Exchange Network Forum

Thursday, February 8, 2024
1:00 to 2:00 PM ET

Agenda:
- Welcome
- Setting the Stage
- Environmental Information & Innovation National Meeting (E2i)
- Water Quality Exchange
- Questions & Answers
- Open Forum

Welcome and Opening Remarks

Alex O'Neill, EPA Office of Information Management (OIM), Office of Mission Support (OMS)

- The Exchange Network (EN) Forum is an open meeting, and the goal is for the meeting to be collaborative.
- Future calls will have breakout groups to brainstorm ideas.
- Polls and surveys will be used to gauge interest in topics.
- Participants are encouraged to ask questions while also being mindful that the audience of these calls is broad with technical and non-technical participants. If a question is too technical, EPA may need to follow-up after with an answer.

Setting the Stage

Alex O'Neill, EPA Office of Information Management (OIM), Office of Mission Support (OMS)

- The April EN Forum will focus on the Combined Air Emissions Reporting System (CAERS). CAERS is an application that allows industry from subscribed state, local and Tribal authorities to report their air emissions data. The presentation will focus on the CAERS Application Programming Interface (API). APIs are an important part of the Network and can be used to streamline data exchanges.
- In the summer, the EN Forum calls will focus on the EN Grants Solicitation Notice and data reporting processes through specific EN data flows. During the Forum calls, EPA programs will share how they make data submitted by EN partners available back to the partners/submitters.

Environmental Information & Innovation National Meeting (E2i)

Alex O'Neill, EPA Office of Information Management (OIM), Office of Mission Support (OMS)

- As background, the EN National Meeting is a key part of an ongoing effort to reinvigorate Exchange Network partnerships.
- For a long time, the EN was a fairly lively and interactive group; however, over the past five years, the Network experienced fewer, direct interactions between the EPA and its Tribal, state and territorial partners. It has been a gradual decline, and the dimming lines of communication were only exacerbated by the pandemic and staff transition among teams.
- For the EN to work effectively, partners need to have more frequent direct interactions. This is part of the impetus for the EN Forum. The EN needs to be reinvigorated by injecting new energy and strength to the program with a positive legacy of impact.
- The National Meetings are a key place where the Network partners can get together to hear from each other and learn what other environmental organizations are doing to improve environmental information exchanges to protect human health and the environment.
- This year's Environmental Information and Innovation (E2i) National Meeting is an opportunity for environmental professionals to come together to collaborate and learn from one another. The meeting will build on the meetings of the past. The theme of this year's meeting is reinvigorate, collaborate and innovate.
- The meeting will be a live, in-person event with some virtual components. The tentative location is Kansas City, Missouri on September 17-19, 2024.
- There's a group of EPA, state and Tribal partners who are working to plan the program. The call for session ideas will go out the week of February 12.
- The following are examples session topics. The sessions will include lightning talks, podium presentations, panel discussions, innovative videos, workshops, etc.:
  - Innovations related to the Exchange Network and its data flows.
  - Telling environmental stories through data visualization.
  - Improving data management processes and procedures.
  - Leveraging geospatial data for environmental management.
  - Improving user experience in web applications.
  - Using Open Data, APIs and analytics to improve decision-making.
  - Applying Artificial Intelligence in environmental protection.
  - Adopting cloud technologies and shared platforms to drive efficiencies.
  - Improving mobile data entry and fieldwork through technology.
  - Building partnerships within and among environmental agencies.
  - Examples of cross-agency and cross-partner collaboration.
  - Finding intersections between data, science and policy.
- Partners who are interesting in getting involved should reach out to Alex O'Neill.

Water Quality eXchange (WQX)

Adam Griggs, EPA Office of Water, Office of Wetlands, Oceans and Watersheds (OWOW), Water Data Integration Branch

- Adam gave an overview and the history of WQX, which is a standardized data format and submission database.
- Please review the [meeting recording](#) and [Adam's presentation](#) about WQX and its data systems, including:
  - **Water Quality Portal**, which is a data warehouse that uses web services to access and provide data from USGS, EPA, USDA and NPS in a standard WQX format.
  - **ATTAINS** (Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System) is an online system for accessing information about the conditions in the Nation's surface waters.
  - **TADA** (Tools for Automated Data Analysis) is a tool for automated data analysis.
  - **How's My Waterway** is the public-facing view of the data. It provides a holistic view of water in communities. The tool accesses, interprets and displays data from over a dozen sources to help make decisions.

Questions and Answers
The following are questions and answers from the session.

Q: What is TADA?
A: TADA is the newest crowdsourced developed tool for automated data analysis. It helps people build and refine queries of data from the water quality portal. It has the capability to do some data reformatting to ensure users are comparing apples to apples, and it does some analysis graphing.

Q: Does the WQX helpdesk also help with TADA questions?
A: Users can contact the help desk for TADA questions. They will submit the inquiry to the TADA lead.

Q: Is the WQX submission's QA/QC automated?
A: Yes, every time a user submits data to WQX they will receive a report back with some additional columns that will flag data where there is not agreement, such as proper units. The system will not change any of the data, but instead will show the user changes that may need to be made. The user can then correct the submission and resubmit the data. This process currently happens after upload, but in the future the team is hoping the QA/QC can be done before upload.

Q: How do people find and download the continuous time series data attachments?
A: The team is currently looking into additional discoverability capabilities so users can find continuous time series data sets. Currently users are going to discover these datasets the same way they'll discover any other data. If they put in the query for data in a location between certain dates, they will get a

result with that data. When looking at the data, they'll see that some data are not measured values, they might be a daily mean or beyond a daily max, and so they'll see the data are not discrete and measurable but a summarized value from a time series. If they scroll over, one of the fields will have a result of an attachment and in that attachment will be a field with a link. When they click on that link, they will get a zipped download of the raw time series data attachment if it was attached. All of this is dependent upon people submitting the data the right way. The team has some activity codes they are suggesting folks use to help in discovery. They also recommend flagging any time series data sets with a particular project to help with data discovery as well.

Q: Are there any advances/updates to WQX handling of continuous datasets?
A: The team has provided a template for summarizing the continuous time series data. It's a little different than some of the other templates in that they've put examples of various ways to summarize the data, such as 30 days, 45 days, daily max, etc. It's up to them in how they want to summarize the data. Depending on what the program is using, they should submit the summary statistics that way. The data will be there for the submitter and anyone who is interested in seeing the program is using their data. The data is being submitted based on how the program is using the data. The team will be posting additional guidance for continuous data.

Q: How many API hits does WQX get in a month or a day?
A: The API for submissions on WQX only has a couple of users right now. They are probably making daily submissions—as they receive data into their systems, they push that data to WQX fairly quickly. On the outbound side through the portal, we get several thousand API hits a day—this gets more usage than submission. The United States Geological Survey (USGS) has the total API hits on the Water Quality portal.

Q: In some cases, people are forced to change their data to fit a data model to make it easier to take the data in. Recently there's been discussion about people using AI to circumvent that transformation to make the connection, like data mapping. Is anyone in WQX looking into this?
A: This is something the team has talked about and are keeping an eye on. It's possible to develop a robust vocabulary and train a system like that. The implementation would be tricky and the devil would be in the details to ensure the connections are correct.

Q: What platforms/technology are used for WQX data warehouse?
A: The team is modernizing systems on USGS's system. They are building out new query interfaces and graphs. The technology is changing but the main warehouse is PostgreSQL.

Q: Is the WQX framework led by a single regulatory ontology? Is there more than one regulatory program with different characterizations of water "stuff" all being submitted to WQX or is there a single ontology for all the water stuff that leads to the WQX model and exchange schema? Is the ability to

manipulate data in WQX for reformatting to meet transmission to accommodate database differences between systems under a single ontology, or is it accommodating different program ontologies? (Ontology is the way we define things in a world view so each regulatory program could have various/conflicting definitions versus simple labeling differences.) Does WQX data have to be transformed during transmission to any internal EPA systems?

A: WQX is currently the steward of over 20,000 characteristics that people can register and submit. One of the things the team can't do is be the arbitrator of the list and determine what's the right name for everything. The team looks for other systems that are defining that ontology. For characteristics and substances, they use EPA's Substance Registry Services (SRS), which has the final name for physical constitutes and synonyms that others may use. They use their services to maintain and determine what is correct. Another example is biological taxonomy. They are not the arbitrators for that taxonomy—they look to other systems for the right names. They look to the ontology leaders and follow them.

Q: Once data gets to WQX, does it get transformed to EPA systems for specific use cases?
A: Yes, for instance, EPA's Office of Enforcement and Compliance Assurance (OECA) absorbs all the WQX data to help them build systems to analyze polluters. They may make changes to the data. The WQX team does not make changes—the data owners are responsible for their data. They are not in the business of transforming the data, but once a system takes the data from WQX, they can transform it.

Q: Is personally identifiable information (PII) an issue in WQX data and if so, how do you manage and protect the data?
A: We do not collect PII other than the submitter's information, but that information is not sent to the warehouse and is not publicly available. We do not have any fields that are collecting PII.

Q: In New York (NY) we have some policies on protecting environmental data with residential properties, which is along the same lines as PII data. Does WQX have any issues like this?
A: We sometimes hear concerns about property water quality data, but nothing has been raised to a high-level of concern. It is something that could be a concern when looking at water quality, but we haven't had any issues. This is more of an impact on flood plains and if a property is a threat from water resources, and WQX doesn't have that type of information.

Q: How do citizens promote/encourage more reporting to WQX for How's My Waterway so people can get updates for their neighborhood?
A: Some small non-government organizations (NGOs) do monitoring in neighborhoods. These smaller organizations may need support and can reach out to the helpdesk. If they haven't put their data in WQX before, they may not be aware of some of the benefits of doing so and how to use the data from the water quality portal and How's My Waterway. The helpdesk can help them see the benefits of submitting the data and resources available to them. It's a community of support for them.

Q: Does WQX handle nearshore/intertidal data?

A: Yes, WQX can take data from anywhere, across the globe. It has a fair number of data, even open ocean data. WQX can take data from wells, streams, the tap, lakes, groundwater and the ocean.

Q: Where WQX leverages curated information sets (such as SRS), what were the technical challenges you encountered where users may find reference set mapping that appeared in WQX but needed to be communicated to the reference stewards? Do you have any advice for other systems/agencies for coordinating issue resolution?
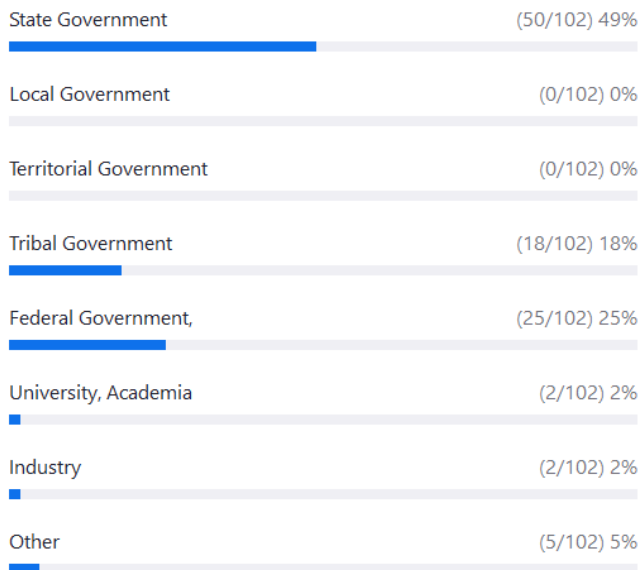
A: The WQX team does not necessarily steward the stewards. When WQX does receive feedback that one of the registry systems (like SRS) may have an edit, they tend to recommend that the user communicates that directly to the system steward. The WQX team generally does not have the background/SME to argue one way or the other, so really rely on our User Community to be the experts, and in turn communicate any discrepancies they find.

Poll Questions

The following are the results from the poll questions during the meeting.

1. What type of organization do you represent?   (Single Choice)

102/102 (100%) answered

State Government	(50/102) 49%
Local Government	(0/102) 0%
Territorial Government	(0/102) 0%
Tribal Government	(18/102) 18%
Federal Government,	(25/102) 25%
University, Academia	(2/102) 2%
Industry	(2/102) 2%
Other	(5/102) 5%

2. From the list of possible workshops or training sessions, please select topics that most interest you: (Multiple Choice)

74/102 (72%) answered

Data Management Systems   (60/74) 81%

Agile Project Management   (15/74) 20%

Data Analytics Tools   (45/74) 61%

GIS and Geospatial Analysis   (30/74) 41%

Digital Storytelling, Communication   (20/74) 27%

API Strategy, Development & Management   (26/74) 35%

Quality Assurance Project Plans   (23/74) 31%

Cybersecurity   (18/74) 24%

User Experience, User-centric Design   (21/74) 28%

Other   (4/74) 5%