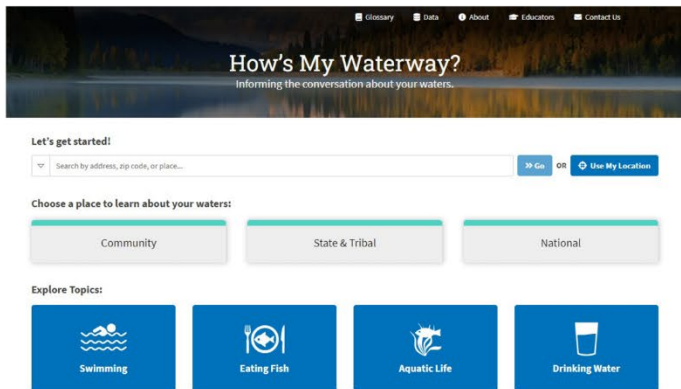


Tribal Data Management for WQX Submission

Supplement to the Clean Water Act Section 106 Tribal Guidance



How's My Waterway?
Informing the conversation about your waters.

Let's get started!

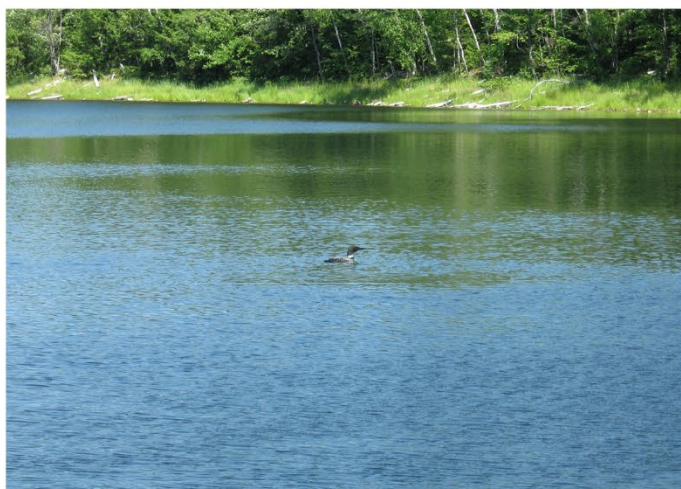
Search by address, zip code, or place... [Go](#) OR [Use My Location](#)

Choose a place to learn about your waters:

- Community
- State & Tribal
- National

Explore Topics:

- Swimming
- Eating Fish
- Aquatic Life
- Drinking Water



Water Quality Data Upload with WQX

There are two options for you to share your data using WQX. You can choose a standard web-based application (WQX Web) that uses Microsoft Excel spreadsheets or you can choose to create a custom submission application using WQX XML schema through Exchange Network Nodes or Node Clients. The best approach for your organization depends on in-house technical expertise, data volume and resources. The WQX Web application requires less technical expertise and manual steps to upload. Custom WQX XML submission applications can be more efficient for organizations with larger databases and a need for automated submissions.

- [WQX Web - Resources for a Standard Upload](#)
- [Exchange Network - Resources for a Custom Upload](#)

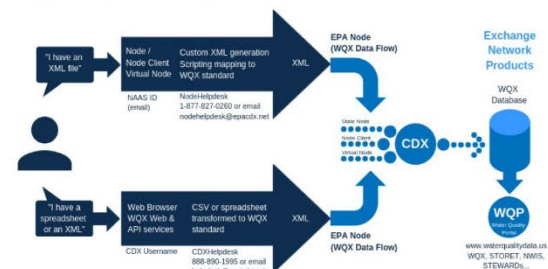


Table of Contents

NOTICE	ii
INTRODUCTION.....	1
DATA MANAGEMENT.....	1
DATA TYPES	2
RECORDING AND MANAGING METADATA	3
GENERAL DATA MANAGEMENT CONSIDERATIONS	4
GENERAL DATA MANAGEMENT BEST PRACTICES	7
MODEL DATA MANAGEMENT THAT FACILITATES WQX SUBMISSIONS	9
DATA MANAGEMENT CONCLUSIONS.....	11
DATA SUBMISSIONS TO WQX.....	12
ESSENTIAL ELEMENTS AND DESIGN OF WQX.....	13
WAYS TO SUBMIT DATA	20
RETRIEVING DATA FROM THE WATER QUALITY PORTAL	22
WQX SUPPORT AND RESOURCES	22
APPENDIX A: KEY TERMS.....	25

Notice

This document has been reviewed and approved for publication by the Environmental Protection Agency (EPA). Mention of trade names or commercial products or services does not convey and should not be interpreted as conveying official EPA approval, endorsement, or recommendation. Any use of words such as “should” or “may” are not intended to be mandatory language, rather, these are statements of suggestion that a tribal program may or may not adopt depending upon available resources.

Cover Photo Credits (Clockwise from Top Right)

Mole Lake WI wild rice, photo courtesy of Robyn Delehanty

Screenshot of EPA’s Water Quality Data Upload with WQX page

Forest Lake Potawatomi WI, photo courtesy of Robyn Delehanty

Screenshot of the homepage of EPA’s How’s My Waterway application

Introduction

The [Clean Water Act Section 106 Tribal Guidance](#) provides a flexible framework for Tribes and EPA regions to ensure that Tribes' water quality program goals are met while being consistent with Section 106 of the Clean Water Act (CWA), the Governmental Performance and Results Act (GPRA), and other federal requirements. The *CWA Section 106 Tribal Guidance* outlines water quality reporting requirements and data expectations for all tribal programs receiving Section 106 funds. These requirements help Tribes collect and manage critical data and information for effective management of water quality programs.

This document is intended to supplement the CWA Section 106 Tribal Guidance by providing useful suggestions and examples for Tribes on how to establish a data management system that reflects tribal water quality goals and manage the system, so the data are accessible and reliable. Good data management empowers Tribes to use the data to determine the physical, chemical, and biological conditions of a waterbody to measure and identify water quality trends to improve and maintain water quality. An organized and thorough data management program provides the foundation for measuring and achieving water quality environmental results while helping Tribes meet EPA data reporting requirements. Appendix A: Key Terms provides a glossary of terms used throughout this document.

This document provides basic data management principles that will be helpful to all Tribes and all water quality programs. This document presents data management principles to assist Tribes with getting data into a usable format to begin putting together data assessments (such as statistics and data summaries). This document will center around data management practices that will help prepare data sharing through the Water Quality Exchange (WQX). For information related to assessment data and assessment decisions which are reported in Water Quality Assessment Reports or placed into the Assessment and Total Maximum Daily Load Tracking and Implementation System (ATTAINS), please see the [Tribal Assessment Modules](#).

Data Management

Data management:

1. Is the process of recording, storing, organizing, maintaining, and safeguarding data and information.
2. Allows programs to review, summarize, and analyze results to make decisions.
3. Saves time and helps ensure data integrity and quality.
4. Is a crucial part of a Monitoring Strategy as it lays out how the water quality program plans to manage data long-term.

Good data management can aid in the primary goal of monitoring water quality to protect traditional tribal lifeways and leave a legacy of quality data for future data practitioners (Figure 1). In this process, Tribes manage the data to make water quality decisions. These data can be shared with others but does not dictate the process of managing it.

Figure 1. Components contributing to protection of traditional tribal lifeways.

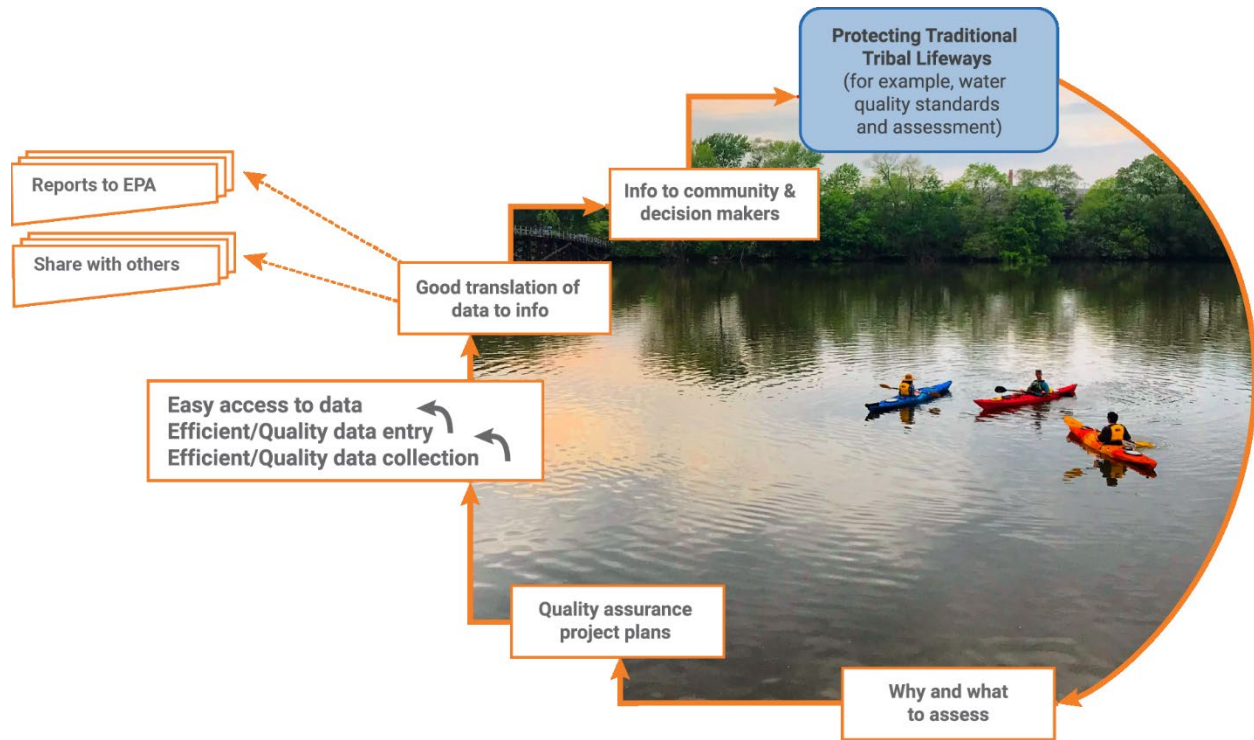


Figure Credit: Figure was provided courtesy of Penobscot Nation Water Resources Program. Photo Credit: Laura Shumway

Data Types

The management of data depends on the types of data a program collects, and the goals for reporting or analyzing those data. For example, Tribes and Nations can collect water quality data in a variety of ways, including:

- *In-situ* via field probes and other sensor devices.
- Observations of habitat and stream condition.
- Collection of biological samples such as benthic macroinvertebrates or fish.
- Measures of physical parameters such as flow.

These data are often used in concert with other data from landscapes or dischargers for example or are used to calculate additional measures in the form of metrics or indices which can indicate a water's condition. Each of these distinct data types will look a little different, as they are capturing various kinds of information. However, to use the data to help answer management questions, or to draw conclusions for formal assessments, different data types may need to be brought together in some way. Generally, the more complex a monitoring and assessment program is, the more that program will have to invest in data management and planning to produce useful information. As shown in Figure 1, ensuring quality data, and documenting that process, allows for easy understanding and greater utility of the data and their conclusions used by decision makers and the community. This, in turn, allows for successful protection of water quality and feeds information back into the water quality protection cycle.

All water quality monitoring data types (observations, measurements, and biological data) are shareable through [WQX](#) and are made available for download or retrieval via the [Water Quality Portal \(WQP\)](#). Sharing data comes with its own data management requirements that can also benefit from upfront planning for the intended data use goal. Regardless of the type of data collected, each sort of data should encompass metadata, or “data about the data.” The Recording and Managing Metadata section discusses general metadata requirements and specific metadata collected for each data type.

Additionally, the assessments and conclusions drawn from the data themselves also become data that are managed and shared accordingly. This includes sharing information on water quality conditions with tribal members or the tribal council, submitting a Water Quality Assessment Report or by generating assessment decisions that can be loaded into ATTAINS and shared through How’s My Waterway (see the [Tribal Assessment Modules](#)). The conclusions drawn from the data, and the results that are shared, allow Tribes to reconsider their current water quality monitoring goals, and adjust their sampling collection and analysis to meet new goals to protect the community based on information learned during the previous iteration of sampling (Figure 1).

Recording and Managing Metadata

Data management is more than just ensuring that the values of measurements and observations are properly captured. To be good stewards of data, metadata or “data about data” are documented and stored about those measurements, the methods deployed, and the programs in general. This helps others confidently use the data and future counterparts maintain continuity in data use and management. Additional metadata typically include:

1. Where was the sample or measurement collected?
2. When was it collected?
3. What was collected?
4. How was it collected?
5. Why was it collected?
6. Who collected it?

While the questions above describe common metadata collected for all sample types, each individual data type may require additional metadata to help describe the samples collected. Please see Table 1 for major data types and differences among their associated metadata.

Collection of metadata ensures that the information provided by the data is accessible and understandable. The purpose of data management is to convert monitoring results into a useful set of data to use for making decisions about the Tribe’s water quality monitoring program and protecting traditional tribal lifeways.

Table 1. Various data types and their different data elements or metadata requirements

Data Type	Location (Datum, Equipment)	Sample Methods and Equipment	Analysis Methods	Species Info
In-Situ Sensors	X	X		
Field- Biological	X	X	X	X
Field- Physical	X	X		
Field- Chemical	X	X	X	
Laboratory- Chemical	X	X	X	
Geospatial	X			

A note about use of Indigenous Knowledge and the collection of data

Tribes may elect to use Indigenous Knowledge to inform aspects of their tribal water quality program. The Section 106 Program is committed to respecting tribal Indigenous Knowledge and acknowledge that this information is owned by the Tribe. Tribes that use Section 106 funds to collect Indigenous Knowledge will not be required to report that knowledge as part of their grant requirements. Tribes are expected to meet the reporting requirements of the Section 106 grant but are not expected to share the underlying Indigenous Knowledge used to inform water quality objectives and management practices. Programmatic activities that reflect decisions based on Indigenous Knowledge, such as location of sampling sites and monitoring data, are still subject to grant reporting requirements; however, the Indigenous Knowledge-based rationale for site selection does not have to be reported to EPA. Where applicable, Tribes should engage with their Project Officers early to discuss how Indigenous Knowledge can be protected if it is used to influence policy and regulatory decision making. For more information on the elective inclusion of Indigenous Knowledge within the Section 106 Program, please see the Indigenous Knowledge section of the [CWA Section 106 Tribal Guidance](#).

General Data Management Considerations

Effective data management procedures should be considered and planned for during the sampling design phase of a monitoring program (please see Chapter 5: Development and Implementation of a Monitoring Program of the [CWA Section 106 Tribal Guidance](#)), and should be documented in the Monitoring Strategy. For more information on how to create a Monitoring Strategy, please review the [Developing a Tribal Water Quality Program Monitoring Strategy Supplement](#). Without a plan in place, data may begin to accumulate in an unmanaged way, making the eventual use of that data more challenging in the future.

Data management procedures are processes that allow for a controlled and standard way of:

- Deliberating,
- Managing,
- Manipulating,
- Organizing,
- Storing, and
- Sharing water quality monitoring data.

It is often helpful to map out the lifecycle of data (See Figure 1 and Figure 2). Many programs submit samples to a water quality laboratory for analysis and receive results in a different format than they might produce from their field forms, logger files, or biological samples. Programs with multiple datasets will benefit from a well-defined process for combining these data, helping them avoid challenges when dealing with large amounts of data later. To ease the data combination process, a program may consider supplying the laboratory with a data template to help return laboratory results in a format that will be more easily incorporated into their data management system or by stipulating the format of the data deliverables received. Programs facing these data combination challenges may consider allowing WQX to combine their data for them, by creating different custom imports for each reproducible dataset. Then a Tribe can download a combined program dataset in a single table to store, analyze, or use further.

Figure 2. Lifecycle of Data



Similarly, many programs evaluate and interpret data to determine decisions based on the attainment of thresholds or water quality standards. How those datasets are assembled for analysis of those thresholds or standards should be part of the planned lifecycle of the data. Often a program can access a prescribed roadmap of data management in a Data Management Plan (DMP), a voluntary document that can be a helpful tool to document how to deal with data after collection. For example, if a program collects data using GIS enabled apps, or plans to bring data into a GIS application for analysis, then mapping the steps for that data in a DMP would be beneficial to ensuring that data path can be navigated smoothly. Effective DMPs will also help current and future staff better understand the purpose for the data and methods for collecting, maintaining, storing, and sharing data. Again, a DMP is not a requirement of the Section 106 Program.

Successful management of data begins before samples are collected. To determine what data management procedures may be right for the data, consider the following questions:

1. What are the goals for this project?
 - i. What questions do the decision makers or tribal administration have?
 - ii. What are the research questions?
 - iii. How will the program know when enough data have been collected to answer the questions asked?
 - iv. When and how often will the information need to be produced?
 - v. What are the formats that information will ultimately need to be presented in?

2. What different data types will be produced by the project?
 - i. What are the different paths these data will take from the field or laboratory to the management platform?
 - ii. What datasets need to be combined for analyses? Where and when does this happen?
 - iii. What metadata will need to be managed alongside the result data?
3. Who will be responsible for data management and data analysis?
 - i. What support will staff need to properly manage the data?
 - ii. Which pieces do you want to manage in-house, and which pieces can be done by WQX, WQP, or other systems?
4. What resources are available to support the data management program?
 - i. What software options or approaches are best suited to the data management needs?
 - ii. Is there a budget to support data management?
 - iii. What are the best options given resource constraints?
 - iv. Are there additional grants available that are specific to data management?¹
 - v. Are there communities of practice (other WQX users for example), Consortia, or workgroups available to help answer questions and discuss approaches?²
5. What are your program's data management requirements?
 - i. Who will need access to review or enter data?
 - ii. Does the program need a data management system? (For database type definitions, please see the Data Management Procedures and Considerations section.)
 - iii. How will data be prepared, extracted, and submitted to WQX?
 - iv. What other systems might the data need to be able to integrate with?
 - v. Will there be an electronic form to support data entry, or will data first be entered directly onto paper spreadsheets?

Working through the above questions will help determine which procedures are needed to best manage data. Tribes may also want to consider developing written procedures for data handling, records management, and documentation to help track how the water quality monitoring program will operate and how they will record and track the information. For records management, it is good practice to keep electronic backup copies of all files. This is helpful in case any physical or digital copies of the data and reports are lost.

Once a Tribe has decided what data management procedures to follow, one approach is to write out all the steps of the data management process as if they were explaining the steps to a new person, which can be formalized as a Standard Operating Procedure (SOP) or as part of the overall DMP. The DMP can

¹ The Exchange Network Grant Program is available for Tribes managing data to submit to WQX. For more information please visit, <https://www.epa.gov/exchangenetwork/exchange-network-grant-program>

² Tribes may apply for assistance via the Exchange Network Mentorship Program – Request Tribal Assistance from the TXG Google Doc, <https://docs.google.com/forms/d/e/1FAIpQLSd3TJLv1xsGFeqUZLdpG4yyTzSTwep7fzJiz-VfpGfchXErrw/viewform>

be a living document and updated anytime the procedures change. This can remind staff what the procedures are and how to communicate them to others.

General Data Management Best Practices

As data management procedures are implemented, a few principles are important to keep in mind:

Data Requirements and Integrity

Within databases and spreadsheets, it is important to agree on what types of data to track and which data elements to require. For example, if there are two different pieces of equipment that measure the same parameter, staff may want to record which piece of equipment took each measurement. Sometimes, two different methods might even produce different units. If these kinds of situations arise, staff may want to require that these fields be filled out in the field or once back in the laboratory. In a database, enforcing these requirements is easy. In spreadsheets, lookup ranges and data validation rules can be supplied that limit entry options, or data integrity rules can be recorded on a document and made available as a reference for staff completing data entry or quality review functions.

Manage Primary Data Elements in Unique Tables

Data Managers may consider managing different data elements such as project information, sampling locations, methods, equipment, and personnel in unique lookup tables where detailed information about those entries can be stored. This process allows the user to store all the data that should be documented about a monitoring program in compact, efficient tables. The use of Identifiers, such as Monitoring Location IDs, can be used to link the related information to results by including only the IDs in the results table. In a relational database, these relationships can be strengthened and enforced by table connections and the use of those Identifiers. Managing metadata in separate but related tables will also help in the data transfer process to WQX, as this is how WQX collects and stores information internally.

Naming Conventions

It is an important practice to begin using common naming conventions, especially in tracking what types of water quality parameters are sampled. If the Tribe tracks dissolved oxygen by a consistent name, it will be much easier to find, manipulate, and analyze all the measurement values associated with this parameter. Another example may be making sure a consistent naming scheme is in place for identifying all the monitoring stations. In a database, drop-down lists based on valid values can make it very easy to implement common naming conventions.

Standard Data Formats

Having standard formats for data elements will also make it easier to manipulate data once data are ready for the user to analyze. For example, all dates need to have a standard format of YYYY-MM-DD or all latitude and longitude coordinates should be in a standard format (for example, decimal degrees). Additionally, to ensure consistency in datasets, use standardized column headers (for example, do not use "TEMP" in one spreadsheet and "TEMPERATURE" in another). It is helpful to also define a consistent standard format for the spreadsheets (for example, using the same column order).

Standard Data Structures

The overall structure of datasets should also be a consideration. Certain standard data structures are referred to as "machine-readable" meaning they follow a consistent standard layout that allows

computer programs to read and interpret the information. Adopting a machine-readable structure will help ensure that datasets can be easily shared or used in applications that can help turn them into useful information. “Matrix” style datasets (aka “tidy”, “wide”, “flat” data are one such standard format, very common to analysis datasets. In this format each row represents a single observation, and each column a different result from that observation (see Table 2 for an example). These formats are great for quickly analyzing and viewing results, but they don’t usually leave room for the metadata. Tribes may also choose to store data alongside their metadata in a “stacked” format like the format of WQX Results table, where each row represents a single result, and each column is a piece of information about that result (for example, a value, unit, or method) (see Table 3). The latter structure is very useful for data management and storage of associated metadata, whereas the wide formats are useful for analyses that explore relationships between parameters in the data, such as stressor-response analyses. Avoid structures that needlessly separate data, like placing data from each station in a separate section or tab of a spreadsheet. Such separation will make sharing or analyzing the dataset across stations more challenging.

Table 2. An example of the same dataset in a “tidy” or “wide” format

Activity Identifier	Activity Start Date	Monitoring Location Identifier	Iron	Lead	Manganese
nwiswi.01.99208821	8/28/1992	USGS-04072050	1400	1.5	510
nwiswi.01.99208822	8/28/1992	USGS-04072050	1800	1.2	650
nwiswi.01.99208826	8/28/1992	USGS-04085110	2200	1.2	890
nwiswi.01.99208856	8/29/1992	USGS-04085475	1600	1.4	480
nwiswi.01.99407330	9/19/1994	USGS-04063700	120		250
nwiswi.01.99407332	9/19/1994	USGS-04063700	5000	1.2	2100
nwiswi.01.99407338	9/21/1994	USGS-04080798	1400	1.2	820
nwiswi.01.99407340	9/21/1994	USGS-04080798	1200	1	2400

This format is often used for data analysis. In this style, each record contains all the observations of a single sampling activity, and often holds only values and not metadata.

Table 3. An example of data in a “stacked” or “tall” format

Result Identifier	Characteristic Name	Sample Fraction	Measure Value	Unit
NWIS-114877794	Stream width measure		3	ft
NWIS-114877795	Temperature, water		16.8	deg C
NWIS-114877797	Stream flow, instantaneous		0.19	ft3/s
NWIS-114877798	Specific conductance	Total	696	uS/cm @25C
NWIS-114877799	Acidity, (H+)	Total	0.00001	mg/l
NWIS-114877800	Oxygen	Dissolved	11.4	mg/l
NWIS-114877801	pH	Total	8.3	std units

This is how data are stored in WQX Web templates and served by the WQP. In this style, each record contains a single observation or measurement and its associated metadata.

Model Data Management that Facilitates WQX Submissions

Spreadsheets

It is possible for a program to properly manage its water quality data using spreadsheets in preparation for submitting data to WQX. For best results, it is recommended that spreadsheet users adhere to the general data management best practices laid out in the General Data Management Best Practices section of this supplement.

Spreadsheet users should ensure their data structure is “machine-readable.” This generally means that datasets have: a) a single header row at the top of the dataset that names each field; and b) a standard structure (see descriptions and examples of both “tidy” and “stacked” formats; WQX Web templates are in a “stacked” format) (See example in Table 3). By default, spreadsheets allow for direct manipulation of the data in the tables, so this can lead to both mistakes or a breakdown of entry and storage conventions. Programs should try to be consistent in the naming conventions applied across stations, parameters, and other data. Avoid using any abbreviations or alternate naming styles across datasets or programs as it will lead to difficulties in later integration of the data. As a failsafe, users may consider creating a sheet in the spreadsheet file that holds allowable value lists for each field. Users can then reference this sheet to create dropdowns in the spreadsheet, helping to avoid data entry errors or inconsistent entries.

Programs that are just getting started with their monitoring program, or are looking to improve their data management approach using spreadsheets, may consider adopting the [WQX Web Template Files](#) (or a simplified version) for their data management programs. By using the templates, users can ensure that all WQX-required data elements are complete at the time of entry, and that value-naming conventions match from the onset. Once a user enters all tribal data into the WQX Web templates, then the data will be ready for submission through the WQX Web application.

Relational Databases Lite

A relational database stores data that are related to one another in a series of tables. Examples of

relational database software include Microsoft Access, Oracle, SQL Server, and PostgreSQL. The entry point for someone starting out with relational databases is typically Microsoft Access.

Relational databases can ensure that users enter data accurately, completely, and consistently. By storing independent tables of information for each data type, and connecting those tables through ID values and relationships, all the information becomes integrated and accessible to answer questions about the data. This structure also allows them to store data alongside other related data or metadata elements in the smallest overall size possible.

Enterprise-Level Data Management Systems

Enterprise Data Management Systems are custom data management systems designed to meet multiple data management needs of their users. These types of systems typically combine a back-end relational SQL database with a front-end graphical user interface that allows users to enter, query, and manipulate the data. Users may build these systems entirely in-house, through contracted development, or purchase them through a subscription service. These systems often allow monitoring staff to enter data without having to enter all requisite metadata or manage it. Many of these systems deliver queries, graphical views, or facilitate WQX exchanges of data directly.

Each type of electronic data management system has advantages and disadvantages based on the complexity of data collected and the intended use (Table 4). As the Tribe collects more diverse types of data, they may consider expanding their data management program to incorporate more complex systems.

Table 4. Differences in the various data management techniques from more basic to more advanced

	Spreadsheets	Relational Database lite	Enterprise Database
Data entry	✓	✓ (Can add forms)	Requires forms
Stores data and relationships	Some (Power Pivot, vlookup)	✓	✓
Run queries	Limited	✓	✓
Manual data fixes	✓	✓	Backend Access (Developer) is required
Expertise required	Low	Medium	Low (Enter data) High (Manage data)
Costs required	Low	Low-Medium	Medium-High
Ensures data integrity	Some	✓	✓
Publish to WQX	WQX Web	WQX Web/Exchange Network	Exchange Network or API

Data Management Conclusions

Data management is an often overlooked yet significant portion of a monitoring program. When properly planned and implemented, a data management program can become a powerful tool that enables success for the goals of a monitoring program. When not planned and implemented ahead of time, data management can become a barrier to reaching those goals. Considering the goals of the Tribe's monitoring program and identifying each step in the path that data must take is important to generating information that is reported, used, shared, or acted upon. It is important for programs to identify vital functions, users of the data, and steps to fully document and integrate the data to inform the ideal requirements of the Tribe's data management system. Creating standardized systems and practices in the form of a DMP can help ensure users can navigate the path without surprises or unforeseen consequences.

For more information on data management options that will facilitate a WQX submission, or for any questions regarding WQX, please contact the WQX Helpdesk at [wxq@epa.gov](mailto:wqx@epa.gov), the regional Tribal Coordinator, or the Data Management Coordinator.

Data Submissions to WQX

At any point in this process, contact the WQX support line to leave a message at 1-800-424-9067 or e-mail wqx@epa.gov if there are any questions about how to organize the data, what are the allowable values, or other topics.

WQX is the mechanism for tribal environmental organizations, state environmental agencies, and other organizations to submit water quality, biological, and physical data to EPA. One of the three programmatic reporting requirements for the Section 106 Program is that water quality data are submitted to WQX annually. Uploading monitoring data to WQX not only benefits EPA, it also benefits Tribes, other agencies, and the public. Once data are in WQX, users can retrieve them from the WQP for use in a variety of studies and analyses. For instance, the Tribe can acquire their own data and data from surrounding jurisdictions to identify upstream issues that might be of concern. Data published to the WQP is also now made available for viewing in How's My Waterway, alongside any assessment decisions that have been uploaded to ATTAINS about the waterway. The WQP can also be useful in the event of an internal data loss. See Table 5 for a list and description of the different EPA programs and how they work together to collect, document, and provide water quality data.

Table 5. A list of EPA programs and products that support the exchange of water quality data

EPA Product	Explanation
WQX	Data Standard WQX is the mechanism for data partners to submit water quality monitoring data to EPA.
Exchange Network	Data Flow The Exchange Network is the virtual platform where Tribes, states, territories, and EPA collaborate on technological innovations, solutions, and services for sharing data.
Central Data Exchange (CDX)	Data Reporting Site CDX is EPA's electronic reporting site and has been defined as a central point for receiving legally acceptable data.
The Water Quality Portal	Data Retrieval The WQP is the mechanism for anyone, including the public, to retrieve water monitoring data submitted to WQX.
How's My Waterway	Data Integrator How's My Waterway was designed to provide the public with information about the condition of their local waters based on data

	that tribal, state, federal, and local agencies and others have provided to EPA.
--	--

In addition to the Exchange Network listed in Table 5, Tribes may contact the Tribal Environmental Exchange Network for tribal specific data management support and mentorship programs (<http://www.tribalexchangenetwork.org/>).

Essential Elements and Design of WQX

WQX is a standards-based data format for exchanging water quality data. It defines a standard set of allowable fields (or elements) that can capture and describe water data and metadata. The WQX standard also includes a standard set of domain values (allowable values) for a subset of those fields that help to harmonize things like parameter names (for example, to harmonize the parameter names for dissolved oxygen, which can include “DO” and “oxygen, dissolved”). Lastly, a set of business rules ensures that the data are complete, and that data of certain types have the associated metadata fields that they require to be accurately described. See Table 6 for a summary of WQX essential elements.

Table 6. Essential elements of the WQX data standard

Description	Source
<u>Glossary of Terms and WQX Tips</u>	This glossary of terms describes the name and definition of important WQX elements and terms. This page also includes many useful WQX Tips.
<u>Business Rules</u>	The Business Rules provide the validation checks that the system will perform against incoming data. The rules specify permissible attribute configurations and general relationships on an element, ensuring that data meet the standards before the user can save the record.
<u>WQX DET (Data Exchange Template)</u>	Provides a directory of allowable fields provided by the WQX schema, element definitions, and indicates if fields are required, conditional, or optional, and identifies those with associated domain lists.
<u>WQX Domain Value Services and Downloads</u>	Many data elements in WQX have specific lists of allowable values associated with them. Domain values are provided as URL-based services in several formats, including comma-separated values (CSV), eXtensible Markup Language (XML), and zipped files.
<u>Flow Configuration Document</u>	This document contains data validation rules, services, and schema. The principal document captures the detailed data exchange processing rules that govern WQX using narrative text, diagrams, and examples.

Data submitters may first want to determine which water quality data they have to submit. For example, data submitters document water quality data collected by a handheld meter differently than the data from a laboratory analyzed sample and shared via report. For a field measurement, WQX needs to know the equipment deployed and the sensor type; and for a laboratory sample, preservation, analytical

method, and analysis date would be important pieces of information to document and share. Similarly, the fields that describe macroinvertebrate data are different than those for metrics, habitat, or fish data. For each data type, WQX requires that users report certain information in a specific way. These data types answer the questions defined in the Recording and Managing Metadata section. Table 7 through Table 9 show some of the information needed to submit data to WQX and where to get the information.

WQX arranges data in a relational format that is hierarchical (Figure 3). Understanding this hierarchy will help one understand how to properly submit the data. All data submitted to WQX is owned and organized under an Organizational Account which can be requested through the WQX Helpdesk. Only registered users associated with the Organizational Account may add data in the form of Projects, Locations, and Activities/Results. In the hierarchy, data describing Projects and Locations must come before data describing the Monitoring Activities and Results Data. Users must associate Activities and Results with a properly documented Project and Location before they can be stored. This prevents errors such as duplicates or missed associations in the data.

Figure 3. Graphical representation of the WQX relational data structure

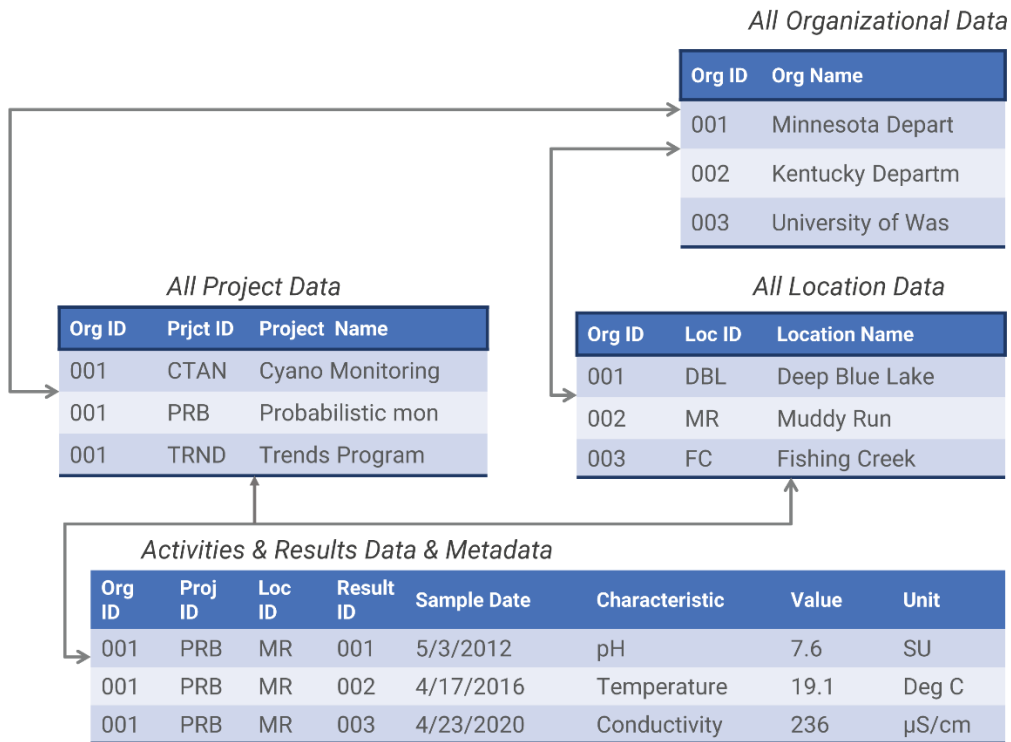


Table 7 through Table 9 come from the WQX physical/chemical Data Template. A Microsoft Excel file is available from the [WQX Web Template Files](#) website. The website contains other worksheets and lists of definitions which can be very helpful to understanding WQX. The template was created as a result of consultation with Tribes. Tribes can use the template to manage data over time, only for occasional uploads to WQX, or as a guide to what WQX is looking for in the data.

- *Field Name* is the data parameter that can be uploaded to WQX. Tribes will likely have similar

names in their own datasets.

- *Description* explains what the field name means and other helpful hints.
- *Source* provides information on where to get the data from.

Table 7. Project information in WQX

Field Name	Description	Source
Project ID	A unique ID for the Project to which data will be assigned.	User created.
Project Name	A unique name for the project.	User created.
Project Description	Project description, which may include a description of the project purpose, summary of the objectives, or brief summary of the results of the project.	User created.
QAPP Approved Indicator	Indicates whether the QAPP for this project has been approved by EPA. *All data collected by a Tribe with CWA Section 106 funds must have an EPA approved QAPP.	Either “Yes” or “No.”
Project Attachment File Name	The file name of any attached files providing additional information about the project.	User created.
Project Attachment Type	The file type of the attached file, such as “pdf.”	From allowed values list.

Table 8. Station information in WQX

Field Name	Description	Source
Monitoring Location ID	An ID used to describe the unique name, number, or code assigned to identify the monitoring location.	User created.
Monitoring Location Name	A name specified by the sampling organization for the site at which sampling or other activities are conducted.	User created.
Monitoring Location Type	The descriptive name for a type of monitoring location.	From allowed values list.
Tribal Land Indicator (Yes/No)	An indicator denoting whether the location is on a tribal land.	Either “Yes” or “No.”
Tribal Land Name	The name of an American Indian or Alaska Native area where the location exists.	User created.

Field Name	Description	Source
Monitoring Location Latitude (DD.DDDD)	The latitudinal coordinate measure of the given monitoring location in decimal degree.	User created.
Monitoring Location Longitude (-DDD.DDDD)	The longitudinal coordinate measure of the given monitoring location in decimal degree.	User created.
Monitoring Location Source Map Scale	The scale of the map used to identify the coordinates of the location (if a map was used).	User created.
Monitoring Location Horizontal Collection Method	The method used to determine the latitude and longitude coordinates for a point on the earth.	From allowed values list.
Monitoring Location Horizontal Coordinate Reference System	The name that describes the reference datum used in determining latitude and longitude coordinates.	From allowed values list.
State Code	A code designator used to identify a principal administrative subdivision of the United States, Canada, or Mexico.	From allowed values list.
Monitoring Location County Name	A code designator used to identify a U.S. County or county equivalent.	From allowed values list.

Table 9. Results information in WQX

Field Name	Description	Source
Activity ID	A unique identifier for each activity (for example, a unique laboratory ID, a suite of on-site field measurements, or an individual sample).	User created.
Activity Media Name	Medium in which this activity occurs.	Listed in template allowable values.
Activity Type	Type of activity.	Listed in template allowable values.
Activity Start Date	Date activity began (MM/DD/YYYY).	Date samples or measurements taken.
Activity Start Time	Time activity began (HH:MM).	Time samples or measurements taken.
Activity Start Time Zone	Time zone of the Activity Start Time.	Three letter time zone code.

Field Name	Description	Source
Sample Collection Method ID	Valid WQX sample collection method, required if activity type is a sample.	User-created based on QAPP.
Sample Collection Equipment Name	Equipment used to collect the sample.	Listed in template allowable values.
Characteristic Name	Valid WQX characteristic name.	Listed in template allowable values.
Method Speciation	Identifies the chemical speciation in which the measured result is expressed.	Listed in template allowable values.
Result Value	Measured value of characteristic.	From laboratory results or field sheet.
Result Value Units	Units for characteristic result.	From laboratory results or field sheet.
Result Sample Fraction	Fraction for characteristic, required for certain characteristics (see Non-Taxa Characteristics for details).	Laboratory results.
Result Status ID	Indicates acceptability of the result with respect to Quality Assurance/Quality Control (QA/QC) criteria.	Listed in template allowable values.
Result Analytical Method ID	The identification number or code assigned by the method publisher.	Listed in template allowable values or created by organization staff based on QAPP.
Result Analytical Method Context	Identifies the source or data system that created or defined the identifier.	User created.

To manage and report data in a WQX format, Data Managers will need to provide at least the data fields listed in Table 7 through Table 9. The WQX Physical/Chemical Results Template provides detailed guidance on the requirements for each field that may need to be provided. For example, make sure that:

- Field length of data in a data field is no longer than what WQX will accept.
- Provide information in the defined formats or select the appropriate allowable value from the WQX domain lists, as defined in the template.
- Provide the unique identifier that WQX requires for all Projects, Stations, and Activities within the organization.

Metadata are also very important for telling potential data users how data were gathered and why certain data may be unreliable. For this reason, it is encouraged that all relevant QC data are entered

into WQX, as well as specific sample collection, field, and laboratory procedures.

Data standards are consistent ways to report data in formats that are easy to share. When reporting results to WQX, the Data Manager will need to use the appropriate allowable value for characteristic names. WQX characteristic names are based on an EPA data standard. Note, there may be multiple WQX characteristic names available for parameters depending on the form of the sample. For example, the Data Manager can select from dissolved oxygen (DO), dissolved oxygen saturation, or dissolved oxygen uptake within WQX to best fit the sample collected. Table 10 shows the list of parameters most reported within WQX and their corresponding WQX characteristics. Allowable WQX values for parameters that are not listed in Table 10 can be found using the WQX Domain Service. Please review all available WQX characteristic names to best suit the sample collected.

Table 10. The 15 most reported parameters from Tribes in WQX and their allowable values

Commonly Reported Parameter	Corresponding WQX Characteristics
Water temperature	Temperature, water
Dissolved oxygen	Oxygen, dissolved
pH	pH
Turbidity	Turbidity
Specific conductance/conductivity	Specific conductance
Salinity	Salinity
Total suspended solids	Total Suspended Solids
Temperature	Temperature, air
Fecal coliform	Fecal Coliform
Flow	Flow (other forms exist as well)
E. coli	Escherichia coli
Phosphorous	Total Phosphorus, Mixed Forms*
Total nitrogen	Nitrogen
Benthic macroinvertebrates	Count (also metrics)
Fish assemblage	Count (also metrics, fish tissue)

*Total phosphorous is calculated as the sum of both orthophosphate and phosphorous, depending on the methodology, Tribe’s may need to include both characteristics to accurately describe total phosphorus. Alternatively, the Tribe can use the characteristic of “Phosphorus” combined with the sample fraction of “Total” to describe total phosphorus. The other individual forms such as “Orthophosphate” are also available.

Good QA and QC ensures that the data can be reused by another user for analysis, as well as safeguarding against misuse. The following documents will help improve the quality of the data and help meet WQX submission requirements for these data types

- [WQX Nutrients Best Practices Guide](#).
- [WQX Metals Best Practices Guide](#).
- [Best Practices for Sharing Benthics Data](#).

There are also standard ways to report the procedures used when collecting data in the field and the laboratory used to evaluate samples. For example, when a Tribe reports a result to WQX the Tribe will need to select an Analytical Method Context, such as the National Institute for Occupational Safety and Health or EPA. The Tribe will also select the Analytical Method ID, such as 150.1 for pH or 150.2 for pH by continuous monitoring. All these allowable values are in the WQX templates. If the Tribe uses unique procedures, they can ask they be added to WQX by contacting the WQX Helpdesk. One of the first activities the Data Manager will do when they plan to submit data to WQX is to identify the sources and IDs for the field and laboratory procedures used for each characteristic. Tribes can work with their laboratory to find out procedures for analyzing certain parameters. Tribal QAPPs, SOPs, and even data logger manuals are also good resources for procedural IDs.

Overall, the following steps will help get data into WQX. At any point in this process, contact the WQX support line to leave a message at 1-800-424-9067 or e-mail wqx@epa.gov if there are any questions about how to organize the data, what are the allowable values, or other topics. The WQX templates on EPA's [WQX Web Template Files](#) website show specifics such as allowable and required values for the following steps.

1. Ensure that all required data fields listed in Table 7 through Table 9 are collected, as shown in the template on EPA's website.
2. Determine what values WQX will allow for the selected parameters and field and laboratory procedures. Refer to the WQX templates for help.
3. Format the data according to WQX's specific required fields, field length, and field type. Tribes may want to use the WQX templates, which lay out these items.
4. Before submitting data to WQX, Tribes need to request an Organization ID by e-mailing wqx@epa.gov or leaving a message with the WQX Helpdesk at 1-800-424-9067.
5. The General Data Management Best Practices section reviews how to put data into a format to be uploaded to WQX. Once data are in this format, the next step is to submit the data electronically to WQX using the different tools and mechanisms available.

EPA requires all data collected using Section 106 funds to be uploaded to WQX annually. In some circumstances, the EPA regional contact or Project Officer may waive this requirement. However, the data will still need to be submitted to the Regional Office or Project Officer in a WQX compatible format, along with a justification (see the [CWA Section 106 Tribal Guidance](#), Appendix C: Sample Justification for WQX). A WQX Web template can be used as a starting point for data, even if the data are not submitted to WQX or when an Excel spreadsheet with the required metadata are not provided. If there are issues with getting the data into WQX, please contact the WQX Helpdesk (wqx@epa.gov).

As described in the Appendix C: Sample Justification for WQX in the [CWA Section 106 Tribal Guidance](#), the justification for waiving WQX submission should include: a description of the limitation or impediment to water quality data entry directly to WQX (see text box for examples), a proposed plan to address the issue, and a proposed timeframe to meet the programmatic reporting requirement. Information in the justification will assist EPA in addressing barriers to data submission to WQX, where possible. Tribes should send WQX compatible data and justification details to their Project Officer in accordance with the annual milestone due date in their yearly or multi-year grant work plan.

Justification examples:

- Inadequate internet access
- Funding limitations
- Loss of key staff
- Training needed
- Other

Ways to Submit Data

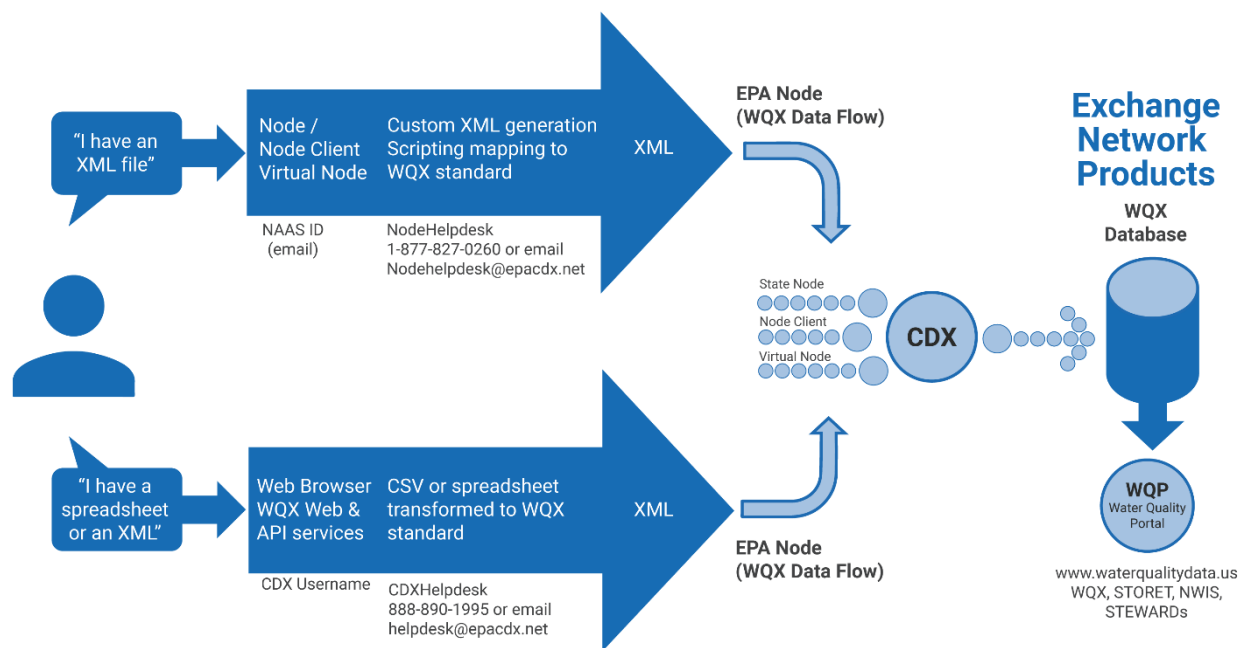
To upload data into WQX, determine what mechanism is best. Each approach has advantages and disadvantages, but all are acceptable means for meeting the data submission requirements set forth in the [CWA Section 106 Tribal Guidance](#). The list below includes options (described in more detail in Figure 4):

1. Submit data through the WQX Web Submission Tool (WQX Web).
2. Submit data to WQX using an Exchange Network node or node client.
3. Third-party WQX-compliant data management software for purchase.

1. Submit Data through WQX Web

WQX Web is a web-based user interface for uploading text and spreadsheet files of Project, Station, and Results data to WQX. Standardized templates, as well as tutorials, are available to assist Tribes in formatting data correctly for data submission. Users can submit data using the format in the templates or custom import configurations can be created to submit data in other formats. The tool also allows Tribes to correct data errors upon submission. Examples of Common Errors and Resolutions is found in the Glossary and WQX Tips Page.

Figure 4. Visual example of data flow from data practitioner to the WQP



- API:** Application Programming Interface
- CDX:** Central Data Exchange
- CSV:** Comma-Separated Values
- NAAS ID:** Network Authentication and Authorization Services Identity
- NWIS:** National Water Information System
- STEWARDS:** Sustaining the Earth’s Watersheds Agricultural Research Data System
- STORET:** STORage and RETrieval Data Warehouse
- XML:** eXtensible Markup Language
- WQP:** Water Quality Portal
- WQX:** Water Quality Exchange

Once data are submitted to WQX, the data are available to the Tribe, EPA, and the public within one week’s time. As a part of the Section 106 Program’s reporting requirements, the data submitted to EPA will enable water quality managers to have a better picture of how water quality is improving in Indian Country.

Errors may be found after having submitted data to WQX. All the tools for submitting data described above allow for changing any submitted data. For example, if a set of results taken during a sampling event or season was incorrect because the equipment used to collect it was found to be defective, then the submission tools can be used to delete or update the erroneous data. Resources available to help with this submission method are available in Table 11 below.

2. Submit Data to WQX Using an Exchange Network Node or Node Client

WQX uses EPA’s [Exchange Network](#) to flow water quality monitoring data to WQX from a Tribe’s data system via Exchange Network nodes or node client. Tribes that have been awarded Exchange Network grants use this method to submit to WQX. Most Tribes submitting data through WQX directly have database systems for managing data. Some Tribes have begun working together toward utilizing this method for submitting data. If considering this route, Tribes should stay aware of activities going on in

their region through the Exchange Network grant representative, Section 106 grant representative, and regional monitoring coordinator. More information about WQX can be found on EPA's [Water Quality Data](#) website.

EPA's Exchange Network is a way to exchange environmental information securely over the internet, using an approach based on standard technology and protocols.

Exchange Network Grants are awarded to federally recognized Tribes, states, and territories to assist in the development and implementation of Exchange Network nodes, exchanges, tools, and related backend systems.

3. Third-Party WQX-Compliant Data Management Software Services

Since many water quality data producers need or want to report data through WQX, several providers of enterprise data management systems have added WQX compatibility to their software systems. Enterprise level data management systems are customized data management platforms that often provide functionality for several data management functions, like entry, data integration, analysis, import, and export, in a variety of formats. These systems may either prepare datasets for WQX Web submission, submit data through the WQX Web Application Programming Interface (API), or publish to the CDX directly through an Exchange Network. These systems are typically purchased through licenses and may still require a certain amount of technical expertise to implement successfully.

Retrieving Data from the Water Quality Portal

Tribes may wish to perform analyses on data loaded to WQX by their programs or by other relevant data producers. Data uploaded to WQX may be downloaded via the [WQP](#) website. The WQP offers several query options. To retrieve all data uploaded, Tribes may search by the provided WQX Organization ID. Note – It's often best to keep queries simple as any conflicting parameters will limit or prevent the data from being returned. Once the query options have been selected, select a profile to download, such as site data or physical/chemical results. Once a profile has been selected, click "Download" and a box will pop up giving a count of records in the query. If this number is correct, click "Continue." Detailed instructions for downloading data from the WQP are available from the [WQP User Guide](#). A video tutorial on [Searching for Data in the Water Quality Portal](#) is also available on the EPA YouTube channel and WQX website.

A second option for retrieving data from the WQP is through [WQP Web Services](#). Web Services offer a way to allow computers to retrieve data directly from the WQP in predefined formats using a computer-to-computer protocol for directly sharing information.

WQX Support and Resources

If there are questions about getting data into or out of WQX, there are ways to seek help as shown in Table 11.

The WQX helpdesk is a useful resource to assist with any issues that arise with WQX. Users may leave a message with the WQX support line 1-800-424-9067 or e-mail [wxq@epa.gov](mailto:wqx@epa.gov). EPA's WQX helpdesk can set up Organization IDs and enter field and laboratory procedures into the system.

A plan to submit data and frequency of submittal is helpful in managing staff schedules. Additionally, if submittals occur frequently, it may help ensure familiarity with the system. Once Tribes have decided on a plan for submitting data, it can be noted in workplans and the Monitoring Strategy and revised as necessary.

Table 11. WQX resources for Tribes

Topic Area	Resource Links
Open Water Data Resources	Water Quality Data Homepage [Data In] The Water Quality Portal [Data Out] How's My Waterway [Info Out] Central Data Exchange [Data Held] Exchange Network (Node Submissions)
WQX Nuts and Bolts	WQX Web - Resources for a Standard Upload Glossary of Terms Business Rules WQX DET (List of Elements) WQX Domain Value Services and Downloads (Acceptable Values)
Using WQX Web – Getting Started	Introduction to WQX (5-minute video) WQX Web Template User Guide (PDF Guide) User Guide Version 3.0 for WQX Web Getting Started with WQX Web: How to Gain Access WQX Web Basics – Two-part session – Day 1 WQX Web Basics – Two-part session – Day 2
Using WQX Web – Templates	WQX Web Template User Guide WQX Web Template Files Web Templates Overview Video
Using WQX Web – Import Configurations	Import Configurations Overview Video Import Configuration Options Video
Using WQX – Best Practices	Best Practices for Sharing Benthics Data WQX Metals Best Practices Guide

Topic Area	Resource Links
	<u>WQX Nutrients Best Practices Guide</u>
Using WQX - Other	<u>WQX Factsheet</u> <u>WQX/WQP Training Resources Site (Includes 101 Course)*</u> requires Tribes to request permission to access <u>WQX Web Error Resolution</u> <u>Learn More about Water Quality Data</u> <u>Exchange Network Grant Program</u>

Appendix A: Key Terms

Allowable Values – Acceptable data values that can be input for a given data element. These values provide information on the exact spelling and capitalization needed for acceptance into the Water Quality Exchange (WQX).

Assessment Data – Decisions about water quality that are made based on comparisons of water quality monitoring data against water quality standards.

Business Rules – Rules that ensure completeness and consistent structure of data input into WQX.

Central Data Exchange (CDX) – The Agency’s electronic reporting site. The Central Data Exchange is a central point which supplements EPA reporting systems by performing new and existing functions for receiving legally acceptable data in various formats, including consolidated and integrated data.

Characteristic Names – Standard names used to describe measured or analyzed parameters.

Conditional Fields – Data fields in WQX that are required only when certain conditions are met in other data fields. For example, a time zone must be entered if the time of an Activity is entered but is otherwise not required.

Data Elements – A basic unit of data within a data flow or transfer. It is often called a data field in a database.

Data Field – A column in a database that can hold one specific data parameter. For example, the station ID for all stations would be contained within one data field.

Data Manager- Staff person responsible for entering and managing data in a database.

Data Producer- Person who collects or measures data in field and initially reports on

the data.

Data Standards – Consistent ways to report data so that they are easily understood and shared between data users.

Data User – A person who has an interest in using the data at some point (Tribal constituents, EPA, or other).

Exchange Network Node – Software that facilitates the transfer of data between an organization’s database and other databases on EPA’s Exchange Network. Nodes use standards-based web services and eXtensible Markup Language (XML) to allow for the seamless exchange of data between organizations regardless of hardware, operating system, or programming environment.

Extensible Markup Language (XML) – An open standard language that describes data using rigid syntax rules. XML allows data to be transferred between different data systems, regardless of computer system or platform. XML overcomes system incompatibility by translating information into a common data structure and format.

Field Length – The maximum number of characters or numbers that can be placed in a data field.

Field Name – The name of a data field, which provides a common reference for a data field.

Field Type – Term referring to whether a data field can contain freely formatted text data, data of a defined format, or must conform to a list of allowable values.

GIS Data – Data of a spatial nature, such as a map associated with information about the waterbodies, produced using a Geographic Information System.

Metadata – Information about a specific water quality sample that provides context for the sampling activity. Metadata contain information about when, where, why, and how the program collected a water quality sample.

Monitoring Strategy – Serves as the foundation for a monitoring program and addresses how the program will meet water management needs. Please see Chapters 5 and 6 in the [CWA Section 106 Tribal Guidance](#) and the [Developing a Tribal Water Quality Program Monitoring Strategy](#) for more information.

Naming Conventions – A standardized way of creating IDs and names that allow for easy identification and organization of data.

Node Client – Software that can be downloaded onto individuals' computers and can initiate Exchange Network requests for data or a submission of data to other nodes. Unlike nodes, nodeclients cannot respond to requests for data.

Quality Assurance (QA) – An integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is intended and of high quality.

Quality Control (QC) – The overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established; operational techniques and activities that are used to fulfill requirements for quality.

Relational Database – a database that relates data in one table to data in another table by using matching values in key fields.

Required Fields – Data fields in WQX that must be included in a data submission under all circumstances.

Water Quality Exchange – WQX is the water data schema associated with the Exchange Network used by Tribes, states, and EPA. More information about WQX, WQX Web, and the Water Quality Portal, including tutorials, can be found on EPA's [Water Quality Data](#) website.

Water Quality Monitoring Data – Field data or laboratory results associated with sampling events at a monitoring location.

WQX Data Template – Spreadsheet tool used to facilitate the entry of water quality data for eventual upload to WQX.

WQX Web Submission Tool (WQX Web) – The tool that converts data (from spreadsheets and other flat file formats) to the XML format and migrates the data to WQX.



United States Environmental Protection Agency

Office of Water

Office of Wastewater Management

July 2024