

Continuous Monitoring Data Sharing Strategy

Task 6.1

Prepared by

Michael Baker International

3601 Eisenhower Avenue, Alexandria, VA 22304

LimnoTech

501 Avis Drive

Ann Arbor, MI 48108

MapTech, Inc.

3154 State Street

Blacksburg, VA 24060

Submitted to

US EPA

Washington, DC

Under

EPA Contract EP-C-12-052 Order No. 0005

Dated September 22, 2014

September 21, 2015



Michael Baker
INTERNATIONAL



EP-C-12-052 Task Order No. 0005

Task 6: Continuous Monitoring Strategy Document Outline

Background

This document was prepared to support the goal of the U.S. Environmental Protection Agency (EPA) to promote the use of innovative technology in the sharing and use of sensor data as part of the EPA Office of Water's Technology Blueprint.

This comprehensive Continuous Monitoring Data Sharing Strategy integrates input collected through the following key tasks:

1. Research of current and potential approaches for sharing continuous monitoring data (Task 3)
2. Workgroup meetings (Task 4)
3. State, tribal, interstate, EPA Regional and other Federal agency interviews (Task 5)

This document also includes direct input from the designated EPA Task Order Contracting Officer Representative (TOCOR), and from members of the Continuous Monitoring Data Strategy Workgroup.

Table of Contents

| | |
|----------------------------------------------------|----|
| Background | 2 |
| Table of Contents | 3 |
| Executive Summary | 4 |
| Introduction | 6 |
| Defining Continuous Data | 7 |
| The Need | 8 |
| Difficult Access | 8 |
| Exchange for Discrete Data | 8 |
| Standards for Continuous Data | 9 |
| Methodology and Findings | 10 |
| Workgroup | 10 |
| Research | 14 |
| Interviews | 15 |
| Recommendations | 19 |
| Overview | 19 |
| System Management | 22 |
| Data Types and Sources | 23 |
| Data Quality | 23 |
| Data Submittal and Reporting | 24 |
| Management of Related Information Sources | 24 |
| Sharing/Retrieval Capabilities and Tools | 25 |
| Policy Gaps and System Needs | 25 |
| Collaboration and Implementation Strategy | 27 |
| Program Management and Partners | 27 |
| Implementation at the State/Agency Level | 28 |
| Schedule | 29 |
| References | 31 |
| Appendix 1 – Workgroup Organizations | 32 |
| Appendix 2 – Research Sources | 33 |
| Appendix 3 – Common Questions for Interviews | 35 |

Executive Summary

Water resource managers require access to data to make effective and timely decisions. Discovering and compiling the necessary water data can be significantly burdensome. Within a single watershed, multiple entities collect water data from various resource types for multiple purposes. These entities may include, but are not limited to, local watershed groups, permitted facilities, counties, tribes, state agencies and federal agencies. Example data types used for management decisions may include precipitation, snowpack, stream discharge and stage, reservoir height, groundwater levels and water quality data (both discretely sampled data and data derived from a sensor). To make use of these data, analysts must know where to find the data, go to the website or individual who owns the data, and retrieve the data. Because data are often stored without uniform data standards and terminology, resources must be invested in re-formatting data and ensuring data quality. This may be a very time consuming task, and the process is co-limited by the accessibility of the needed data and the analyst's prior knowledge of available data sets.

In 2007, the U.S. Environmental Protection Agency (EPA) released the Water Quality Exchange (WQX), which became the de-facto data standard for communicating discretely sampled water quality data. WQX has proved to be very successful in enabling the exchange of water quality data between various data providers. It has also led to further integration between federal agencies by allowing the U.S. Geological Survey (USGS), the U.S. Department of Agriculture (USDA) Agriculture Research Service and EPA to share water quality sampling data through the Water Quality Portal (WQP). However, even though WQX has been successful with discretely sampled data, it was not designed to manage continuous data generated by water quality sensors. The use of water quality sensors is becoming much more common, and the development of new sensor technology for nutrients and other parameters of interest is expanding and diversifying their use.

In 2013, the Open Geospatial Consortium (an international data standards development community) released a new data standard designed specifically for continuous monitoring data. This new standard, WaterML 2, may provide a basis for developing a sensor data sharing network that allows all data collected by sensors (both water quantity and water quality) to be shared using a common format. This network would facilitate the discovery and dissemination of water data.

EPA funded this project to define a practical strategy for integrating continuous monitoring data into a sensor data sharing network. This project includes research and review of available standards (including WaterML 2), technologies and activities relevant to management of continuous monitoring data; interviews with organizations already working with such data; and ongoing interactions with a stakeholder group representing potential contributors and users of this data.

The outcome of the project is this strategy document outlining steps for implementation of a continuous monitoring data-sharing network by EPA. The document summarizes the research, interview and discussion processes undertaken to provide a basis for development of the strategy, and presents an initial technical recommendation along with proposed steps towards implementation of a network built around this recommendation.

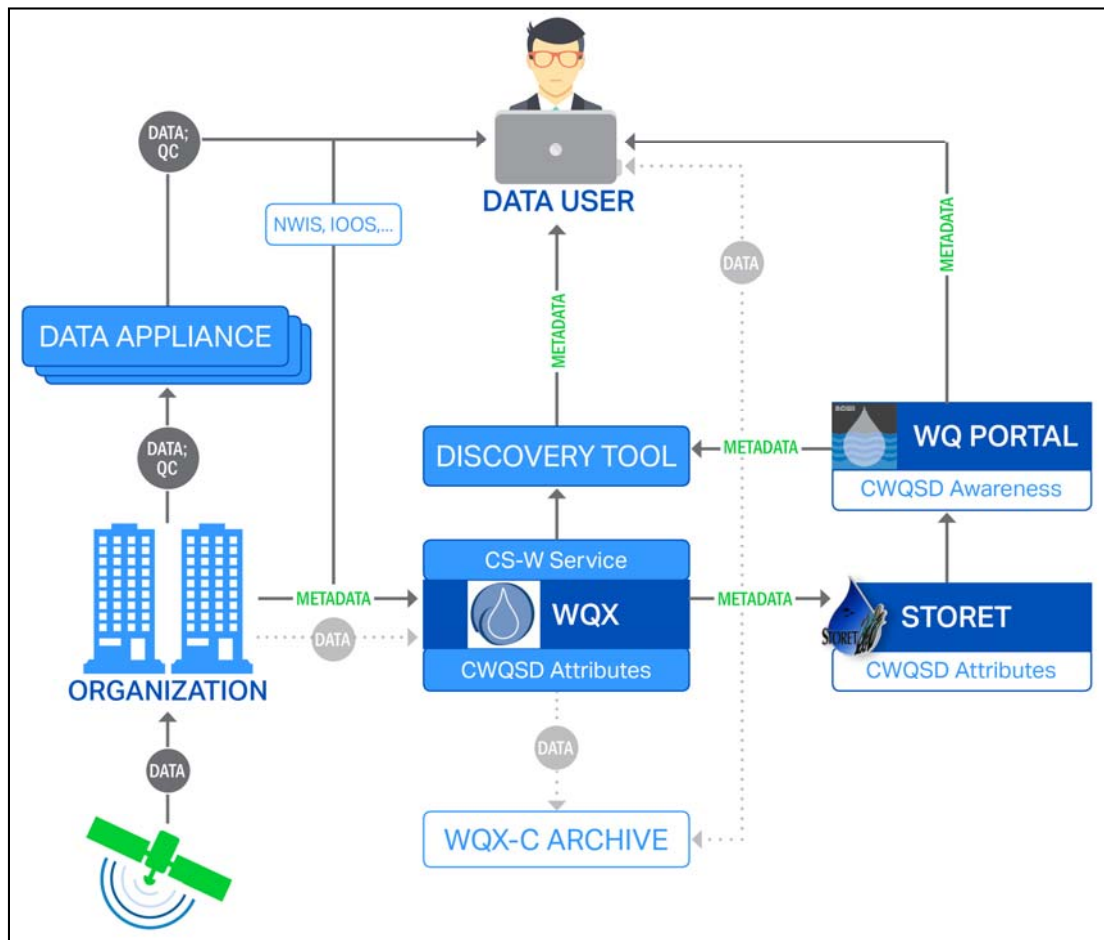
Key findings, the recommendation and implementation strategy are summarized below. The mention of trade names or commercial products and/or vendors does not constitute endorsement or recommendation for use by EPA.

Key Findings

1. Research, interviews and workgroup feedback provided the basis for a number of key findings:
2. Definitions of continuous data vary, but in-situ measurements collected at regular intervals from autonomous sensors form a reasonable definition for the purposes of this strategy.
3. There are both operational and planning/assessment uses for continuous data that call for potentially different management for that data.
4. The existing WQX design framework is not designed for efficient ingestion, management and dissemination of continuous data.
5. There are effective centralized frameworks, such as the USGS National Water Information System (NWIS) and effective distributed networks, such as the Integrated Ocean Observing System (IOOS), which both take advantage of a standards-based approach to implementation.

Recommendation

The recommended system architecture calls for the standards-based implementation of a centralized catalog, data appliances and archive as shown below, where (1) Site and deployment metadata are submitted to an extended WQX serving as a catalog; (2) Data are available through service endpoints exposed by data appliances tied to the organization; (3) Data users discover data of interest with a discovery tool by querying metadata in the catalog, and then retrieve the data from the data appliances; and (4) Data are archived for backup, redundancy and/or regulatory reasons in a modified WQX.



Implementation Strategy

EPA will work with data owners and with key partners such as USGS, NOAA and vendors to define, design and develop a technical implementation of the recommended architecture that provides for scalability and accessibility. Three periods are envisioned for implementation:

- **2015-2016: Pilot and Refine.** EPA will identify a suitable watershed (or watersheds) on which to exercise and test the recommended technical approach on a pilot basis. During this pilot phase, EPA and its contractors will work on detailed specification, development and deployment of the catalog, data appliances, archive and prototype analytical tools. EPA should also engage with the vendors of sensors, data loggers and sensor data management tools, and with peer organizations like USGS and NOAA to promote interoperability. Lessons learned during this period will be gathered and reviewed to identify potential adaptive changes to the strategy, and reports made to the stakeholder community through presentations at national conferences, such as AWRA or the National Monitoring Conference.
- **2017-2018: Build Community and Framework.** Following completion of pilot projects and refinement of the data strategy, the focus will be on building tested, scalable infrastructure for an increasing community of organizations sharing continuous data by building and testing production versions of the centralized elements (catalog and archive), further refining data appliance configurations, implementing additional tools, and supporting early adopters through documentation and guidance.
- **2019 and Beyond – Scale Up and Build Out.** Once the tested centralized infrastructure is available, data owners will be able to stand up their own data services to provide access to their continuous water quality sensor data (CWQSD) while placing the associated site and deployment metadata on the WQX-based catalog.

EPA recognizes that the recommended architecture can be used with many other types of continuous sensor data besides water quality, such as water levels, streamflows, precipitation or air temperature. EPA therefore intends to explore options for ingestion of such data or for interoperability with other enterprises aggregating these data.

Introduction

In 2014, EPA began the development of a strategy to manage and share continuous water quality monitoring data. **The goal of this strategy is to define a road map for how EPA and its partners could develop a national data-sharing network for continuous monitoring data that integrates data from multiple sources and is broader than just water quality data. The network would include water quantity data as well (any water data collected via a sensor). The network would also integrate with the discrete water quality sampling data that is already available via the Water Quality Portal.** This effort follows on the success of EPA's Water Quality Exchange (WQX), which enable the sharing of discrete sampling data, and the Water Quality Portal, which provides access to water quality sampling data collected by the USGS, the USDA Agricultural Research Service, EPA, and the 400+ partners that share data with EPA via WQX. EPA embarked on this effort by coordinating with the National Water Quality Monitoring Council (NWQMC) Sensor Workgroup to form a workgroup that would evaluate various options for sharing continuous data. In addition, EPA conducted five in-person interviews with practitioners and data managers to evaluate current approaches and technical needs. Lastly, EPA conducted exhaustive research on current approaches for sharing continuous data, including the evaluation of existing international data standards.

Through this research and discussion, EPA identified several existing approaches being used to manage continuous data for individual organizations, as well as examples for sharing particular types of continuous data with others. Some of these made use of international standards, while others used organization-specific approaches. Through the research, EPA identified existing components along with best practices that could be used in developing a national data-sharing network for continuous data. In addition to the recommended approach for the data sharing network, this document provides an overview of the research that was conducted and of the options that were considered.

Defining Continuous Data

It was important to the development effort to provide a working definition of continuous data. The research team worked with EPA and stakeholders throughout to gather perspectives and develop a consensus description of what makes a data collection effort "continuous":

- Continuous data is collected on a regular basis from autonomous sensors;
- Continuous data sampling frequency may be as low as monthly or daily, but typically is on the order of hourly to every few minutes; and
- Continuous data measurements are made in-situ, as contrasted with discrete samples collected in the field manually or with an autosampler and delivered to a laboratory for analysis.

Other characteristics associated with continuous data include low cost and effort per measurement, serial correlation between measurements, and less voluminous metadata requirements as compared to discrete data.

In establishing the working definition, EPA recognized that continuous data may be collected from sensors mounted on mobile platforms such as underwater gliders. However, the recommendation presented here is based on consideration of fixed sensors only.

The Need

Environmental data collection and management in the public domain has matured to the point that a bewildering variety of data from across North America and its oceans is immediately available to anyone with internet access. Data is collected by manual or autonomous sensors and uploaded to internet servers where there are one or more quality control steps. From these internet servers or from large, centralized servers, data is readily available from cached or archived storage. Tools and front-end interfaces exist to facilitate the temporal and spatial display of the data by users.

While the maturation of this ad-hoc system continues daily, there are persistent deficiencies in the water resource area that inhibit the collection and presentation of data. Specifically, the majority of water data are instantaneous measurements with associated metadata; discrete samples. Consequently, the data management systems for this data have been designed around discrete sampling and improved to the extent that continuous sampling data either cannot be accommodated, or can only be accommodated inefficiently.

Water resource managers require data to make effective and timely decisions. Discovering and compiling the necessary data can be significantly burdensome. Within the same watershed, multiple entities collect water data from various resource types for multiple purposes. These entities may include local watershed groups, permitted facilities, counties, tribes, state agencies and federal agencies. They may also include departments within an entity. Example **water quantity** data used for management decisions include precipitation, snowpack, stream discharge and stage, reservoir height and groundwater level. Example **water quality** data include measures of temperature, dissolved oxygen, nutrients, metals, organics, conductivity, pH and pesticides. These data exist as both discretely sampled data, and data derived continuously from a sensor.

Difficult Access

To make use of these data, water resource managers must decide what data are needed, determine where to search for the data, locate the website or individual who owns the data, and retrieve the data. This is often a time-consuming task. In addition, because data are often stored without uniform data standards and terminology, more resources are expended in reformatting data and ensuring data quality. The process is co-limited by the accessibility of the needed data as well as the manager's prior knowledge of available data sets. With limited resources available to conduct water quality monitoring, no one organization can reasonably be expected to monitor all waters. Sharing the cost and effort of monitoring across multiple organizations provides significant benefit, but in order for this benefit to be fully realized, common data sharing approaches must be taken. Otherwise the burden of discovering, reformatting, and making use of others' data becomes too great and outweighs the value of using that data.

Exchange for Discrete Data

In 2007, EPA released the Water Quality Exchange (WQX) (EPA, 2015), which became the de facto standard for communicating discretely sampled water quality data. WQX has proved to be very successful in enabling the exchange of water quality data between various data generators and providers. It has also led to data integration between federal agencies by allowing the U.S. Geological Survey, the USDA Agriculture Research Service, and EPA to share water quality sampling data through the Water Quality Portal (USGS et al., 2015). However, although WQX has been successful with discretely sampled data, it was not designed for high-frequency continuous data generated by water quality sensors.

Approaches for Sharing Continuous Monitoring Data

The use of water quality sensors is becoming much more common, and the development of new sensor technology for nutrients and other parameters of interest is expanding and diversifying their use. This is leading to a proliferation of continuous data.

Standards for Continuous Data

In 2013, the Open Geospatial Consortium (OGC) released a new data standard that is designed specifically for continuous monitoring data. This new standard, WaterML 2 (OGC, 2015), provides one possible basis for developing a sensor data sharing network that allows all data collected by both water quantity and water quality sensors to be shared using a common format. This network would thereby facilitate the discovery and dissemination of water data. The OGC, an international consortium of companies, agencies and universities developing publicly available interface standards that improve data exchange and analysis on the Internet, has also promulgated other standards that are relevant to this effort, such as Sensor Observation Service (SOS) and Catalog Service for Web (CS-W). These services together provide the basis for existing networks such as the Integrated Ocean Observing System (IOOS) and the Consortium of Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI) HydroServer, and are allowing commercial vendors to readily feed organizations' sensor data into such networks.

Methodology and Findings

Activities undertaken for this project included research and review of available standards (including WaterML 2), technologies, and activities relevant to management of continuous monitoring data, interviews with organizations already working with such data (e.g., agencies, vendors, and academic consortiums), and ongoing interactions with a stakeholder working group representing potential data contributors and data users.

Workgroup

In order to gain stakeholder input on requirements and feedback on proposed approaches, EPA invited federal, state and tribal staff along with representatives of academic institutions and vendors from across the nation to participate in a Continuous Monitoring Data Strategy Workgroup organized under the NWQMC Sensor Workgroup. The 70+ workgroup members were given access to written summaries of research and interview activities, and participated in nine teleconferences during which they were first briefed on project activities and then given the opportunity to discuss and comment on all matters. A project collaboration site was established for use by the team and workgroup members to share project materials and information. Represented organizations are listed in Appendix 1 – Workgroup Representation.

Conference calls with the workgroup and related comments submitted by e-mail provided useful perspectives on real-world workflows and applications for continuous data, such as:

Definitions vary. Stakeholders and interviewees defined continuous data in a number of ways. The most common first response was by sampling frequency. Sampling frequency generally was thought to be continuous if samples were taken at least every fifteen minutes. However, some defined continuous data with much lower frequencies (e.g. daily or monthly) if the sampling site was monitored over a long period of time (e.g. years). For example, the Texas Water Development Board (TWDB) monitors groundwater wells once a month and has a long history of data for each of their wells. From their perspective, this data is continuous. Spatially, data was considered continuous if the sampling objective was to describe the spatial variability of the monitored parameter. The New Jersey Department of Environmental Protection (NJDEP) deploys gliders to describe, spatially, the DO variability in their coastal waters. The participants also recognized the difference in the amount of metadata between continuous and discrete sampling. Discrete sampling is considered to have a great deal more metadata associated with a single sample. Discrete samples are generally collected in the field and delivered to a laboratory for analysis. Other qualifiers to continuous data definitions included the “effort” required in collecting data, low cost per sample, “serial correlation” between samples, data from sensors and autonomous in-situ analysis.

When considering the wide-ranging thoughts, the one that captured the general consensus is the concept autonomous in-situ analysis. Data collected in this fashion require sensors to measure the parameter of interest. The ratio of metadata to sample data is low as compared to discrete data. Sensors available on the market today generally will allow for a wide variety of sampling frequencies.

Uses Vary. Stakeholders and interviewees identified a wide range of use cases that could be characterized as operational (where immediate action may be dictated by observed conditions) or assessment (where no immediate action is dependent on measurements, though future actions may be informed by the measurements). For example, the Texas Commission on Environmental Quality (TCEQ) measurements of turbidity, pH and conductivity in the Rio Grande are used operationally to inform irrigation decisions (certain water quality conditions can lead to unaesthetic water spots on leafy greens), while NJDEP’s optical measurement of nearshore chlorophyll from the air is used to direct grab sampling protective of shellfish

Approaches for Sharing Continuous Monitoring Data

harvesting operations. TCEQ water quality observations are also used to monitor and characterize long-term trends in watershed health and restoration, and NJDEP uses continuous water quality data to support water quality assessments and the listing of impaired waters.

Figure 1 presents a categorization of use cases based on latency (time elapsed between measurement and use) and level of quality control. A high level of quality control, perhaps encompassing manual review both of the continuous sensor data and associated pre-, intra- and post-deployment discrete samples, is deemed necessary for use of continuous data in a regulatory or assessment context where comparison to water quality standards might come into play. In an operational context, a somewhat lower level of quality control – likely automated comparisons checking that measurements lie in a valid range, are not “stuck”

or flatlined, and similar – may be acceptable when measurements are compared to a control value that triggers a (non-regulatory) action. Operational applications have a short data latency requirement, as the operational actions are generally required to take place as soon as possible, while assessment applications do not have the same short-term data requirement and may be undertaken weeks, months or even years after the sensor measurements are made. Two other combinations are possible in this categorization scheme – low latency with higher QC, and high latency with lower QC. However, the former is unrealistic, as higher QC cannot be achieved in combination with the low latency requirement, and the value of the latter is uncertain, as operational applications will not be feasible because of the high latency, and assessment applications will not be feasible because of inadequate QC.

| | LOWER QC | HIGHER QC |
|------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| LOW LATENCY Data Available Within Minutes | OPERATIONAL <ul style="list-style-type: none"> Automated QC Useful for trends Exceedance of an action limit may not be a violation | UNREALISTIC |
| HIGH LATENCY Data Released Weeks or Months After Measurements | LOW VALUE | ASSESSMENT <ul style="list-style-type: none"> Manual QC May be comparable to WQ standards |

Figure 1. Categorization of continuous data according to latency and level of quality control.

Operational uses identified by the workgroup included:

- Identification of water quality conditions impeding desired use
 - Additional treatment of drinking water, or temporary intake shutdown
 - Recreational warnings based on empirical relationships between observed parameters and bacteria
 - Rescheduling of irrigation water intake for aesthetic or human/crop health reasons
 - Rescheduling of irrigation water release to protect ecosystem health
- Screening of water quality to prioritize detailed discrete sampling

Assessment uses include:

- Direct comparison to standards to support listing/delisting of impaired waters
- Characterization of receiving water quality and/or sources for TMDLs
- Establishment of baseline or reference conditions for restoration
- Monitoring of long-term water quality trends
- Monitoring of progress and/or BMP effectiveness in restoration projects
- Support development of detailed water quality models for use in any assessment context

Approaches for Sharing Continuous Monitoring Data

Discussion of Figure 1 and its implications with the workgroup identified a useful detail in that some data owners use measurements from a given sensor for both operational and assessment purposes. In this use case, data initially taken from the sensor and subjected to automated QC are used to support operational decisions, and the data are then later subjected to additional QC review and used for assessment purposes.

Access to Data Varies. Workgroup members exhibited a wide range of discoverability (capability for others to learn of the existence of a dataset) and interoperability (capability of others to readily acquire and use datasets) for their data. This diversity was also seen in interviews. At the lower end (Figure 2), some stakeholders described their data as often stored in Excel spreadsheets, others only publish continuous water quality data embedded in tables in reports. Many agencies make their data available for viewing and download through their own or others' data aggregation websites or web portals, but discoverability – particularly for self-run portals – can be low. USGS NWIS is an existing example where data interoperability and discoverability are high; however, NWIS is currently an option only for data collected by USGS through cooperative agreements.

WQX is not currently a viable alternative. Handling of continuous water quality data in the WQX (Figure 3) is not adequate for stakeholder needs, as WQX is designed for use with discrete water quality samples.

- Continuous and discrete data are structurally different, so use of a system built around discrete data to handle continuous data may be inelegant and inefficient.
- The current recommended practice is for partners to summarize continuous data by creating daily, weekly or monthly averages, and then uploading the raw file as a binary large object ("BLOB"). This approach does not make the continuous data discoverable or viewable in the Water Quality Portal, a commonly used tool for accessing water quality data. An additional challenge is that the continuous data stored as "BLOBs" are often not in consistent formats, which further limits their usability and increases the barriers for using the data.
- An alternative approach of storing the continuous observations as a series of discrete points is inefficient in storage, because of the large volume of supplemental information (or "metadata") stored with each discrete point. Ongoing cost reductions for mass storage devices may make this point moot, but there are also concerns about bandwidth requirements necessary for the transfer of data from WQX to other, more accessible data stores, such as the STORET Data Warehouse – particularly if the existing WQX data structure with its high metadata density is maintained.
- Water Quality Exchange is not currently well suited for ingestion and dissemination of operational data that must be accessed in near-realtime (within an hour or less of measurement).



Figure 2. The “As-is” scenario for much of the continuous data sharing is “none” – where the organization collects data and does not make it discoverable and interoperable.

Approaches for Sharing Continuous Monitoring Data

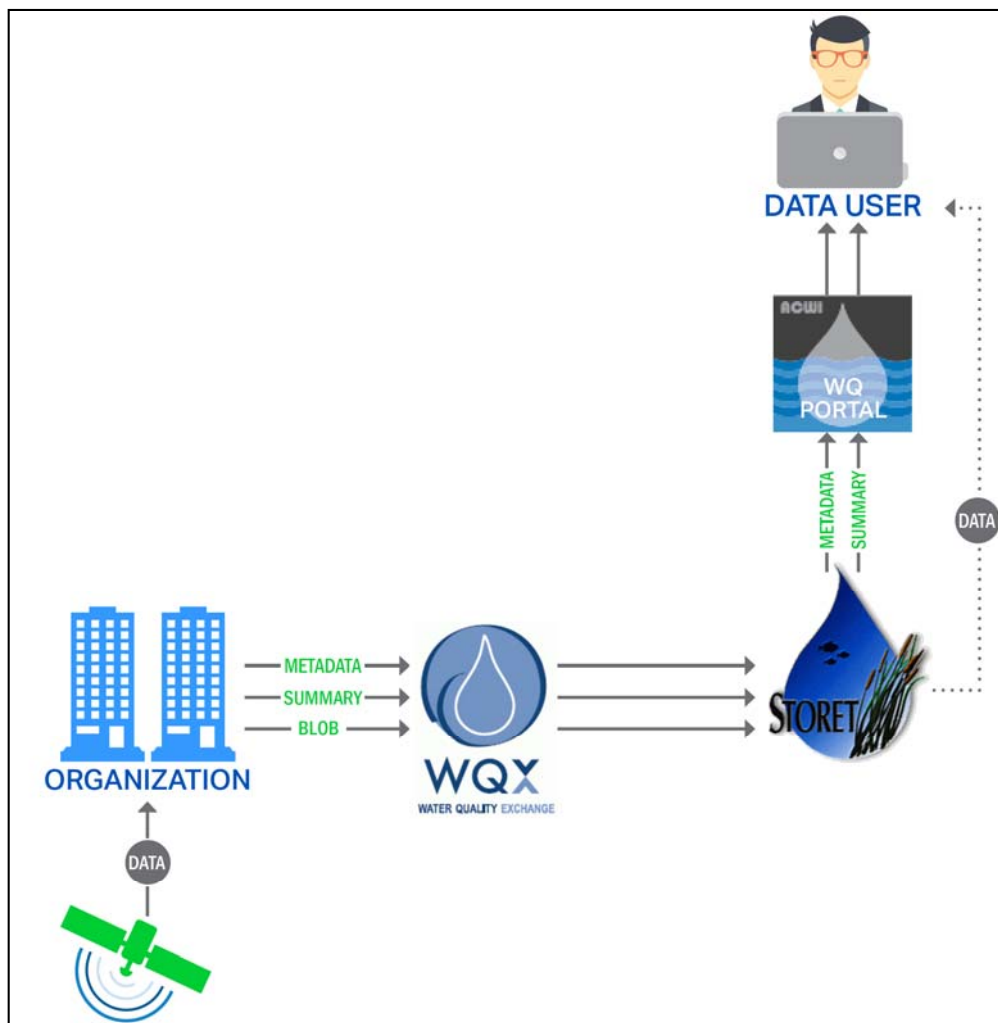


Figure 3. The existing configuration of Water Quality Exchange does not make continuous data readily accessible, requiring familiarity also with Water Quality Portal and STORET.

Interest in compatibility. Stakeholders are accustomed to using tools like the Water Quality Exchange and the Water Quality Portal, and would appreciate minimization of changes in their existing data management workflows under any new strategy for sharing continuous data. Stakeholders would also like to see interoperability between these systems.

Quality Assurance and Quality Control are important. Unsurprisingly, stakeholders and interviewees expressed a strong interest in ensuring that any continuous water quality data collected from in-situ sensors should be subjected to appropriate quality assurance (QA) and quality control (QC) procedures, and that documentation of these procedures should be easily accessible so that users of the data have an assurance of the quality of the data. Data collected for operational purposes might be initially subjected to somewhat less rigorous, automated QC than data used for assessment purposes, but the full range of QC measures may be eventually applied to all data. QA/QC should be standardized wherever possible to encourage those practices and make them easier to implement and document. Raw, “provisional” (less rigorously checked, mostly with automated QC) and “final” (fully QC’d, including manual review) versions of datasets may all need to be stored for reference, regulatory and research purposes.

Approaches for Sharing Continuous Monitoring Data

Research

Research consisted primarily of a review of online resources identified in project staff discussions supplemented in certain cases with telephone conversations to elicit detail missing from available materials. Sources included various entities working with continuous data:

- Existing programs (16) run by federal agencies or offices, such as the U.S. Geological Survey's National Water Information System (NWIS) and the Integrated Ocean Observing System (IOOS) led by the National Oceanic and Atmospheric Administration (NOAA).
 - Existing programs (4) run by other entities, such as the Global Lake Ecology Observation Network.
 - Groups (4) working on development and implementation of standards and tools for interoperable data management, such as the OGC and CUAHSI.
 - Equipment and data management software vendors (8) such as Hach and Aquatic Informatics.
 - Professional associations (11) such as the Ocean Engineering Society of the Institute of Electrical and Electronic Engineers.
6. The full list of resources consulted is included as Appendix 2 - Research Sources.
 7. Many of the agencies, professional associations and vendors consulted in the research subtask were found to participate directly in OGC processes to help shape present and future interoperability standards to better conform to their needs.
 8. The online review and follow-up calls confirmed that there are existing approaches to management of continuous data that may serve as a model for EPA's proposed data sharing network, including centralized systems like the EPA AirNow system (Figure 4). The AirNow system is well designed for its purpose of aggregating and disseminating data that are collected with consistent protocols nationwide.

The research also suggested that standards-based approaches are a promising basis for development of a data sharing network. Much thinking has been put into the development of standards like WaterML 2 (an extensible markup language or XML template for water-related data) and Sensor Observation Service (SOS) (packaging for characterization, discovery and access to continuous data measured by a sensor). A data sharing system built on such standards is conceptually feasible, and has been demonstrated as technically feasible by CUAHSI and IOOS among others. Commercial vendors of sensors and of enterprise data management systems have expressed interest in supporting these standards to increase the utility of their products, and some vendors have already demonstrated interoperability of their data management subsystems, including:

- Lake Simcoe Conservation Authority publishes near-realtime temperature and flow data from selected in-stream sensors to the Great Lakes Observing System data catalog and portal using Kisters' KiWIS product, which provides WaterML 2 and SOS services.
- The LOBOViz portal developed by Satlantic now exposes data from its Land-Ocean Biogeochemical Observatory platforms through SOS. This capability is used by the Gulf of Mexico Observation System to harvest and display data.

Approaches for Sharing Continuous Monitoring Data

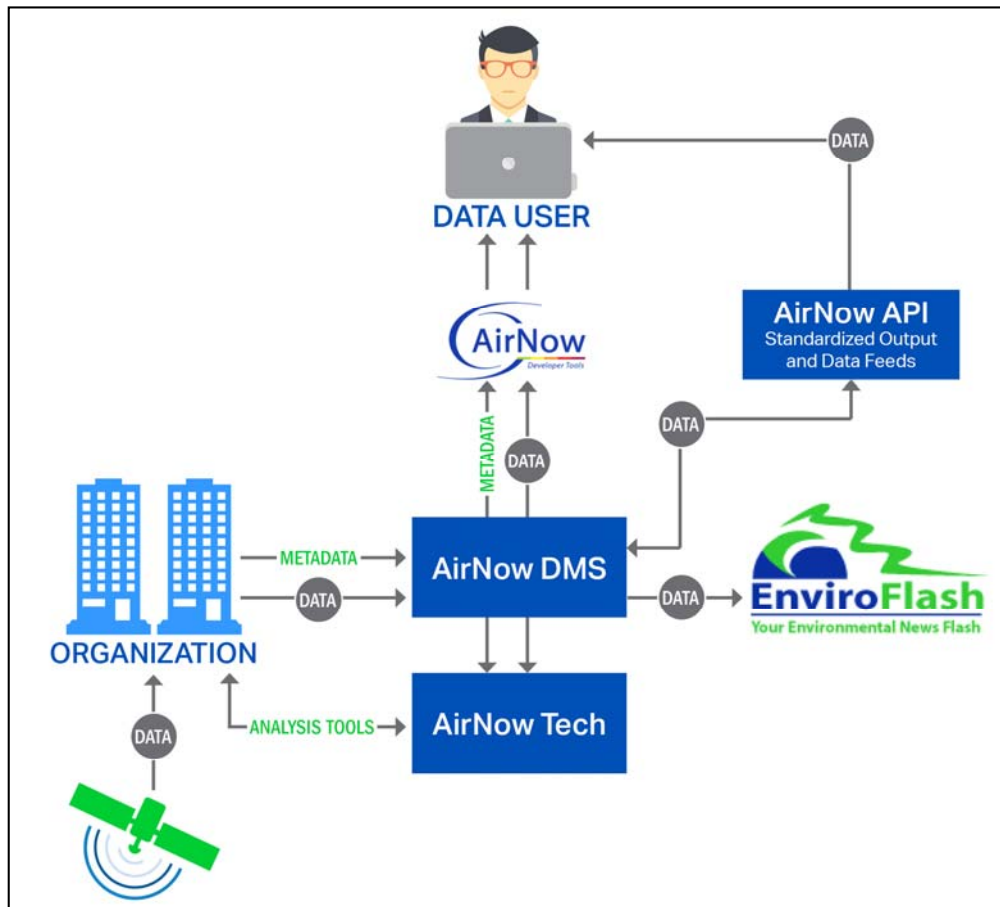


Figure 4. The EPA Office of Air's AirNow system provides centralized ingestion, management and dissemination of ambient air quality measurements whose collection and processing protocols are highly homogenous.

Interviews

In-person interviews were performed with five organizations already working with continuous water quality data in order to gain perspective into use cases, existing practices, directions being considered for future development, and more. The common list of questions used as the starting point for the interviews is included in Appendix 3 – Common Interview Questions.

Three of the organizations were state agencies collecting continuous data to support various operational and assessment needs. Although all three agencies had relatively mature data collection and storage protocols, data access, exploration and visualization were identified as key areas for improvement. These common needs for improved data portals were also identified in workgroup discussions.

The interviewed organizations, relevant network information, and some specific findings are listed below.

Approaches for Sharing Continuous Monitoring Data

- Texas Water Development Board (TWDB) (Austin, TX) operates a network of 600+ sensors focused on water level and conductivity in aquifers.

TWDB's water level data is typically measured hourly and collected semi-daily. Latency is not an important issue because of the long response times of groundwater systems to management actions.

- Texas Commission on Environmental Quality (TCEQ) (Austin, TX) uses in-situ water quality sensors operationally (e.g. to alert irrigators of potentially unsuitable water conditions) and to support assessment.

TCEQ has a mature approach to management of continuous data that includes coordination with partners. Although data are not currently available through OGC web services, TCEQ's upcoming transition to AQUARIUS software for data management will likely realize this capability. TCEQ emphasizes delivery of data of known quality, and appreciates discussions of EPA expectations for submittal of metadata and QA/QC information, especially in the context of TCEQ's existing efforts in recording this information.

- New Jersey Department of Environmental Protection (NJDEP) (Trenton, NJ) deploys sensors directly for assessment and modeling support, and works with partners for source water protection, shellfish bed management, and other purposes.

NJDEP's use cases include mobile airborne and underwater sensor platforms. There is strong interest in making continuous data holdings more widely available through a common portal, and in leveraging national efforts to reduce implementation costs for data sharing. NJDEP is aware of many entities in New Jersey besides themselves who collect continuous data and would welcome improved access.

Approaches for Sharing Continuous Monitoring Data

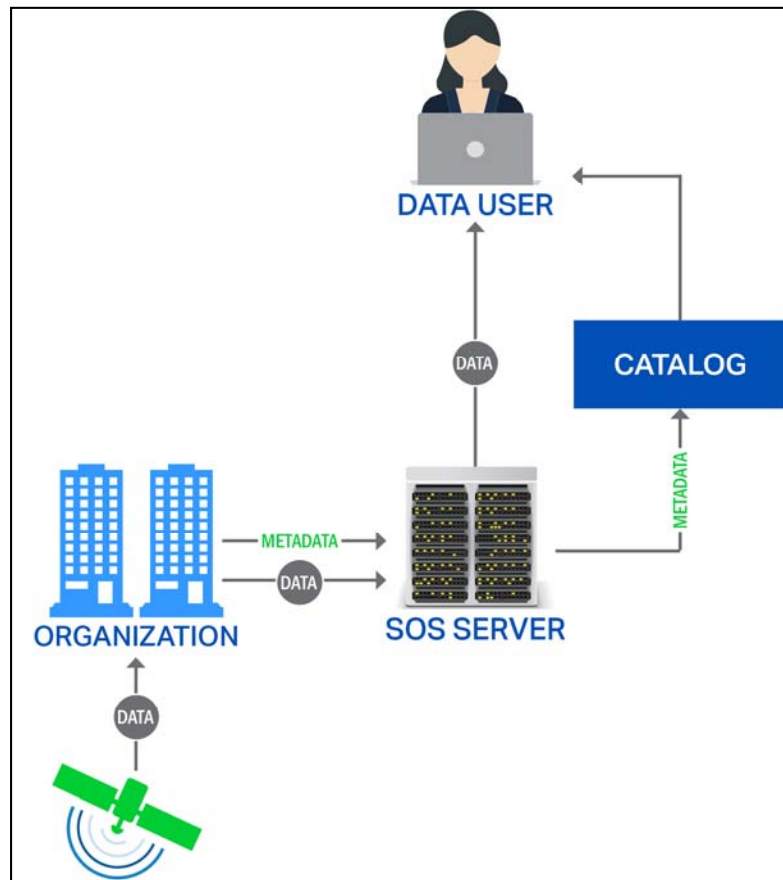


Figure 5. The IOOS infrastructure is based on a centralized catalog linking to an assortment of individual data servers running the OGC Sensor Observation Service (SOS).

9. The Integrated Ocean Observing System (IOOS) Program Office (Silver Spring, MD) coordinates ocean, coastal and Great Lakes observations from over 5000 buoys, gliders and other sensor deployments.

The IOOS data management model (Figure 5) relies on standards-based operability to provide access to heterogeneous data services through a centralized catalog. The data services are implemented, operated and maintained independently by eleven regional associations. Data interchange is facilitated by adherence to standards.

Approaches for Sharing Continuous Monitoring Data

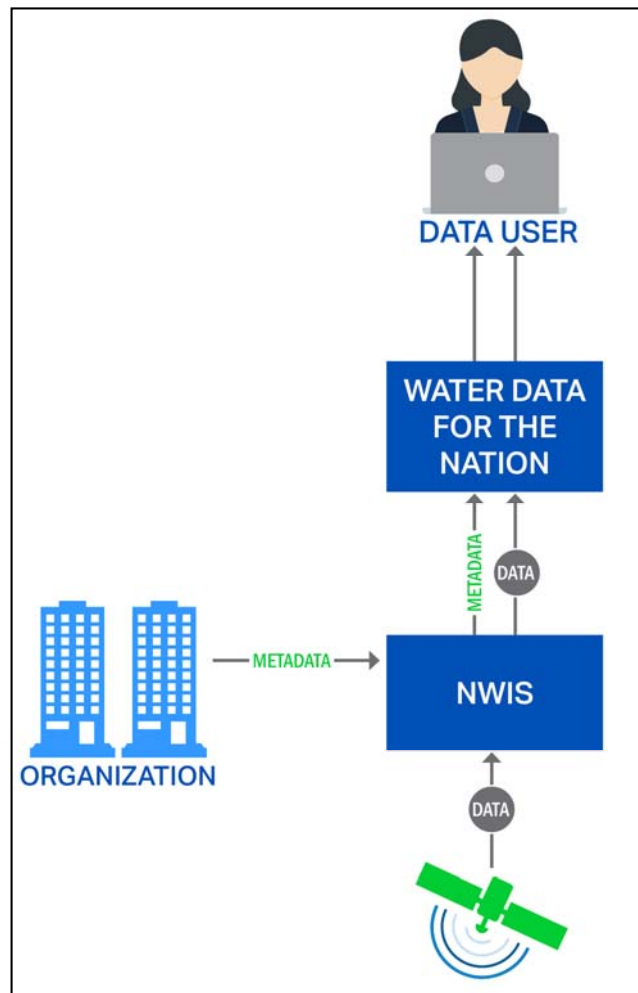


Figure 6. The future, centralized NWIS will keep continuous water quantity and water quality data collected by USGS and cooperative agreement partners readily discoverable and accessible.

- The USGS Center for Integrated Data Analytics (CIDA) and Wisconsin Water Science Center (Middleton, WI) are responsible respectively, for handling of continuous water quality data within USGS, and for application of those data to science problems in the Great Lakes area such as quantification of tributary nutrient loads.

CIDA staff contribute to the development and implementation of the USGS NWIS update, which will provide a simple, effective centralized system for continuous data sharing (Figure 6). Like AirNow, NWIS is effective in part because it is based on homogenous data whose formats and handling are controlled by the funding or regulating agency. However, organizational standards and logistical/funding requirements for now preclude adoption by EPA.

Recommendations

This section presents the recommended technical approach for handling continuous water quality data from sensors. The technical approach is based on the attributes of current and emerging options identified in research, input from stakeholders, and identified use cases for operational and assessment purposes.

Overview

The base recommendation is to develop a system with a centralized catalog and distributed data services that supports discoverability, accessibility and usability of data. This standards-based approach is somewhat similar to the CUAHSI and IOOS systems reviewed in research.

Recommended Standards-Based Approach

During the research phase, standards-based approaches to implementation of continuous data sharing systems were found to be effective for IOOS and CUAHSI, whose universe of data contributors is somewhat similar to the heterogeneous collection of likely contributors to EPA's network. The standards-based approach will allow data owners to select from a range of different solutions to best meet their needs while supporting data streaming to EPA. Two specific standards are recommended as keystones for implementation:

- **Catalog Service for Web (CS-W).** CS-W is an OGC standard for publishing and searching collections of metadata that describe geospatial data and services together with related information objects. Assignment and use of a single common CS-W template will provide a common basis for data contributors to submit site and deployment metadata and links to service endpoints to the centralized catalog, and for data users to search the catalog and discover usable datasets.
- **Sensor Observation Service (SOS).** SOS is an OGC standard that defines a web service interface for discovery and retrieval of time series data from a sensor. SOS provides access to sensor metadata and to observations, delivering the latter in OGC WaterML format. Specification of SOS as the interface for data access allows the data contributor to choose between vendor-supplied or open-source implementations while allowing development of discovery and analytical tools that are unfettered by data format concerns.

Recommended System Architecture and Design

The recommended system architecture calls for the implementation of three components: a centralized catalog, data appliances and an archiving mechanism. The relationship between these components and existing data systems is shown in Figure 7, where:

- Site and deployment metadata are submitted by an organization to an extended WQX serving as a catalog;
- Data are available through service endpoints exposed by data appliances tied to the organization;
- Data users discover data of interest with a discovery tool by querying metadata in the catalog exposed either directly from WQX or passed through the STORET Data Warehouse and the Water Quality Portal, and then retrieve the data from the data appliances;
- Data are archived for backup, redundancy and/or regulatory reasons in a modified WQX.

Data discovery and access from similar sensor data sharing networks, such as IOOS and NWIS, is enabled by the use of similar standards-based approaches in those networks.

Approaches for Sharing Continuous Monitoring Data

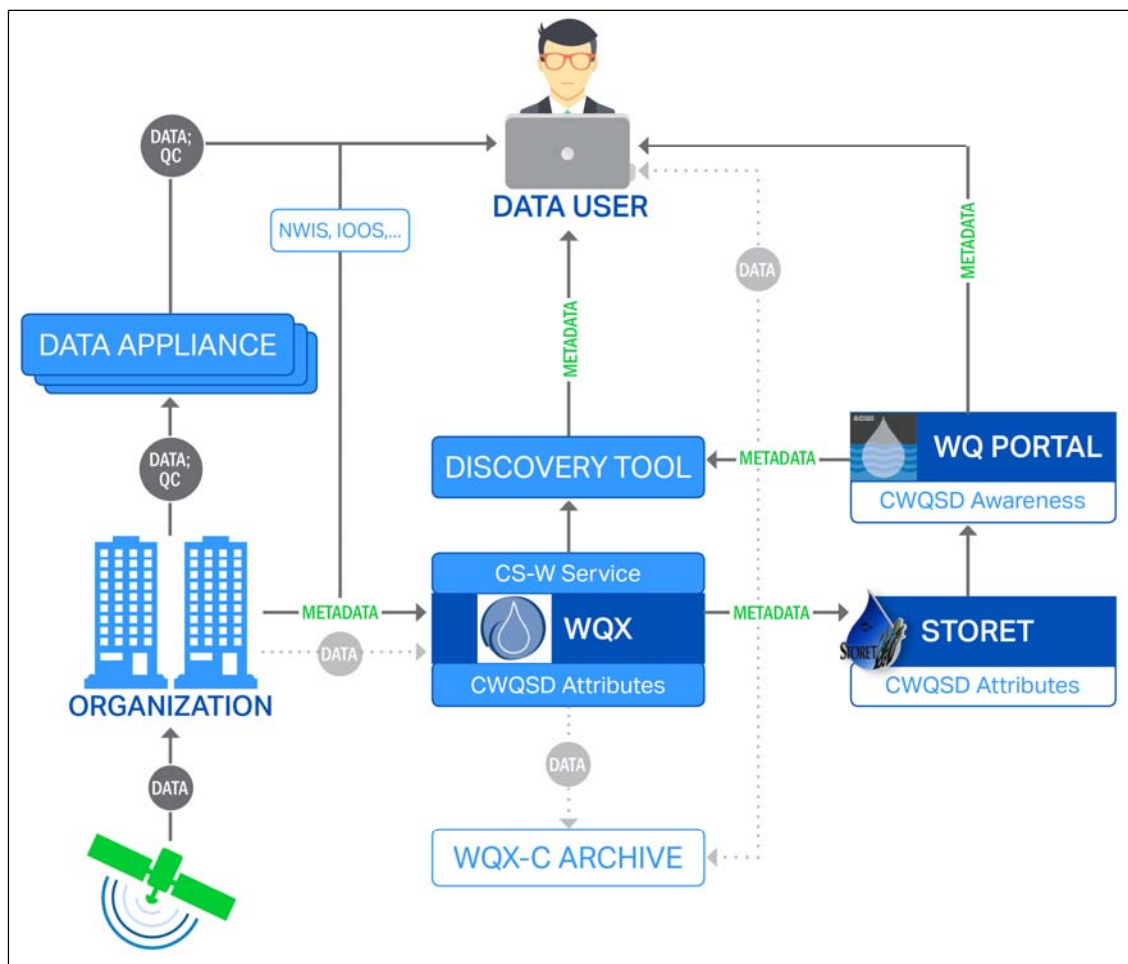


Figure 7. The recommended technical approach includes distributed data services provisioned through data appliances and a centralized catalog based in WQX.

This recommended approach leverages existing WQX infrastructure to support efficient centralized discovery of continuous water quality datasets while promoting data accessibility through standards-compliant data delivery services tied to the data owner.

Catalog

The current WQX/STORET/WQP system provides effective storage, discovery and data access for discrete sample data submitted to WQX by data owners, but is not well structured to store continuous data. However, site metadata for continuous data sensors and, possibly, QC-related lab and discrete sample data, could be stored in WQX either by overloading existing fields (e.g., AlternateMonitoringLocationIdentity group) or by adding continuous data-specific fields. This will allow data owners familiar with WQX to continue submitting station data in a well-understood manner, and provide access for discoverability purposes through the widely used WQP. Discoverability will require that STORET and WQP be aware of the attributes added to WQX to handle CWQSD.

Approaches for Sharing Continuous Monitoring Data

| |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| WQX |
| <Organization> |
| ..<MonitoringLocation> =>Would be the same structure for sensor locations and may even use the same station as discrete samples. Would include additional metadata to indicate that the monitoring location is used as a sensor station |
| ..<Activity> =>Used to describe the deployment of a sensor including additional metadata about calibration and other QA/QC |
| ...<WaterML2> =>Used to store all the sensor data |
| ..<ActivityGroup> =>Used to correlate a sensor deployment with QA/QC samples or calibration samples. |

Figure 8. The WQX schema can be extended to capture site and deployment metadata as well as observations for continuous data.

Station metadata would therefore flow from the organization to a CWQSD-aware WQX (Figure 8), which would then push station metadata on through the STORET warehouse and into the WQP for users to discover. Once a station and associated CWQSD metadata are discovered, the user will be able to access the actual data from the data delivery services for the owning organization.

Data-owning organizations will be responsible for submittal of station and deployment metadata to the centralized catalog, which will be managed by EPA. The metadata about stations and sensors can be submitted to an enhanced WQX to promote discoverability, accessibility and usability of CWQSD:

- Discoverability will be realized either through addition of Catalog Service for Web (OGC, 2015a) service endpoint to WQX or alternately by flow-through of metadata to Water Quality Portal via STORET Data Warehouse.
- Data access will be provided through Uniform Resource Identifier pointing to associated data appliance included in metadata.
- Usability of data will be facilitated by descriptive information about station purpose and location, and sensor details such as linked calibration data.
- WQX enhancement would also include updates and modifications to WQX node clients to facilitate data owners' handling of CWQSD data and metadata, such as recognition and parsing of new attributes.
- Guidance and processing for metadata will emphasize station georeferencing and harmonization of duplicate locations.
- Metadata submittals may be automatically checked for compliance with applicable requirements for conformance to standards.

Data Appliances

Continuous water quality stations registered in WQX will include a link to appropriate data delivery services for the dataset. These data delivery services will reside on a "data appliance" that may be administered directly by the data owner or optionally by EPA, and may be located on-site or in the cloud. The data appliance will ingest and store observational data, perform a series of automated quality control steps, and make the data and relevant metadata available through data services compliant with the Open Geospatial Consortium (OGC)

Approaches for Sharing Continuous Monitoring Data

Sensor Observation Service (SOS) standard. OGC SOS provides access to data through the WaterML 2.0 standard and to metadata through the SensorML standard.

Organizations will be primarily responsible for the publication of their data using OGC SOS services to realize a distributed data architecture. The services will be made available using one of the following data appliance platforms:

- An on-site server owned, configured, operated and maintained by the data owner.
- An on-site or cloud server running a vendor-supplied data management system that exposes data through an installed SOS capability.
- A cloud server instance configured by EPA or an EPA contractor to act as a data management and aggregation center for one or more data owners.
- Open-source technology stacks are recommended for EPA-configured data appliances.

All data contained in a data appliance should be discoverable. Organizations would set-up automated procedures to publish data that they wish to share to the data appliance on a regular schedule, if not in real time. If organizations have data that they do not want to share, they should not publish that data to the data appliance.

Archiving

A complementary centralized archiving component for long-term storage of datasets could serve as a permanent repository of record and also provide additional data redundancy. This storage system may be managed by EPA or by a third party:

- The archive could be implemented by EPA as an enhancement to WQX or as a new, separate data store optimized for long-term compact storage of continuous water quality data.
- The archive could alternately be made part of NOAA's National Centers for Environmental Information, with submittal protocols and processes jointly developed and managed by EPA and NOAA.
- Other considerations for archiving include: Archival submittals could be on a "push" basis with data owners explicitly transmitting datasets to the archive or on a "pull" basis with automated harvests of datasets from the organizations' data appliances.
- Storage and archiving of continuous data in netCDF format may provide benefits due to the wide range of tools that can be applied to this format.

10. Additional exploration is needed before final recommendations are made for implementation.

Peer systems research and interviews have found that there are already successful sensor data sharing networks, including USGS NWIS and NOAA IOOS. Because of these networks' standards-based approaches to implementation, there is a strong expectation of interoperability with the recommended EPA network, with the key step being crosswalk and harmonization of site and deployment metadata to support service-based injection or federation of other networks catalogs with the new EPA catalog.

System Management

The responsibilities for management of the recommended system will be shared between EPA and the organizations that own the data.

- Organizations will be primarily responsible for the publication of their data using a data appliance.

Approaches for Sharing Continuous Monitoring Data

- Organizations will be responsible for submittal of appropriate site and deployment metadata to a centralized catalog managed by EPA.
- EPA is anticipated to manage archival systems.

Data Types and Sources

Data considered under this recommendation consists of measurements collected from stationary in-situ instruments at regular time increments (typically a few minutes to a few hours) over an extended deployment period. Water quality measurements based on lab analysis of samples collected with an automated sampler are not considered sensor data, and are therefore not included. Data collected from sensors installed on a moving platform such as an underwater autonomous vehicle (UAV) is also not considered under this recommendation, although this type of data should be considered for future enhancements. Remotely sensed raster (gridded) data were also not considered, but should be a consideration for future expansion of the data sharing network.

The data source for a given data set is considered to be the organization owning the data. Instrument details associated with the collection of the data, such as particular sensor types, etc. are not considered to be data sources, although instrument make and model can be entered in the deployment metadata to improve transparency.

Data Quality

Effective use of collected CWQSD datasets is dependent on the quality of the data. The technical recommendation includes the following data quality suggestions:

- The data appliance used to serve up measurements should provide a set of automated QC tests that identify observations that may be suspect due to (at a minimum) data loss, out-of-range conditions, stuck sensors, or spikes.
 - The tests may be based on existing quality control processes for low-latency data, such as the NOAA IOOS Quality Assurance of Real-Time Oceanographic Data (QARTOD) or USGS Continuous Record Plan (CRP) protocols.
 - Data owners and EPA should identify opportunities to work with the sensor manufactures to encourage them to make additional sensor diagnostic data available that are being recorded by the sensor, but are not readily available to the sensor user. This diagnostic data may provide insight into the quality of a result.
 - The data owning organization may choose whether the automated QC tests flag or censor suspect data.
 - Data that has received no quality control testing will be considered “raw” data; data that has received automated QC testing will be considered “provisional” data.
- More rigorous quality control based on comparison of raw or provisional data to data collected at other sensors and to lab and discrete in-situ grab samples may be used to develop “final” data.
 - Lab control samples and discrete grab samples collected for CWQSD sensor calibration will ideally be stored in WQX and linked to a sensor deployment record.
- The data owning organization may choose which version(s) (raw, provisional or final) of a dataset to publish through its data appliance and which version(s) to archive.

Approaches for Sharing Continuous Monitoring Data

- If data are used operationally, the provisional version of the data is expected to be published and archived.
 - If data are used for assessment, the final version of the data is expected to be published and archived.
 - Provisional and final versions of a dataset may be simultaneously available from the data appliance and both be archived.
- The data appliance must support the simultaneous publication of raw, provisional and final versions of a dataset.
 - The data appliance may provide tools to support quality control processes, such as retrieval from the catalog or the Water Quality Portal of related calibration data, charting of calibration data versus CWQSD and correction or censoring of measurements.

Data Submittal and Reporting

Under this recommendation, data-owning organizations will be responsible for registering the sensor and providing deployment metadata, and discrete quality control measurements, to WQX:

- Station metadata will describe the location where a continuous water quality sensor is deployed. Station locations will be checked against the locations of other stations to identify possible duplicate entries, and will be georeferenced to National Hydrography Dataset addresses.
- Deployment metadata will identify the sensor deployed and the dates of deployment, possibly accompanied by calibration and maintenance notes
- Quality control measurements, possibly including lab control samples and pre-, intra- and post-deployment grab samples will be entered into WQX and linked either directly to a deployment using a deployment ID or indirectly using deployment dates and location or via Activity Groups in WQX.
- WQX will provide standard reports describing the success of data submittals.
- If a “push” model is adopted for archiving, organizations will be responsible for submitting data to the archiving system, which may be an extension to WQX, a NCEI repository or a new data store. Alternately, organizations would be responsible for defining the frequency of harvesting for each archived CWQSD dataset and for indicating which versions (raw, provisional, or final) of datasets to harvest. The archive system will provide a report acknowledging each archived dataset based on that defined schedule. The archive system would harvest data using the services available on the data appliance.

The organization may also receive status reports from the data appliance (if used) summarizing data ingestion and automated QC results.

Management of Related Information Sources

This recommendation aims to avoid duplication of existing data holdings such as NWIS and IOOS. Although the management of these holdings is independent of EPA, steps can be taken to integrate and harmonize with existing continuous data sharing networks:

- The centralized catalog system must be aware of sampling locations in each different network. This awareness may come into play at submittal time or during discovery.

Approaches for Sharing Continuous Monitoring Data

- Harmonization of CWQSD metadata template with templates from peer networks will minimize the overhead of mapping between standards.

Sharing/Retrieval Capabilities and Tools

The mechanics of retrieving published CWQSD datasets will be inherently supported by the use of OGC SOS-compliant services on data appliances. Sharing of CWQSD datasets will be driven by discovery through query of the centralized catalog system:

- Direct query of CWQSD station and deployment records in WQX may or may not be implementable through provisioning of an OGC CSW-compliant service
- The existing WQP may more easily be extended to work against an enhanced STORET Data Warehouse that receives CWQSD site and deployment data and support discovery of CWQSD deployments based on parameters, deployment period, geography and more.
- The extended WQP should also recognize and discover CWQSD sites in other systems such as NWIS and IOOS.
- Additional WQP enhancements could support retrieval of CWQSD data from organizations' data appliances or from archives.
- Retrieval of CWQSD data combined with retrieval of coincident or nearby grab sample data will support direct comparison of continuous and discrete data for research and assessment purposes.

Policy Gaps and System Needs

Five areas of the recommendation require additional specification that may best be explored in the pilot phase proposed in the next section's implementation strategy:

1. The recommendation calls for the publication of data by organizations using the OGC SOS standard. One option is to use a data appliance configured by EPA. There are two leading alternatives for delivering SOS on such an appliance – 52 North SOS and ncSOS. The chief difference between the two is that ncSOS stores data directly in netCDF format, which is convenient for some archiving options, while 52 North uses a relational database as the primary store but can export readily to netCDF.
2. The recommendation identifies three options for long-term archiving of data, but data are currently insufficient for a specific recommendation between the three options (enhance WQX, leverage NCEI, build a new system). Further investigation is needed to flesh out archiving requirements and evaluate suitability of available options to fulfill those requirements.
3. A list of recommended automated quality control tests is not specified here. A crosswalk of USGS CRP and IOOS QARTOD protocols and review of available tools during the pilot is suggested to establish this list.
4. Interoperability with existing systems, such as IOOS and NWIS, calls for a mechanism to identify duplicate sites. It is currently unclear whether the identification of duplicate sites should happen when a CWQSD site is submitted to WQX or during the discovery process. It is also unclear how the identification of potential duplicates will happen, whether the mechanism will identify both potential intra- and inter-organizational duplicates, and what action, if any, is to be taken if a duplicate, coincident or co-located non-WQX site is found.

Approaches for Sharing Continuous Monitoring Data

5. Additional input will be sought from data-owning organizations on desirable optimizations and enhancements to WQX node clients on the Exchange Network to minimize disruption to existing workflows for submittal of site and deployment metadata.

Also noted are potential or anticipated needs to address continuous data from mobile platforms, and to engage stakeholders in discussions about the use of this data for Clean Water Act purposes.

Collaboration and Implementation Strategy

This section presents a suggested collaboration and implementation strategy for a continuous water quality data management system. The strategy considers impacts on data owners, identifies likely partners and presents an initial timeline for implementation.

Program Management and Partners

As described under the recommendation section above, data management responsibilities will generally reside with the data-owning organization, while metadata for sites and deployments will be managed by EPA through the enhanced WQX. Management of archived data will be the responsibility of EPA (enhanced WQX or new store) or NOAA NCEI.

Overall, EPA must be the program manager for the sharing of water quality data, with the following key responsibilities:

- Development and operation of the centralized catalog as a repository for the CWQSD site and deployment metadata.
- Development of tools and guidance to facilitate data owners' submittal of CWQS metadata.
- Development and implementation of a mechanism allowing access to centralized CWQS metadata for discovery.
- Development and (potential) operation of cloud-based data appliance instances that ingest, store and serve up CWQS data.
- Monitoring of centralized catalog and data appliances to ensure correct operation.

A number of partners will be involved and EPA should expect to partner with several entities:

- USGS will collaborate on adding CWQSD awareness to the Water Quality Portal; development of specific mechanisms for harmonization of NWIS site metadata and continuous water quality data; and identification of appropriate automated QC procedures for inclusion on data appliances.
- NOAA will collaborate on the development of specific mechanisms for harmonization of IOOS site metadata and continuous water quality data, and identification of appropriate automated QC procedures for inclusion on data appliances.
- OGC will collaborate on the development of mechanisms for effectively tying CWQSD datasets to related QC data, such as pre-, intra- and post-deployment lab and site discrete sample measurements.
- Vendors will be encouraged to provide standards-based access to sensor data through their proprietary tools in a manner consistent with the desired goals of discoverability, access and usability for data sharing.

As a major funder of monitoring programs, EPA should also consider the development of guidelines that promote contribution of data to the continuous water quality data sharing network.

Implementation at the State/Agency Level

An important consideration in the development of the technical recommendation and the collaboration and implementation strategy is the impact on data owners. For evaluation of potential impact, a representative workflow for deployment of a continuous water quality sensor and collection of its data was considered (Table 1):

Most of the steps in the workflow remain unchanged with the following exceptions:

- Creating station metadata (step 4). Station metadata would include identification as a CWQSD station and a link to the standards-compliant web service for data access.
- Creating deployment metadata (step 6). Deployment data would likely include at a minimum instrument identification and deployment dates, and a link to a station.
- Recording deployment QC data (step 15). QC data would be tied to deployment by deployment ID or by instrument/date combination, and then stored in WQX.
- Posting near-realtime data (step 9). Data would be posted, if appropriate, to the organization's associated data appliance and made available through standards-compliant web services.
- Performing QC (step 16). Quality control may be facilitated by tools packaged with the data appliance.
- Replacing provisional data with QC'd data (step 17). The data appliance will be able to differentiate between raw, provisional (e.g. subjected to automated QC) and final (subjected to full QC procedure including manual review) datasets, and to control access to each type of data.
- Submitting data to WQX (step 18). Continuous water quality sensor data may only be submitted to WQX for archival purposes. The submittal may a manual "push" with the data owner directly uploading through the Exchange Network, or an automated "pull" from the owner's data appliance on a schedule set by the data owner. The submittal could also be to an independent data store or to NOAA NCEI.

Table 1. Representative Workflow

1. Identify monitoring needs (parameters, frequency, period)
2. Identify station location
3. Prepare site
4. Create station metadata
5. Prepare deployment
6. Create deployment metadata
7. Pre-deployment lab measurement
8. Begin recording observational data
9. Begin posting near-realtime data
10. At-deployment in-situ grab
11. Intra-deployment cleaning and in-situ grab(s)
12. Post-deployment in-situ grab
13. Stop recording observational data, posting NRT data to local site
14. Post-deployment lab measurement
15. Record deployment QC data
16. Perform QC
17. Replace provisional data with QC'd data

Feedback will be sought from data owners during the proposed pilot phase (see below) on how best to integrate new functionality into existing WQX node tools.

Approaches for Sharing Continuous Monitoring Data

Schedule

The schedule below outlines the principal activities anticipated in the implementation of the recommendation.

2015-2016: Pilot and Refine

EPA will identify a suitable watershed (or watersheds) on which to exercise and test the recommended technical approach on a pilot basis. During this pilot phase, EPA and its contractors will work on detailed specification, development and deployment of the necessary components:

- Centralized catalog framework and submittal procedures
- Distributed data appliances, including automated QC processes and reports
- Archive repository and procedures
- Prototype analytical tools that provide discovery and access

During this period, EPA should also engage with the vendors of sensors, data loggers and sensor data management tools to inform commercial software's capabilities for interoperability within the data sharing strategy recommendation. EPA should also engage with peer organizations like USGS and NOAA to harmonize metadata requirements and develop procedures for identifying and handling duplicated, coincident and co-located stations and their associated data in related data sharing systems like NWIS and IOOS.

Lessons learned from the pilot implementation and from interactions with vendors, partners and data owners should be gathered and reviewed to identify potential adaptive changes to the strategy. Reports to the stakeholder community through presentations at national conferences, such as AWRA or the National Monitoring Conference, are also advisable.

EPA recognizes that the recommended architecture can be used with many other types of continuous sensor data besides water quality, such as water levels, streamflows, precipitation or air temperature. EPA therefore intends to explore options for ingestion of such data or for interoperability with other enterprises aggregating these data.

2017-2018: Build Community and Framework

Following completion of pilot projects and refinement of the data strategy, the focus will be on building tested, scalable infrastructure for an increasing community of organizations sharing CWQSD. The proposed activities for 2017-2018 are:

- Build, and test centralized elements (catalog, archive)
- Refine data appliance configuration
- Implement additional analytical tools
- Develop and promulgate documentation and guidance
- Identify and support early adopters
- Build community through sharing at conferences.

Approaches for Sharing Continuous Monitoring Data

2019 - Scale up and Build Out

Once the tested centralized infrastructure is available, data owners will be able to stand up their own data services to provide access to their CWQS data while placing the associated site and deployment metadata on the WQX-based catalog.

References

EPA, "STORET/WQX: What is WQX?". <http://www.epa.gov/storet/wqx/>. Accessed July 13, 2015.

Open Geospatial Consortium. "Catalogue Service". <http://www.opengeospatial.org/standards/cat>. Accessed July 13, 2015 (a).

Open Geospatial Consortium. "OGC® WaterML".
<http://www.opengeospatial.org/standards/waterml>. Accessed July 13, 2015.

USGS, EPA, USDA. "What is the WQP". http://www.waterqualitydata.us/wqp_description/. Accessed July 13, 2015.

Additional references to be included in final version.

Appendix 1 – Workgroup Representation

The following organizations were represented on the Continuous Monitoring Data Strategy Workgroup formed for stakeholder input into the development of a continuous monitoring data sharing strategy:

- Aquatic Informatics
- Bad River Band of Chippewa
- Bay Mills Indian Community
- California State Water Resources Control Board
- California Department of Water Resources
- Consortium of Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI)
- Delaware River Basin Commission
- U.S. Environmental Protection Agency (EPA) (HQ and seven Regions)
- Fond du Lac Reservation
- Great Lakes Indian Fish and Wildlife Commission
- Hannahville Indian Community
- Houlton Band of Maliseet Indians
- Idaho Department of Environmental Quality
- Indiana Department of Environmental Management
- Lac du Flambeau Band of Lake Superior Chippewa Indians
- Lac Vieux Desert Band of Lake Superior Chippewa
- Massachusetts Department of Environmental Protection
- Maine Department of Environmental Protection
- Minnesota PCA
- New England Interstate Water Pollution Control Commission
- National Park Service
- Oklahoma Water Resources Board
- Oregon Department of Environmental Quality
- OTT Hydromet
- Pennsylvania Department of Environmental Protection
- Penobscot Indian Nation
- Pokagon Band of Potawatomi
- Red Lake Nation
- Susquehanna River Basin Commission
- St. Croix Band of Chippewa
- Susquehanna River Basin Commission
- Texas Council on Environmental Quality (TCEQ)
- University of Michigan
- University of Texas-Austin
- U.S. Geological Survey (USGS)
- Virginia Department of Environmental Protection
- Wisconsin Department of Natural Resources
- Woolpert LLC
- West Virginia Department of Environmental Protection

Appendix 2 – Research Sources

Six categories of sources were consulted for insight into continuous monitoring data sharing strategies:

Sources initially identified in EPA’s scope of work for the project

The seven sources identified by EPA in the scope of work for this project were treated as high-priority and therefore given the greatest level of scrutiny.

- Consortium of Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI)
- National Oceanic and Atmospheric Administration (NOAA) Integrated Ocean Observing System (IOOS)
- National Groundwater Monitoring Network (NGWMN)
- National Water Information System (NWIS)
- Open Geospatial Consortium (OGC)
- Water Quality Portal (WQP)
- Water Quality Exchange (WQX)

Additional higher-priority sources

These sources were identified by the subject matter experts working on this effort as also highly relevant. Their inclusion was accepted by EPA.

- NOAA Environmental Data Management Framework
- NOAA Meteorological Assimilation Data Ingestion System
- NOAA National Weather Service
- Alliance for Coastal Technologies
- OceanSITES

Additional medium-priority sources

These sources were identified by the subject matter experts as relevant and accepted for inclusion by EPA.

- EPA Emissions Measurement Center
- EPA Ambient Monitoring Technology Center
- NOAA National Data Buoy Center
- NOAA National Estuarine Research Reserve System
- NOAA National Oceanographic Data Center
- National Science Foundation (NSF) National Ecological Observatory Network
- European Union – U.S. Cooperation (COOPEUS)
- Group on Earth Observations

Additional lower-priority sources

These sources were identified by the subject matter experts as potentially relevant and accepted for inclusion by EPA. These sources received less scrutiny than high- and medium-priority sources.

- Commonwealth Scientific and Industrial Research Organization

Approaches for Sharing Continuous Monitoring Data

- NOAA Animal Telemetry Network
- NSF Critical Zones Observatory
- Global Lake Ecological Observatory Network
- Global Rivers Observatory (low)

Vendors

The subject matter experts identified a number of equipment and data management software vendors as being potentially relevant. Inquiries were made as to vendor awareness of and support for OGC and similar standards. The mention of trade names or commercial products and/or vendors does not constitute endorsement or recommendation for use by EPA.

- Campbell Scientific
- Hach
- Satlantic
- Aquatic Informatics
- ERAMOS
- Kisters
- OneRain
- Weatherflow

Professional Associations

With EPA's agreement, the subject matter experts looked into the level of interest and expertise of a number of professional associations with respect to continuous water quality data.

- American Water Works Association
- American Water Resources Association
- American Society of Civil Engineers - Environmental and Water Resources Institute
- Institute of Electrical and Electronics Engineers - Oceanic Engineering Society
- Instrumentation Testing Association
- International Society of Automation
- International Water Association
- Marine Technology Society
- Water Environment Federation
- Water Environment Research Foundation
- Water Research Foundation

Appendix 3 – Common Questions for Interviews

The following questions were posed to each interviewee as a starting point for discussions:

1. Base Questions
 - a. How do you define continuous data?
 - b. What are your applications for continuous data?
 - c. What problems are better answered with continuous data?
 - d. What is driving your use of continuous data?
 - e. What equipment are you using to collect continuous data?
 - f. How are you sharing the data and with whom?
 - g. How are you assuring and documenting data quality?
 - h. Other entities?
2. Data Latency
 - a. How long from collection to publication?
3. Data Retention
 - a. How long are data kept?
4. Metadata and QA/QC Data Management
 - a. Protocols and procedures
 - b. Storage
 - c. Dissemination/distribution
5. Funding sources?
6. Data Collection and Analysis
 - a. What are the regulatory or other drivers for collection and analysis of data?
7. Challenges
 - a. What obstacles do you face, if any, to making data generally available (technical or administrative)?
8. Publication Locations
 - a. Are there multiple publication locations? What are they?
9. Data Assembly
 - a. Do you have or use local or regional data assembly centers?
10. Instrumentation
 - a. What specific instrumentation is being used to collect continuous data?
 - b. How long are the sensors deployed (days, weeks, permanently, other)?
 - c. Is there data processing performed at the sensor?
 - d. How is data transmitted/received?
 - e. What is the sensor maintenance/calibration frequency?