately is wait until all my data is in STORET so I can pull it out in a single format then get it ready for analysis. Its makes entry EASY but it compounds analysis…quite a little paradox.

2. Yes. Originally, I couldn't figure out the Activity ID aspect of things. It had to be a unique number for every sample. I talked with several different folks who assigned a random number for the activity ID. This is what I did also initially. I didn't like the fact that you could not determine which activities were what when downloading information from the internet so I actually did start over and now give my samples an activity ID based off of the parameter code and a random number.

3. Don't get me started. I would have dumped the migrated stations from LDC and imported new stations off of GIS coverages since I later found the coordinates were off. We did the former before SIM, when it took forever to create new stations. I would have created one Characteristic Group per lab and one for Field Measurements/Observations, one for fish and one for bugs "macroinvertebrates." I would have been consistent with Projects (match QA Plan) and Trips. I would have used STORET parameter codes for Row IDs since other systems (PCS, USGS) still use them. I would not have gotten so hung up about filling in every field I could on stations. I would have "broken" ties with LDC and established multiple Orgs, one for each physical Bureau in NJDEP, so they could administer their own data and I wouldn't have to look at their stations or Characteristic Groups when searching for information (and get faster response times).

4. Only as noted above, under advantages and disadvantages. Just as important as defining these concepts and IDs was to develop our "pre-STORET" Access database to create SIM files based on them.

5. I wish I could start over with the establishment of the Char. Groups for the Benthic Macroinvertebrates, now that the ITIS list has grown, and the C. Group can take a larger number of characteristics. But looking back, I would not have done anything differently, since I worked with what I had at the time. I wish we had utilized personnel names more frequently in our spreadsheet days. I have gotten confused with some of our data that was designated non-detect while other data was labeled present below quantification limit; I wish

I had labeled it all non-detect, since now I have to verify which is which, and confront what this difference truly means for when we serve the data.

6. [respondent left blank]

7. Not really. We are doing the best we can to implement the system with the resources and staff we have. As time goes on we hope to improve not only our business process, but also enhance the metadata we are providing with the data, and more fully utilize the potential of the STORET system. The "trade-offs" we have made to simplify use are seen as necessary to get us going on the system.

8. [respondent left blank]

9. Each migration gets a bit easier. We have loaded data many different ways depending on the needs of the program and the quality of their data. My big advice would be that you need to approach each project based on its merits, and not try to force a one-size-fits-all approach.

10. I would not necessarily use characteristic groups because querying data from STORET becomes more complicated and there are errors in the database that are introduced by the use of characteristic groups on occasion.

11. [respondent left blank]

12. None as of yet. There may be things discovered in the future, as we get into this a little more.

13. No.