RECLAMATION

Managing Water in the West

Open Water Quality Data Initiative (OWQDI) Scoping Report

Research and Development Office
Science and Technology Program
Final Report No. ST-2016-1415-01

Irrigation canals in Utah (left) and Oregon (right).

U.S. Department of the Interior
Bureau of Reclamation
Technical Service Center

September 2016

Mission Statements

The mission of the Department of the Interior is to protect and provide access to our Nation's natural and cultural heritage and honor our trust responsibilities to Indian Tribes and our commitments to island communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

REPORT DOCUMENTATION PAGE						Form Approved OMB No. 0704-0188
T1. REPORT DATE September 2016		**T2. REPORT TYPE** Research				**T3. DATES COVERED** 2016
T4. TITLE AND SUBTITLE Open Water Quality Data Initiative (OWQDI) Scoping Report						**5a. CONTRACT NUMBER** RY1541WD201611415
						5b. GRANT NUMBER
						5c. PROGRAM ELEMENT NUMBER 1541 (S&T)
6. AUTHOR(S) Merlynn D. Bender email: mbender@usbr.gov office telephone: 303-445-2460 Denver Federal Center Technical Service Center (86-68210) P.O. Box 25007 Denver, CO 80225						**5d. PROJECT NUMBER** 1415
						5e. TASK NUMBER 1
						5f. WORK UNIT NUMBER 86-68210
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Department of Interior, Bureau of Reclamation, Technical Service Center, Water Resources Planning and Operations Group 86-68210 Denver Federal Center, Denver, CO						**8. PERFORMING ORGANIZATION REPORT NUMBER** ST-2016-1415-01
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office U.S. Department of the Interior, Bureau of Reclamation PO Box 25007, Denver CO 80225-0007						**10. SPONSOR/MONITOR'S ACRONYM(S)** R&D: Research and Development Office BOR/USBR: Bureau of Reclamation DOI: Department of the Interior
						11. SPONSOR/MONITOR'S REPORT NUMBER(S) NA
12. DISTRIBUTION / AVAILABILITY STATEMENT Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/						
13. SUPPLEMENTARY NOTES The guidance, support, and review by TSC RESC IT staff is greatly appreciated.						
14. ABSTRACT *(Maximum 200 words):* Ways to process and save water quality data were reviewed. Initially, add-on packages to database software are the recommended procedure to accommodate immediate needs; a structured non-proprietary opensource software virtualization platform format is recommended for future needs. Add-ons developed for a particular piece of software should be modular and flexible to accommodate software changes over decades. Specialized add-ons for enhanced automation might be developed for a future opensource software package or future virtualization techniques such as software defined anything (SDx). Development of a coordinated open water quality data initiative (OWQDI) program and a demonstration project are recommended. Hierarchical Data Format version 5 (HDF5) format might be used for such a demonstration project.						
15. SUBJECT TERMS data life cycle, data processing and archival, water quality						
16. SECURITY CLASSIFICATION OF:				**17. LIMITATION OF ABSTRACT** U	**18. NUMBER OF PAGES** 28	**19a. NAME OF RESPONSIBLE PERSON** Merlynn D. Bender
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U				**19b. TELEPHONE NUMBER** 303-445-2460

S Standard Form 298 (Rev. 8/98)
P Prescribed by ANSI Std. 239-18

Open Water Quality Data Initiative (OWQDI) Scoping Report

Research and Development Office
Science and Technology Program
Final Report No. ST-2016-1415-01

PEER REVIEW DOCUMENTATION

Project and Document Information

Project Name: Scoping the Open Water Quality Data Initiative WOID: Z1415

Document: Open Water Quality Data Initiative (OWQDI) Scoping Report

Document Author: Merlynn D. Bender Document date: September 2016

Peer Reviewer: Kurt Wille

For Reclamation disseminated reports, a disclaimer is required for final reports and other research products, this language can be found in the peer review policy:

Review Certification

Peer Reviewer: I have reviewed the assigned items/sections(s) noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.

Reviewer:________________ Date reviewed: 7/11/2016
(Signature)

Open Water Quality Data Initiative (OWQDI) Scoping Report

Research and Development Office
Science and Technology Program
Final Report No. ST-2016-1415-01

Prepared by Merlynn D. Bender, Hydraulic Engineer

Technical Service Center
Water Resources Planning and Operations Support Group

U.S. Department of the Interior
Bureau of Reclamation
Technical Service Center

September 2016

Acronyms and Abbreviations

2-D	two dimensional
AGPM	Animator Graphics Portfolio Manager
ALG	algae
ALK	alkalinity
BETTER	Box Exchange Transport Temperature Ecology Reservoir
BOD	biochemical oxygen demand
BRACS	Brackish Resources Aquifer Characterization System
cfs	cubic feet per second
D	dissolved
DET	detritus
DISA	Defense Information Systems Agency
DO	dissolved oxygen
DOR	derive dissolved organics
DQOs	Data Quality Objectives
DSAP	drought sampling analysis plan
EC	electrical conductivity (specific conductance)
EPA	Environmental Protection Agency
ft	feet
FY	fiscal year
GIS	Geographical Information System
HDF5	Hierarchical Data Format version 5
HUC	Hydrologic Unit Code
IEEE	Institute of Electrical and Electronics Engineers
INL	Idaho National Laboratory

Interior	Department of Interior
ISS	inorganic suspended solids
IT	information technology
L	labile
MS	Microsoft
NH4	ammonia
NOX	nitrate plus nitrite
NWQMC	National Water Quality Monitoring Council
NWIS	National Water Information System (USGS)
QC	quality control
OM	organic matter
OWDI	Open Water Data Initiative
OWQDI	Open Water Quality Data Initiative
P	particulate
P	parameter, or
P	phosphorus (STORET)
pH	numeric scale to specify acidity or alkalinity
PN	Pacific Northwest
PO4	bioavailable phosphorus
QA	quality assurance
QAPP	QA Project Plan
R	refractory
Reclamation	Bureau of Reclamation
RMS	River Modeling System
S&T	Science and Technology

SAP	Sampling Analysis Plan
SDN	software-defined networking
SDS	software-defined storage
SDx	software defined anything
SOPs	Standardized Operating Procedures
SRAO	Snake River Area Office
STEWARDS	Sustaining the Earth's Watersheds, Agricultural Research Data System (USDA)
STORET	Environmental Protection Agency database to store data
T	temperature
TDS	total dissolved solids
TSC	Technical Service Center (Reclamation)
TVA	Tennessee Valley Authority
U.S.	United States
USGS	United States Geological Service
W2	CE-QUAL-W2 model
WaterSMART	Sustain and Manage America’s Resources for Tomorrow water program (Interior)
WEEG	Water and Energy Efficiency Grant
WQP	Water Quality Portal
WQX	Water Quality eXchange
YSI	Yellow Springs Instrument

Executive Summary

Ways to process and save water quality data were reviewed. Initially, add-on packages to database software are the recommended procedure to accommodate immediate needs; a structured non-proprietary opensource software virtualization platform format is recommended for future needs. Add-ons developed for a particular piece of software should be modular and flexible to accommodate software changes over decades. Specialized add-ons for enhanced automation might be developed for a future opensource software package or future virtualization techniques such as software defined anything (SDx). Development of a coordinated open water quality data initiative (OWQDI) program and a demonstration project are recommended. Hierarchical Data Format version 5 (HDF5) format might be used for such a demonstration project.

Contents

Figures

Tables

Introduction

A majority of water quality data being collected are not used, because the data are not typically stored in a user-friendly format for future use. The goal of this scoping research project is to help identify and understand existing water quality database issues while providing a foundation for developing guidance that can simplify data archival for basin-wide drought, climate change, and watershed projects relating to water quality improvements. The research question is "How can the existing water quality data and future water quality data integrate with the U.S. Department of Interior (Interior) Open Water Data Initiative (OWDI)?" This Open Water Quality Data Initiative (OWQDI) scoping proposal reviewed options for development of a simple and consistent water quality database in terms of metadata, Technical Service Center (TSC) and Bureau of Reclamation's (Reclamation) regional office repository format, access of the data, and analysis of the data for development of a data product. Water quality data for surface water and groundwater should likely not be stored in the same database because integration of that information is challenging. Therefore this scoping report focusses on surface water quality data for use in specific riverine and reservoir models rather than watershed models which might be better addressed with a land-use Geographical Information System (GIS) approach.

Reclamation is responsible for and needs access to a vast amount of surface water quality data in the 17 Western United States. That water quality data was often collected using public funding. The ultimate goal would be to provide better data for public projects including environmental projects. The reoccurring droughts in the southwest United States are providing a greater need for a consistent water quality database storage and analysis model. The mission of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public. Riverine and reservoir water quality data is a valuable resource that needs to be organized and stored efficiently.

Insight was gained by investigating the data systems developed by the Environmental Protection Agency (EPA), Tennessee Valley Authority (TVA), the U.S. Geological Service (USGS), Northern Colorado Water Conservancy District (http://www.northernwater.org/DynData/WQDataMain.aspx) and other agencies. Attempts to store "all the data" in one database tend to fail. Therefore, scoping identified the following focused and specific needs for riverine and reservoir modeling:

1) There is a need to review the available historical Reclamation water quality data sources and databases before initiating each Reclamation riverine and reservoir water quality modeling project. This task would need to concentrate on available historical riverine data, reservoir water

quality profile data, the vertical datum used, and the formats used to store the point and profile data.

2) There is a need to develop a proposed simplified format for storing both the riverine mixed point data and reservoir depth profile data on an open source platform that will be consistent with the Interior OWDI. The format used by the Animator Graphics Portfolio Manager (AGPM) pre- and post-processing modeling software is an example of proprietary software (http://www.loginetics.com/) that demonstrates the utility of consistency for rapid plotting with closeness-of-fit statistics for comparing reservoir profiles or fish refuge volumes under various operational scenarios. A simplified drought sampling analysis plan (DSAP) is an example of the need to store water quality profiles for multiple reservoirs.

3) There is a need to develop a strategy to prioritize collection of dry, median, and wet year data sets for an early season (high pool) versus late season (low pool) reservoir water quality comparison and to more automate plotting and quantification of water quality degradation based on water volume and fish refuge habitat. Such a procedure might be used to provide guidance on ranking drought emergency management grant applicant projects or WaterSMART Water and Energy Efficiency Grant (WEEG) projects in the Western United States.

4) There is a need to develop a proposal to support a subsequent conducting funding request in fiscal year 2018 (FY18) for investigation of a focused, simple, and easy to maintain interface that is useable by agencies for specific riverine and reservoir water quality models for various time steps. To simplify data storage for one-dimensional riverine flow and water quality models, efforts should concentrate on “hourly” hydraulic data, water temperature data, meteorological data, and parameters specific to flow and water temperature calibration rather than parameters for dissolved oxygen (DO) calibration. Boundary, depth profile, and release data for two-dimensional reservoir models should concentrate on storing parameters for both water temperature and dissolved oxygen calibration. At least daily hydraulic inflow and outflow data should be used unless hourly peaking power or flow fluctuations dominate the modeling; if peaking power needs to be modeled, hourly data are needed. Hourly meteorological data is needed to calculate reservoir modeling inputs for at least a day and a night time step. An example of a user-friendly interface is the Texas Water Development Board's Brackish Resources Aquifer Characterization System (BRACS), http://www.twdb.texas.gov/innovativewater/bracs/.

5) For current riverine and reservoir monitoring and modeling, profile data collected and typically placed in Excel comma separated value (.csv) format by many agencies should import smoothly into a water quality

> database via a user-friendly GIS interface under a research-to-operations transition vision. However, the database needs to be flexible enough to import many formats used for field data now and potentially those unknown formats to be used in the future.

After the EPA STORET database was phased out for smaller and more specific agency databases, there is a need to provide a consistent database format. Currently water quality databases are project specific, spreadsheet-based, and are often lost due to retirements and agency turnover. The benefit of a consistent water quality database is that it allows better data management, comparison between agencies, and greater access by agencies as well as the public.

EPA maintained a stand-alone state-of-the-art centralized water quality database known as STORET up to the turn of the century. EPA now integrates data from other agencies (figure 1). With the National Water Quality Monitoring Council (NWQMC), the Water Quality Portal (WQP) integrates publicly available water-quality data, through use of the Water Quality eXchange (WQX), from the USGS National Water Information System (NWIS), EPA Storage and Retrieval Data Warehouse (STORET), and USDA ARS Sustaining the Earth's Watersheds, Agricultural Research Data System (STEWARDS) (EPA, May 20, 2016, https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange).

With the explosion in amount and types of water quality data, many water quality databases sprung up in federal, state, and local agency offices. Since the turn of the century, the Reclamation regional and area offices in concert with technical support from the TSC developed project specific water quality databases. The Reclamation Science and Technology (S&T) program under the Research Office funded this scoping project to identify an approach to store water quality data for the future. The project was identified by Reclamation's TSC staff as a good candidate for research to attempt to preserve technology transfer. TSC Research Office staff coordinated and contracted with Water Resources Planning and Operations Support Group (86-68210) to scope a future direction for the storage of Reclamation's water quality data.

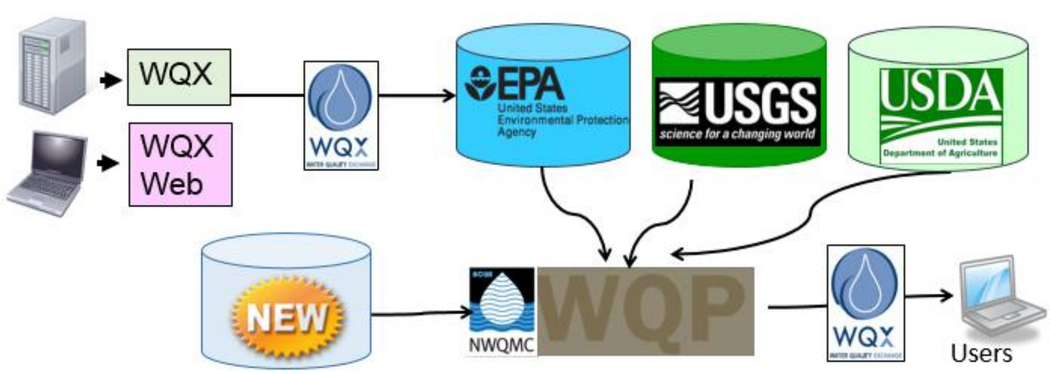

Figure 1.—EPA data portal flow (EPA, May 20, 2016).

Historical Background

At the turn of the century, only two reservoir water quality models were being used extensively at Reclamation. Initially developed at TVA, the Box Exchange Transport Temperature and Ecology of a Reservoir (BETTER) model (Bender, et al. 1990) was used widely at Reclamation before the turn of the century. After the turn of the century, the CE-QUAL-W2 model developed by the U.S. Army Corp of Engineers (Cole and Wells, 2002 and Cole and Wells, 2006) and maintained at Portland State University has been used because it overcame the flat pool assumption of the BETTER model allowing modeling of the slope of a reservoir water surface. The use of both of these models is described in more detail in the Reservoir Water Quality Manual developed under the Manuals and Standards program (Reclamation, May 2009).

In contrast, many one-dimensional riverine water quality models have been developed. The River Modeling System (RMS) developed at TVA (Hauser and Schohl, 2003) is an example of a common tailwater flow and temperature model used downstream of dams. The use of the RMS modeling system is described in more detail in the River Water Quality Manual developed under the Manuals and Standards program (Reclamation, August 2010).

New modelers often ask, "How were data retrieved for riverine and reservoir models years ago?" For either reservoir or riverine models, the following was a typical informal memorandum request for specific data that could be used to develop water quality reservoir model input files:

Informal Memorandum

To: Daryl Ciruli, Data Technician

From: Merlynn Bender, Hydraulic Engineer

Date: February 9, 2001

Subject: Storet INDEX, INVENTORY, and possible RETRIEVAL for Beulah Reservoir (Malheur watershed, Hydrologic Unit Code (HUC) 17050116)

Daryl: Dave Zimmer (Pacific Northwest (PN) regional client) has asked our group to start on the Beulah Reservoir Water Quality Model which may be either a CE-QUAL-W2 (W2) model or a BETTER model. The first priority are the STORET surface water index and inventory for the watershed surrounding Beulah Reservoir which falls in Hydrologic Unit Code 17050116 which includes the Malheur watershed. We will need all years of data in the record. Please first email to me the new index and inventory files with a brief README text file describing the files.

Second priority will be STORET groundwater index and inventories for Beulah Reservoir or Malheur watershed. Use similar STORET command files used on the Cascade and Lowell projects.

Malheur watershed and HUC 17050116 are large. We will not do large data retrievals, unless we have too. The goal is to retrieve what is needed for model calibration. If there is enough data for a dry year data set, the third priority will be two formats of STORET surface water retrievals for the Lake Lowell Watershed with specific parameters (P) numbers in table 1 as shown below. The first format needed is the standard retrieval with parameter headings. A STORET command file will be used as a template for the retrieval. The second format needed is the type where the first columns are the year, month, day, time, and station and the next columns are the parameters. The STORET code is to be placed on the top of each column using one line of code in the STORET command file. Please email to me the two files (one for each format) with a README text file and also please have hardcopies printed. Thank You.

Table 1.—Parameters (P) for STORET Retrievals

For each depth (feet or meters from the water surface) if available	
P=10 Temperature (centigrade)	P=610 NH3+NH4 as N Total (mg/L)
P=300 Dissolved oxygen (mg/L)	P=618 NO3-N dissolved (mg/L)
P=60 Streamflow (cfs)	P=625 Total Kjeldahl as N (mg/L)
P=61 Instantaneous streamflow (instan. cfs)	P=629 Total organic Kjeldahl as N (mg/L)
P=70 Turbidity (JKSN)	P=630 NO2&NO3 N-Total (mg/L)
P=76 Turbidity trbidmtr Hach (FTU)	P=631 NO2&NO3 N-Diss (mg/L)
P=78 Trans Secchi depth (meters)	P=665 Phosphorus Total (mg/L as P)
P=94 Conductivity field (umhos/cm)	P=671 Phosphorus dissolved Ortho (mg/L as P)
P=95 Conductivity @ 25 (umhos/cm)	P=680 Total organic carbon (mg/L)
P=310 BOD 5 day (mg/L)	P=955 Silica dissolved (mg/L)
P=335 Low level COD (mg/L)	P=956 Silica Total (mg/L)
P=400 pH (SU)	P=1045 Total Iron (Fe) (ug/L)
P=403 pH lab (SU)	P=32210 Chlorophyll-a (ug/L)
P=410 Total alkalinity as CACO3 (mg/L)	P=70507 Phosphorus Total Ortho (mg/L as P)
P=530 Residue Total Nonfiltered (NFLT) (mg/L)	P=80154 Suspended Sediment Conc. (mg/L)
P=535 Residue Volatile NFLT (mg/L)	P=70301 Diss Solids sum

Table 1 was not a complete list of the water quality parameters needed nor did it contain the actual input parameters to either a BETTER or W2 reservoir model. A typical W2 water quality input file would have the following input parameters to W2 in mg/L for the first day (jday 1.000) in January (there is no zero time in W2 modeling):

```
#JDAY TDS   ISS    PO4   NH4     NOx   LDOM  RDOM  LPOM  RPOM  ALG1  ALG2   ALG3
1.000   230.5   5.8   0.024 1.200   0.900   0.57   27.70     0.15     2.85    0.000   0.030   0.030
DO      TIC     ALK
8.61   35.67  142.0
```

The parameters historically retrieved from STORET did not match the parameters typically input into the W2 model. The W2 modeler had to derive the inputs and that becomes difficult for labile (L) and refractory (R) dissolved (D) and particulate (P) organic matter (OM) or the LDOM, RDOM, LPOM, and RPOM components. Total dissolved solids (TDS), inorganic suspended solids (ISS), bioavailable phosphorus (PO4), ammonia (NH4), nitrate plus nitrite (NOX), total inorganic carbon, alkalinity (ALK), and different forms of algae (ALG1, ALG2, and ALG3) were derived from available data as well as know relationships under dilute environmental conditions. There have been more recent attempts to develop W2 model inputs by coupling upland watershed and downstream water body models (Debele, et al, 2006).

With some assumptions, the modeled water quality input data sets for modeling were derived from the list in table 1 because that data was the only data typically available of sufficient quantity in STORET or other database. For example, biochemical oxygen demand (BOD) was often used to derive dissolved organics (DOR) and detritus (DET) inputs for the BETTER model; short-term and long-term BOD data in conjunction with a field study of other parameters was often used to derive labile and refractory particulate and dissolved organic matter component inputs (LDOM, RDOM, LPOM, and RPOM) for the CE-QUAL-W2 model (Sullivan, et al., August 3, 2009). Many assumptions were used to develop the model inputs. If BOD data was not available, the crude assumptions used in modeling were developed from sparse data or previous water quality modeling experience from a similar project. Water quality concentration data were often copied from a similar project and then filled in with the sparse data specific to the project being modeled.

A data technician, such as Daryl Ciruli (retired) in the previous example informal memorandum, was trained to index, inventory, and retrieve the STORET data listed in table 1 for a water quality modeler, such as Merlynn Bender. The water quality modeler would then look at the index of data site locations to find the specific input locations for the model. The inventory of data provided the list of number of available data points and basic statistics for each parameter and missing needed model inputs at each specific location. The retrieval of the actual data points was then pulled into a Lotus spreadsheet or a Microsoft (MS)-Excel spreadsheet and manually manipulated by the water quality modeler for input to the specific riverine or reservoir model based on specific locations for modeling and modeling assumptions. The modeler had to separately provide the water surface elevation file (usually derived from a stilling basin water level gauge at the dam forebay) which was tied to project vertical datum, and then manually tie each reservoir profile (and the depths of each vertical data point) to the water surface elevation at each average time of reservoir profile data collection.

The system was time-consuming, worked well, and forced the water quality modeler to look at every water quality data value placed into the chosen water quality model. The art of water quality model calibration and experience provided the rest. And there was much data manipulation, such as averaging, to develop each data point input into the model. Assumptions to fill in gaps were made based on the best available data and engineering judgement. At the end of the process, the modeler was well aware of the quality of the water quality model calibration, as well as the limitations of the data gaps used to develop the model. The result for a reservoir model was model output that reflected the water quality modeler's experience at model calibration to match an observed reservoir profile data set for typically a dry (stagnant), average (median), or wet (flushing) year. Seasonal patterns were adjusted using model coefficients. Sometimes the model was run on an independent verification data set to test robustness of the reservoir temperature and water quality profile calibration. The framework for retrieving data for riverine and reservoir modeling is the focus of a "simplified" methodology that sets up the condition for storing specific and not all data required for modeling.

Methodology

There are two types of water quality data required for reservoir water quality model calibration. These are (1) boundary condition data required for each modeled branch or tributary input location and (2) reservoir depth-profile water quality concentration data required for each model calibration point within the reservoir. Riverine models typically require completely-mixed data rather than a reservoir profile.

The methodology to fill model input boundary condition data sets consists of five main tasks: (1) review of the index list of sampling locations; (2) review of the inventory of the number of available data points for each specific water quality parameter; (3) retrieval of the individual water quality data points; (4) parsing of the data into spreadsheet columns for statistical manipulation such as averaging; and (5) extraction of the time-stamped data for import into the model input files. Based on the required end use of the water quality data, the water quality data should be stored in a format for retrieval of an index file (list of sampling locations), an inventory file (statistical summary of the data at each sampling location), and a retrieval file (individual water quality data points stamped with the field time of collection). To develop loadings, corresponding continuous flow data will also be needed at each specific inflow point to the reservoir or river resulting in larger files. The corresponding water quality input data could be interpolated from the spot data in the sparse input water quality concentration data sets. The water quality model multiplies the flow data by the water quality concentration data thereby providing the loading to the reservoir or river.

The methodology to fill model water quality profile calibration data sets consists of six main tasks: (1) review the index list of reservoir profile sampling locations; (2) review the inventory of the number of profiles and the types of water quality data collected at each reservoir calibration profile location; (3) retrieval of the profiles (both water quality data and the corresponding depth from the water surface) and the corresponding water surface elevation at the time of profile collection; (4) manipulation and interpolation of the observed water quality concentration data into a format that matches the center of the model layer at each profile time; (5) extraction of the time-stamped observed profile data for import into the model input files in a format for plotting against modeled profiles; and (6) development of closeness-of-fit statistics used to compare observed water quality profiles to modeled water quality profiles for model calibration and verification. Two commonly used closeness-of-fit statistics when comparing reservoir profiles are absolute mean error and root mean square error differences between observed and modeled data.

This scoping study will focus on accessing, analyzing, and creating a data product from an index, inventory, and retrieval of the individual data points for water quality modeling. Figure 2 shows where this scoping study fits into the design phase of the data life cycle.

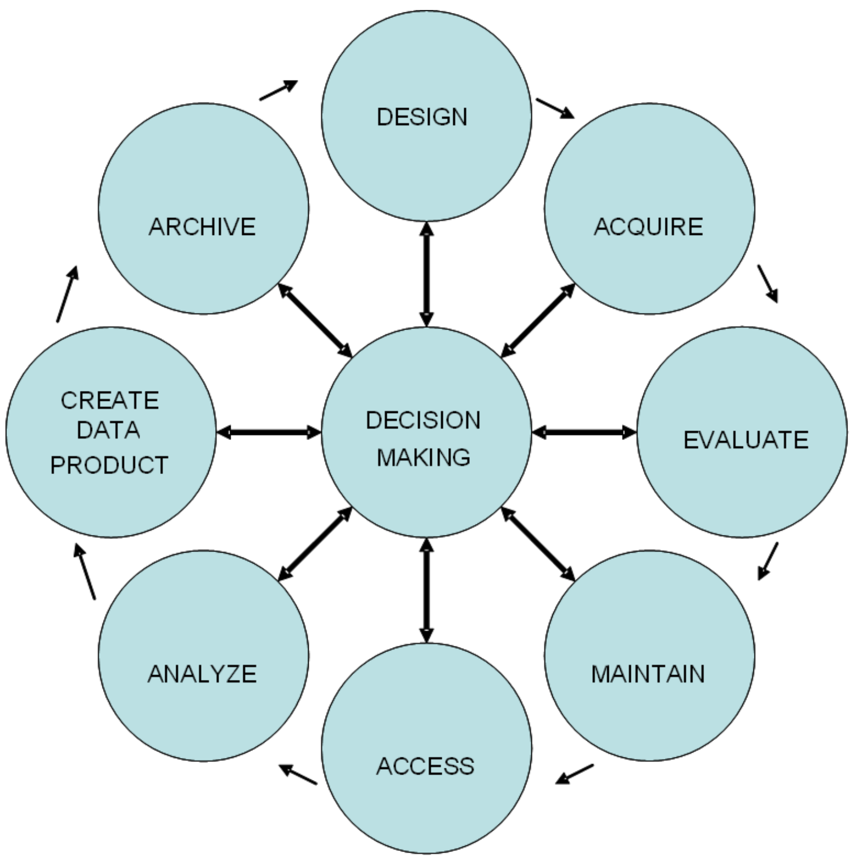

Figure 2.—Data life cycle (FCW, October 2015).

Data Stewardship and Data Life cycle Requirements

- The data life cycle (figure 2) is a central concept within data stewardship. Holistic planning with the business requirements and decision-maker needs at the center is essential.
- Large data collections require collaborative skills with many internal and external institutional entities.
- Data stewardship requires that somebody must be responsible for data management, in other words there must be professional data stewards/managers.
- There is a requirement for communities of interest and practice to develop workflow models, best practices, and data standards.
- There is a need to develop technical skills in the mission areas for managing data and datasets, security measures/access controls, a plan for data quality evaluation, and peer review processes for making data public.

Data Storage Formats and Naming Convention Challenges

There are typically many sampling site locations and many names for each specific sampling site location all within a few hundred feet (ft) of each other. Data are often collected by different agencies for various studies during different years. Often a nearby data collection sampling site has a different name and might be treated as the same site. Table 2 is an example of inconsistent and confusing naming conventions over several years of data collection.

Table 2.—Example of Confusing STORET Stations Names for Beulah Reservoir Water Quality Modeling

These following sites with multiple names were part of a sampling plan.

Beulah Reservoir (a) surface and bottom samples were collected at each site:
MAL013: 100 m upstream of Agency Dam
MAL___: 1000 m upstream of Agency Dam

Outflow (b):
BOI008: Agency Dam release
2040177: Mora Canal Wasteway #1

Inflow (c):

BOI023: New York CL Lake Shore Drive
07F00__: N.F. Malheur River station #6
MAL106: Upper Malheur R, at Peterson Place, 34E 19S 8CAC

In summary, the naming conventions need to be standardized across agencies and agency groups as well as metadata that describes each data collection site. Metadata, including pictures (see figure 3) of the sampling site, are often needed to help describe the location used for sampling. As an example of metadata, depth never exceeded 2.5 ft upstream of the six-foot Cutthroat flume in the cross section in figure 3 which was about 16 ft-wide and had rapid velocities. This site appeared to be completely mixed by the time water flowed over the flume. However, recirculation eddies downstream of the flume could result in warmer water temperatures near the stream banks. Therefore, water temperature data should consistently be collected just upstream of the six-foot wide Cutthroat flume near the most-upstream light-blue stilling basin gauge located in a plastic pipe (which can be seen more closely in figure 4).

Figure 3.—Upstream measurement location just downstream of Upper High Creek Canal headgate and just upstream of six-foot wide Cuthroat flume gauge (flow is from right to left).

Figure 4.—Upstream six-foot wide Cutthroat flume stilling basin staff gauge for determining flow released from the Upper High Creek Canal headgate. This 1.2 foot staff reading corresponds to 32.1 cfs.

Specific or exact sampling site location descriptions are important for accurate and consistent sampling. Database consistency is challenging. That brings this research scoping document to database storage issues that include the following:

1) Grouping multiple nearby site locations
2) Multiple naming conventions for the same sites
3) Different water quality parameters collected at sites over time
4) Quality control of the data collected varies over time
5) Metadata needs to stay with the water quality data

Data collection sites move. For example, one downstream flow measurement on the upper reach was taken 150 ft upstream of the bridge location during April 29, 2015 as shown in table 3 and figure 5. The upper reach downstream measurement site was moved about 300 ft downstream of the previous day's location or 200 ft downstream of the bridge to a location without moving bed conditions during the afternoon of April 30, 2015 and located just upstream of an irrigation turnout. The question is "Should the flow and water temperature data from these nearby sites be stored in the database as one location?" That becomes a judgement call for the project manager.

Table 3.—Richmond Upper High Canal Early-Season High-Flow Inflow/Outflow Measurements during Wednesday and Thursday, April 29-30, 2015

Measurement No.	Flow (cfs)	Estimated Seepage (cfs/mile)
1 Wednesday noon – upstream reach at headgate	30.6	--
2 Wednesday afternoon – downstream end of upper reach above bridge	26.7	2.6
3 Thursday morning – upstream reach at headgate	39.2	--
4 Thursday afternoon – downstream end of upper reach downstream of bridge	36.0	2.1

Figure 5.—Downstream end of the upper reach where sampling occurred both upstream and downstream of the bridge over the canal.

The specific flow and water temperature data collected in the site of figure 5 was used in a WaterSMART report (Bender, January 2016) and was not stored in a database. The more important question is "How and where should this flow and water temperature data be stored for future use and by whom?"

Planning Suggestions for Future Needs

Software will someday manage the entire environment without the dependency on physical hardware such as a computer. What was collected on a computer and then fed into a data base is likely being lost as the hardware and software for that hardware become outdated. The result is lost data unless the data base is carried forward. This is especially true for multi-dimensional data bases such as those used for water quality profiles of many environmental parameters. Software defined anything (SDx) is a software-defined anything environment. Software managed Information Technology (IT) environments are now seen as the large part of the future (see Appendix A). Water quality data collection needs to be focused on specific parameters that can be managed with future SDx in mind. An SDx OWQDI demonstration project is needed to ground proof that concept.

In addition, a hierarchical, filesystem-like data format like Hierarchical Data Format version 5 (HDF5) might be further explored for use in an OWQDI demonstration project. HDF5 simplifies the file structure to include only two major types of objects which are data sets of a homogenous type and group containers. However, because the number of parameters, stations, and types of water quality data often collected over depth adds dimensional complexity, HDF5 format may not be adopted for complicated water quality database structures. Suggested improvement to input data file structure may include making existing text files smaller and organized in a logical directory structure. The HDF5 data storage model is suggested because this format offers flexibility and the data is stored (referenced) in a directory type structure providing a logical extension of the text file directory structure. Furthermore, the HDF5 data model has the notion of attributes, and would be a way to share with others and provide interface to users through commonly used software, for example MS-Excel. A conceptual layout of the data interface provides only a framework (see figure 1 from "Forecasting Crop Irrigation Water Requirements," by Pruitt, Bracken, and Gangopadhyay, 2014, Reclamation S&T Program). Much was learned and gathered from that previous scoping analysis to provide a foundation to identify the greater challenge of storing water quality data for modeling or a potential future OWQDI demonstration project. A Reclamation water quality data base demonstration project (demonstration project) is recommended to work on details.

Due to the continuing drought, a DSAP may allow for the development of a demonstration project for a future OWQDI project. Alternatively, flow and water temperature data collected under the WaterSMART WEEG canal seepage verification program might be used as an OWQDI demonstration project. The need for a demonstration project might be tied to drought and WEEG programs.

Some of the worst water quality conditions within Reclamation project areas occur during droughts. Elevated water temperatures and poor water quality due to low flushing and stagnant low pool conditions affect aquatic life communities, including threatened and endangered species. Water quality modeling data collected during droughts is required for riverine and reservoir model calibration under low flow and low pool conditions. The Reclamation area and regional Offices face water shortages and low reservoir pools which in turn could harm valuable fisheries. Consecutive drought years provide an opportunity to collect extreme drought data. Unfortunately due to tight budgets and the difficulty in predicting drought conditions several years in advance, funding often comes too late to plan for adequate data collection. Monitoring plans describing data needs, data collection requirements and associated equipment, including automated data collection platforms, are needed for targeted watersheds. Water quality parameters, sampling locations, and required frequency for riverine and reservoir environments need to be determined for targeted Reclamation reservoirs and rivers based on technical and political factors. Reclamation's Pacific Northwest (PN) region Snake River Area Office (SRAO) boundary might be an area assessed for automated yet focused data collection needs for a demonstration project as follows:

- Primary Deliverables: 1) A prioritized list of targeted SRAO reservoirs and rivers for which drought data collection is required to address water quality issues affecting threatened and endangered species and other critical aquatic life species according to Reclamation interests and ranking factors. 2) A flexible generic DSAP template describing sample collection design, sample analysis design, data validation, and data analysis in accordance with the Interior, Reclamation "Quality Assurance Guidelines for Environmental Measurements (August 2003)."

- Potential Partners: Co-funding from the Reclamation SRAO or PN Regional Office would be required. Coordination with the USGS, Idaho and Oregon Departments of Environmental Quality, and the Reclamation Committee on Hydrologic Modeling might be necessary. A draft DSAP potential proposal might be developed as indicated in the following example:

Example of a Draft Drought Sampling Analysis Plan Proposal

Overview:

A Drought Sampling Analysis Plan (DSAP) is needed. Data is often collected without proper understanding of how the data will be used. When a drought develops, often riverine and reservoir water quality data is or was collected without adequate quality assurance (QA). QA integrates Data Quality Objectives (DQOs), Standardized Operating Procedures (SOPs), and approved methodologies (protocols) with a written description of details and delineates responsibilities in a QA Project Plan (QAPP). QA is not quality control (QC). QC asks if we are doing things correct. QA asks if we are doing the correct things. One of the first steps in a DQO planning process is development of the Sampling Analysis Plan (SAP). The SAP is a formal stand-alone document or an integral part of a QAPP that specifies the processes and defines the responsibilities for obtaining environmental data of sufficient quantity and quality to satisfy the project objectives. The SAP is the document which specifies the task and provides the technical procedures to be used in collecting samples and performing analysis for environmental measurements so that the quality objectives determined in the DQO planning process are met. Reclamation needs a generic DSAP template that is targeted for rivers and reservoirs affecting threatened and endangered species and other aquatic life species according to

Reclamation interests and ranking factors. The flexible generic SAP to be developed will describe sample collection design, sample analysis design, data validation, and data analysis in accordance with the Interior, Reclamation "Quality Assurance Guidelines for Environmental Measurements," August 2003. These guidelines are often not currently followed due to the cumbersome nature of this policy oriented document. Field personnel need a simplified template that provides adequate complete data sets for modeling studies, especially during special studies such as those for drought conditions.

Reclamation data collection procedures and data warehousing procedures are lacking resulting in mistrust of the data. Modelers are forced to use the sparse monitoring data, fill in the gaps, guess if there is no data, and then stand in front of decision makers defending the model output. Model results are flawed by a long list of errors starting with the sampling analysis plan. There is more error in the data collection process than in the error in mathematical regressions, yet more research dollars are spent on getting the equations right. It is time to improve the basics of systematically collecting trustworthy data for investigation of more adequate use of over-allocated water and energy resources. Good science needs good data.

Water quality data and models can be used to verify and validate operational quantity models. Often operational model output is found inadequate to drive water quality models at low flows indicating operational model inadequacies. Water quality models concentrate on details, gains, and losses in shorter reaches which can often be used to pinpoint where the more coarse operational models have errors. Reclamation concentrates on accounting of water budgets and largely ignores clues from water quality models. Water quality models are typically used for low flow analysis. With the increasing concerns of global climate change and more severe droughts, water quality models will become more important to assessing water quality for support of endangered species in riverine and reservoir habitats. A standardized drought sampling analysis plan placed in service today will serve to benchmark current conditions for future historical studies within and across basins.

One of the most basic elements to gaining the confidence of decision makers in model results is improvement of the data being collected. If the data can be trusted, the model results will be used with more confidence. Using research dollars to improve the data collection process is essential for good research. How the collected data is stored is a question that should be addressed in a future proposal. Data and metadata should be stored in a relational data base that allows selective retrieval, data mining, and can be used by the modeling systems currently available. Much of Reclamation's data is currently on several PCs and disappears during retirements, transfers, or other movement of the workforce.

A future related $75,000 research proposal (table 4) could focus on data and metadata storage and archival. If data collected today is used in a study a thousand years from now, how will it be stored and what confidence will the researchers have in the data? What if the drought data collected today becomes the median clean water benchmark data in 1000 years from now? With anticipated increased future demands, over allocation of the water resource, drier conditions, global climate change, and contaminated return flows, a defensible set of data is needed to benchmark the current conditions for future comparison studies. There are no do-overs if drought data is not collected or collected incorrectly. In many cases, no data is better than incorrectly collected data. Misinformation leads to not trusting the data, the models, and the research. Also data collected in an organized consistent manner is more useful than data collected without much of a plan.

Table 4.—Tasks for a Future Data Demonstration Project

Fiscal Year	Task	Type	Completion Date	Funding Request	Comments
2018	Task 1: Develop prioritized list of reservoirs and rivers and data required for Pacific Northwest Region	Labor	09-30-2018	$15,000.00	
2018	Task 2: Research existing sampling analysis plan information	Labor	09-30-2018	$10,000.00	
2018	Task 3: Draft the DSAP and place on Reclamation website	Labor	09-30-2018	$5,000.00	
2018	Task 4: Peer review by lab analytical specialist	Labor	09-30-2018	$5,000.00	
2018	Task 5: Finalizie the drought sampling analysis plan (DSAP)	Labor	09-30-2018	$15,000.00	
2019	Task 6: Field test the DSAP	Labor	09-30-2019	$10,000.00	
2019	Task 7: Write final S&T report, DSAP template, and advertise to agency field offices	Labor	09-30-2019	$15,000.00	

A simpler example of a DSAP could be the following: “Collect initial conditions on reservoirs – collect top of pool water temperature data around March 30 (one day between February 15 to April 15) to capture initial weakly-stratified conditions for two-dimensional (2-D laterally averaged) water quality models such as CE-QUAL-W2. During winter, initially take data from a marina and assume bottom water temperature is 4 degrees C, or safely cut a hole in the ice and sample from the surface of the ice. If open-water sampling can be done, use a multi-parameter water quality instrument (Hydrolab or YSI) from a boat to measure temperature (T), DO, pH, and electrical conductivity (EC) profiles. Collect end of summer conditions by collecting a reservoir profile at each lake or reservoir during August or September. The low DO profile and associated temperature profile are useful for determining if anaerobic conditions have set in.”

For riverine and reservoir drought data collection in support of water quality and aquatic habitat analysis and modeling, the following should be addressed before going to the field: What, where, when, how, with what equipment, to what standards and quality assurance/quality control, and who should collect the flow, sediment, and water quality data? It's getting dry out there so start up the truck, throw the gear in back, and head to the lake is a commonly used yet inadequate approach to data collection. Inorganics or organics? Glass or plastic? Calibrations? Holding Times? Duplicates, blanks, rinsate blanks, replicates, splits, spikes, round-robins, and references? Half meter, one meter, five feet, surface, grabs, composites, or continuous sampling? Monthly, bi-weekly, weekly, daily, hourly, continuous, or telemetered? USGS, EPA, or standard method protocols and procedures? Meta-data, recording procedures, and chain-of-custody? There are many decisions to make before sampling. And there is also planning required for the laboratory sampling analysis, data processing, and archival of data for future uses.

For any water quality data base demonstration project, the approach needs to be kept simple and easy to communicate. Start with a few data collection sites on a smaller watershed or watershed portion. Then document the data collected with appropriate metadata, such as the example site location picture shown in figure 6. Without a picture, the exact measuring site location in figure 6 would be difficult to locate again on a return field trip. Digital pictures supplement the marking of data collection sites marked by global positioning system coordinates.

Automation of the data collection and data processing system is the ultimate goal. However before automation can be fully implemented, add-ins might initially be used to partially automate the processing of existing and future data sets.

Figure 6.—Example of a metadata site figure for potential data collection for DSAP or WaterSMART that might be included in a water quality demonstration project.

How Add-ins Work

As an example, DataWolff (http://wolffwareltd.com/) is a general data analysis program that runs within Microsoft Excel and is tailored with a data analysis script. Once an analysis script is configured, large sets of data files are consistently processed. The output is contained in one or more Excel workbooks that contain the data, calculations, and charts that enable making timely decisions from the data in an automated procedure. The DataWolff website summarized the advantages of automation and procedures for more automated processing of data as well as examples.

Advantages of Automation

1) Analyze More Data – The benefits gained from automated data processing are much greater than the labor savings alone. With automated processing, it is possible to analyze more data and gain the benefits from the information in data.

2) Simplify Complicated Analyses – In many cases, complex statistical analyses require engineers and scientists to spend a great deal of time with manual spreadsheet operations. Automation eliminates these operations and enables analyzing and gathering more information from data.

3) Eliminate Data Processing Errors – Manual data processing is time consuming and commonly leads to errors. Automating analyses eliminates such error.

Environmental monitoring stations provide large quantities of information with both continuous and periodic measurements. The examples below demonstrate the automated creation of simple and useful workbooks with DataWolff from EPA STORET data and from meteorological data. Other tasks that DataWolff has automated using environmental monitoring data include: computing and plotting exceedance curves; computing hourly and daily averages from data sampled on a five-minute frequency; synchronizing data with disparate time stamps; and providing interactive Excel charts that enable visualizing removal of outliers and correcting dissolved oxygen data that are skewed because of fouled sensors.

EPA STORET Data and Excel Workbook Example

EPA provides water quality data from many monitoring sites located throughout the country that are available for download from their STORET web site (http://www.epa.gov/storet). Manually creating profile and trend plots, and performing other analyses with these data requires significant amounts of time.

Examples of workbooks are described below that were automatically created with DataWolff from a STORET legacy data file. The workbooks contain animated profiles for both dissolved oxygen and temperature in a forebay reservoir.

The Excel workbooks referenced in the links below contain charts with animation, zoom, and scroll features that require macros. Follow these steps to download and run the workbooks:

1) Download and unzip the workbook;

2) Start excel (must be excel 2000 version 9.0 or later);

3) Check your excel security settings which must be set at medium security or lower for the macros to be enabled; and

4) Load the workbooks.

Click here for instructions on adjusting your security settings.

- Dissolved Oxygen Profiles, 1981.xls – Dissolved oxygen profiles in the forebay of a reservoir for one year of data *(236KB .zip archive)*.

- Dissolved Oxygen Profiles, 1981-1984.xls – Dissolved oxygen profiles for four years of data. With DataWolff, once an analysis script is created, creating profiles for many years of data requires no additional time *(204KB .zip archive)*.

- Temperature Profiles, 1981.xls – Forebay temperature profiles for one year of data *(236KB .zip archive)*.

EPA's water quality portal data discovery tool utilizes open source R, a statistical programming language and several add-on packages to visualize the data selected from the portal and to assist users in data analysis.

Meteorological Data Example

Many meteorological stations provide hourly measurements of ambient weather conditions which include air temperature, dew point temperature, cloud cover, and precipitation. Download the attached workbook to view a trend chart created with DataWolff that enables zooming-in and scrolling through three years of data.

- Meteorological Data - Air Temperature vs Time.xls – Contains a DataWolff trend chart of air temperature that enables you to zoom-in and scroll through the data *(796KB .zip archive)*.

Interface Example

One important feature of a good database is an interface needed for creation of a data product that quickly allows retrieval and exportation of specific data. One example of a water quality interface is the Northern Water Quality Retrieval product shown in figures 7 and 8 and found at the following link:

- http://www.northernwater.org/DynData/WQDataMain.aspx (Esther Vincent, Northern Water, Personal email communication or website https://www.northernwater.org/WaterQuality/WaterQualityData.asp.

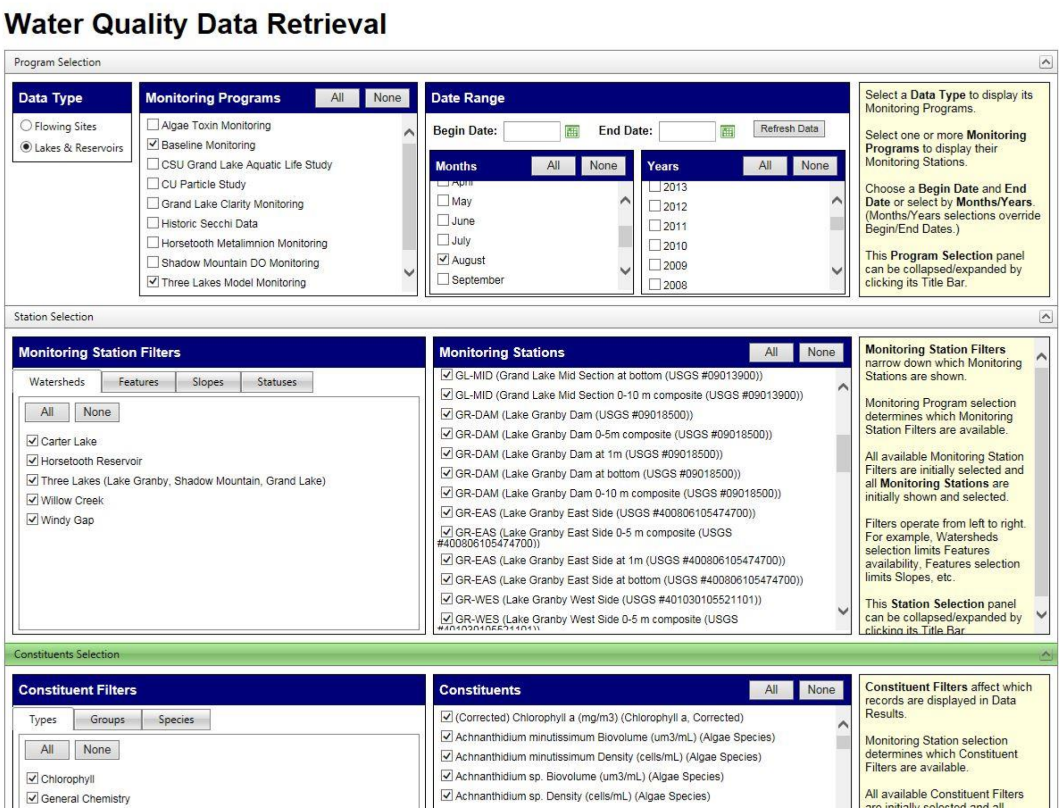

Figure 7.—Northern water quality data retrieval data type, monitoring program, date range, and monitoring station filters.

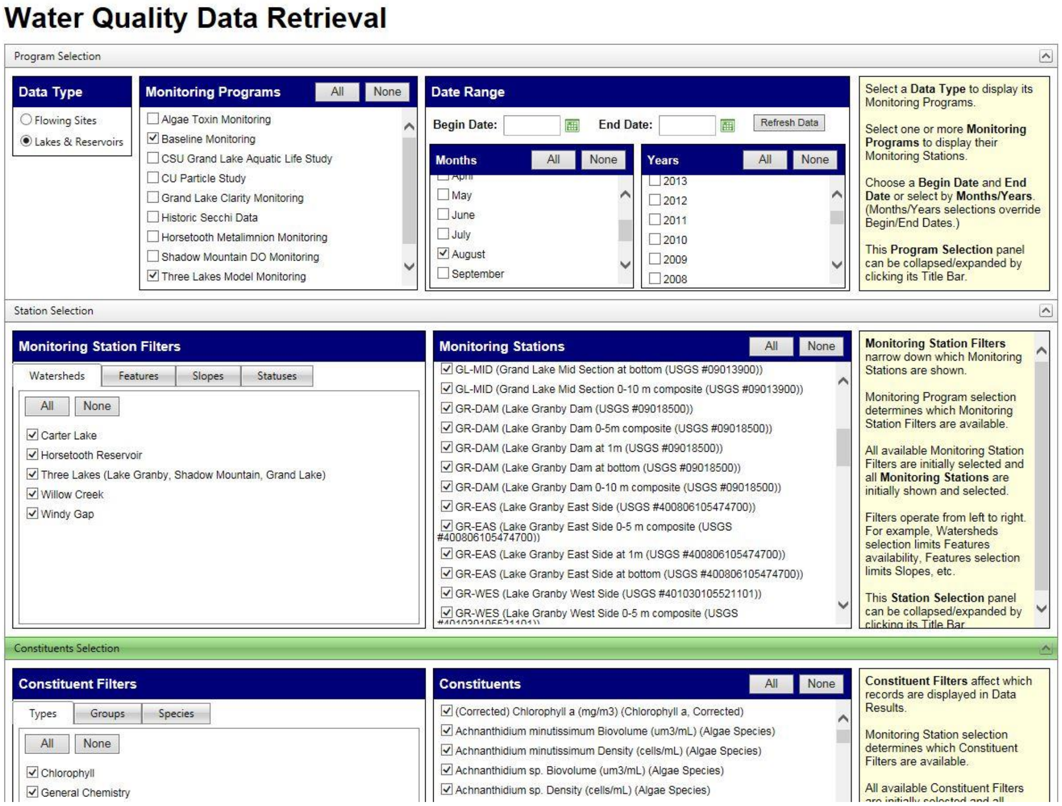

Figure 8.—Northern water quality data retrieval
constituent filters and constituents.

Summary

To maximize the value of data that has and will be collected for riverine and reservoir water quality modeling, a simple consistent data format and a reduced list of parameters needs to be developed. Modelers, engineers and scientists who collect and process data often do not have the necessary computer science and database management skill sets or do not have access to a database management team. That forces the users to resort to commonly-used off-the-shelf Office Suite products such as MS-Excel or MS-Access (personal communication with Dimitri Vidergar, 2015, and Denise Hosler, 2016). The users use available software tools with the intent to transfer the data to a database manager and assume the data will be taken care of and archived appropriately, which rarely happens systematically. Currently, the data owner needs to have access to an alternative to placing the data into MS-Excel spreadsheets or to develop a methodology in MS-Excel or add on to MS-Excel to make that task easy to integrate the data within a designated database. The approach needs to build on and use existing tools and databases such as those being developed in MS-Excel or as MS-Excel add-ons. The plotting capabilities within MS-Excel should be harnessed to provide a consistent look and feel to users. However, additional post-processing tools for specific models might be used and may need to be accommodated with a flexible database structure tool that allows exportation of MS-Excel .csv format or other formats. Future data archival should consider software defined anything virtualization techniques (see Appendix A for an explanation of SDx).

The quantity of online process and test data for all industries has increased exponentially in the past decade requiring automation in a software defined platform such as SDx. Typically a significant amount of data remains unused because performing the analyses to convert the data into useful information requires too much time.

Recommendations

Reclamation needs to develop a demonstration project for a shared centralized transparent data repository with a plan for consistent future data analysis. A structured non-proprietary opensource software virtulization platform format is recommended for future needs. Initially, add-on packages to database software are the recommended procedure to accommodate immediate analysis of data. Add-ons developed for a particular piece of software should be modular and flexible to accommodate software changes over decades.

Initially a propietary post processor, such as DataWolff, might be used to process the data stored in an MS-Excel format for output in a common .cvs format (text file that is comma separate value format for use with multiple opensource software). However, a specialized add-on might be developed for a future opensource software package or future virtualization techniques such as SDx. Much of the world of computing is familiar with MS-Excel which has rapidly turned from a spreadsheet software to a database software package. Initially enhance that capability with MS-Excel add-ons. DataWolff is a propietary software package that allows large amounts of data to be processed quickly. See the following link: http://www.wolffwareltd.com/ portions of which are summarized as follows.

DataWolff is software that streamlines data analysis in various technical fields. DataWolff is general purpose data analysis software that can be tailored to specific needs with an analysis script. DataWolff streamlines and automates large or repetitive analyses to enable maximizing the value of the information contained in the data. This provides insight that may otherwise be lost and maximizes productivity. DataWolff, an Excel add-in, is a general purpose data analysis engine that is configured for a specific application with an analysis script. The analysis script configures DataWolff to import data, perform calculations, and create charts. Once a script is configured, large sets of data files are consistently processed to save time.

References

Bender, M.D., G.E. Hauser, M.C. Shiao, and W.D. Proctor. 1990. "BETTER: A Two-Dimensional Reservoir Water Quality Model, Technical Reference Manual and User's Guide, Report No. WR28-2-590-152." Tennessee Valley Authority, Engineering Laboratory, Norris, Tennessee.

Bender, Merlynn D. January 2016. "Water Savings Verification Results for Richmond Irrigation Company Upper High Creek Canal Project." U.S. Department of Interior, Bureau of Reclamation, Technical Service Center, Special Report, prepared for WaterSMART Water and Energy Efficiency Grant program in cooperation with the Policy Administration Office, Denver Colorado.

Bureau of Reclamation. May 2009. "Guidelines for Collecting Data to Support Reservoir Water Quality and Hydrodynamic Simulation Models." U.S. Department of the Interior, Bureau of Reclamation Technical Service Center, Denver, Colorado. http://www.usbr.gov/tsc/techreferences/mands/mands-pdfs/hydromodels.pdf.

Bureau of Reclamation. August 2010. "Guidelines for Collecting Data to Support Riverine Hydrodynamic and Water Quality Simulation Models." U.S. Department of the Interior, Bureau of Reclamation Technical Service Center, Denver, Colorado. http://www.usbr.gov/tsc/techreferences/mands/mands-pdfs/River_water_quality_manual.pdf.

Cole, T.M., and S.A. Wells, 2002. "CE-QUAL-W2: A Two-Dimensional Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.1, User Manual." Draft Instruction Report EL-02-1, Prepared for the U.S. Army Corp of Engineers, Washington, D.C.

Cole, T.M., and S.A. Wells. 2006. "CE-QUAL-W2: A Two-Dimensional Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.5, User Manual, Draft Instruction Report EL-06-1." Prepared for the U.S. Army Corp of Engineers, Washington, D.C.

Datawolff, 2016, http://wolffwareltd.com/

Debele, B., Srinivasan, R., and Parlange, J.Y. 2006. "Coupling Upland Watershed and Downstream Waterbody Hydrodynamic and Water Quality Models (SWAT and CE-QUAL-W2) for Better Water Resources Management in Complex River Basins." Environmental Modeling Assessment 13 (1): 135–153." http://www.springerlink.com/index/10.1007/s10666-006-9075-1.

Environmental Protection Agency, May 20, 2016, “Data Portal and Exchange,” https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange)

Hauser, G.E., and G. A. Schoh. November 2002. “River Modeling System v4 – User Guide and Technical Reference, Report No. WR28-1-590-164.” Tennessee Valley Authority River System Operations and Environment, Norris, Tennessee, (Revised June 2003, in draft).

Hosler, D. February 3, 2016. “Reclamation Invasive Mussels Research Coordinator.” Personal Communication about mussel database using MS-ACCESS, Bureau of Reclamation Technical Service Center, Denver, Colorado.

Loginetics, 2016, http://www.loginetics.com/

Pruitt, T., Bracken, C., and Gangopadhyay, S. September 2014. “Forecasting Crop Irrigation Water Requirements, Fiscal Year 2014 Project ID 1472.” Bureau of Reclamation, Science and Technology Program, Denver, Colorado.

Sullivan, A.B., Snyder, D.M., and Rounds, S.A., August 3, 2009, “Controls on biochemical oxygen demand in the upper Klamath River, Oregon,” Chemical Geology 269 (2010) 12-21, journal homepage http://www.journals.elsevier.com/chemical-geology and Published by Elsevier B.V., doi:10.1016/j.chemgeo.2009.08.007, pages 12-21.

Texas Water Development Board's Brackish Resources Aquifer Characterization System (BRACS), 2016, http://www.twdb.texas.gov/innovativewater/bracs/.

Vidergar, D. February 3, 2016. Reclamation Biologist’s personal communication about Arrowrock database using EXCEL. Bureau of Reclamation, Pacific Northwest Fisheries.

Vincent, Esther, Northern Water Database Specialist, Personal Communication email (December 22, 2015) with database link http://www.northernwater.org/DynData/WQDataMain.aspx found at website https://www.northernwater.org/WaterQuality/WaterQualityData.aspx

Appendix A – Software Defined Platforms (SDx) for Data Archival

This OWQDI scopes options for development of a consistent water quality database platform structure. By investigating the data systems developed by others, a proposed format for consistently storing reservoir water quality data is being suggested for both current immediate needs and future development. Currently water quality databases are project specific, spreadsheet-based, and inconsistent. The benefit of a consistent water quality database is that it allows better data management, comparison between agencies, and greater access by agencies as well as the public. SDx are the future of data virtualization.

Future of Data Virtualization

Software defined platforms for future data virtualization are discussed as follows:

Literature Research: https://fcw.com/microsites/2015/snapshot-cdwg-virtualization/02-software-defined-platforms-define-future-of-virtualization.aspx. See the file FCW_SDxOct2015.pdf in the same directory as this. *Federal Computer Week*, October, 2015, Volume 29, Number 17, page 12. Title, "Software-Defined Platforms Define Future of Virtualization" SDx article below:

> As government has pushed the mantra of "more bang for the buck," virtualization has become an accepted way of doing IT. Server virtualization is transforming the data center environment. With that comes storage and network virtualization as well. With physical infrastructure now so abstracted, the traditional approach of throwing more hardware into the mix to solve problems is being turned on its head.
>
> Software managed IT environments are now seen as a large part of the future. Software-defined networking (SDN) is an emerging practice. Software-defined storage (SDS) is quickly gaining pace. Emboldened by these innovations, software-defined data centers are just over the horizon.
>
> Inevitably, that has led to thoughts about software-defined anything (SDx). As the dependency on physical hardware is reduced, so the thinking goes, software can manage entire environments. And that vastly increases the flexibility and agility with which agencies use various IT resources.
>
> What once took days, weeks or months to set up and configure with physical IT can be deployed in hours, minutes or, in some cases, seconds with the virtualized world of SDx. It's also much easier to match those resources to the requirements, doing away with the costly over capacity that often has to be built in to physical environments to ensure capacity for expected future demand.

SDx is certainly more concept than reality right now, but the idea is quickly gaining ground. In 2014, market researcher Gartner listed SDx as of the 10 top technologies to watch and include as part of strategic planning. Other technologies include the Internet of Things, mobile, smart machines and various cloud-based infrastructures.

Likewise, the Institute of Electrical and Electronics Engineers (IEEE) Computer Society said interoperability issues and standards for SDx would be a top priority for 2015. Various standards groups such as the Open Networking Foundation, the Internet Engineering Task Force and the International Telecommunication Union are already working on the appropriate specs.

Government agencies are dipping their toes into specific software-defined technologies such as networking and storage. The Defense Information Systems Agency (DISA), for example, has set up a software-defined network working group. It included money in its FY16 budget request to launch pilot programs to see how Defense Department networks can use SDN. Other funds would be used to develop a Technology Environment that will evaluate and characterize new technologies, including SDx.

Researchers at the Idaho National Laboratory (INL) have already gone further. They've developed a proof of concept to see how to apply SDx to the laboratory's business environment. It emulated the use and security of INL business systems accessed by a large number of virtual machines, with software providing control intelligence that would otherwise be embedded in hardware.

In a recent issue of Government Computer News, Wayne Simpson, the INL innovation architect, and research scientist Tammie Borders, described how the prototype solution they developed showed SDx "can be used to improve security, repeatability of process and consistency in results." They concluded that by adopting SDx approaches, organizations could reduce employee workload, improve security controls and optimize existing IT investments.

"As the dependence on hardware for the intelligence to implement access and security controls diminishes, organizations must overcome traditional thinking and drive changes in regulatory restrictions," according to Simpson and Borders. "As these challenges are addressed, SDx will become more widely adopted and will change how information is accessed and consumed worldwide."