

Exchange Network Forum

Thursday, May 9, 2024

1:00 – 2:00 PM ET

Welcome

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

- Alex O'Neill welcomed participants to the meeting and noted that the format is designed to be an open meeting to encourage interaction between participants.
- The meeting materials, including the summaries from the Forum meetings are available on the [Exchange Network website](#).
- Alex reviewed the agenda:
 - Meet Your EN Coordinator: Region 1
 - FY24 EN Grant Solicitation Notice
 - Environmental Information & Innovation (E2i) National Meeting
 - Data Exchanges 101
 - Open Forum

Meet Your EN Coordinator: Region 1

Steven Zailskas and Ed Kim, EPA Region 7

- Steven Zailskas and Ed Kim introduced themselves:
 - Steven is an IT specialist and has been with EPA for four years and the Regional Exchange Network Coordinator for four years.
 - Ed Kim is also an IT specialist and has been with EPA for over 35 years and worked in the RENC role for 10 years.
- Region 1 currently has 13 open grants and 85 total grants since 2002. Region 1 has about three to four EN grants awarded per year. The grants are three years in duration. They have three to four grants that rollover each year and three to four grants that close out each year.
- The top three EN Project opportunities in Region 1 from 2019-2023 include:
 - Water Quality Exchange
 - Open Data Projects
 - Individual Capacity Building
- 87 percent of Region 1 EN grants are awarded to states and 12 percent to Tribes. In 2023, three grants were awarded to the Tribes and one to a state university. The Tribal participation is encouraging.

- Steven and Ed introduced Jill Carr who is a coastal data scientist from MassBay National Estuary Partnership/UMass Boston. They are one of 28 estuary programs from across the country.
 - The National Estuary Programs are required by the Clean Water Act to report on the water quality of estuaries. Local agencies work with a robust community of scientists through volunteer organizations to sample the waters. The MassBay National Estuary Partnership develops tools to assist the community scientists with collecting and sharing the data.
- MassBay received an EN grant in 2021 for “Leveraging Citizen Science Data to Improve Massachusetts Waters.”
 - Through the project, MassBay developed MassWaterR, an open-source R-based package for conducting QA/QC, analysis, and WQX upload of community-based water quality data for fresh and marine discrete surface water quality monitoring data.
 - Roll-out of the package included trainings, support and a Community of Practice forum. The tool has been used by dozens of groups and agencies resulting in new data in the WQX.
 - The package provides a streamlined and repeatable means of:
 - Screening data for QC
 - Preparing QC summary report
 - Creating graphics for analysis
 - Formatting data for upload to WQX.
 - No R experience is necessary to use the package.
- Jill Carr provided a demonstration of the package, which can be viewed in the meeting recording.
- The team provides training events, online training recordings and a web page with a user guide. They have a web-based community of practice where developers and other users answer questions.
- The MassWaterR GitHub website and user guide can be found at <https://massbays-tech.github.io/MassWaterR/articles/MassWaterR.html>.
- They also have a community of practice with a help forum that can be found at <https://massbays.discourse.group/login>
- For more information, contact Jill Carr at Jillian.Carr@umb.edu

FY 24 EN Grant Solicitation Notice

- The EN Grant Solicitation Notice is live and the deadline has been extended to May 23, 2024.
- Since the last forum there have been three webinars. Recordings are available and will be posted on the EN website.

- For more information about the EN SN, reach out to the EN Grant Program Leads:
 - Erin McGown (mcgown.erin@epa.gov)
 - Aimee Awad (awad.aimee@epa.gov)

Environmental Information & Innovation (E2i) National Meeting

- The E2i National meeting will be in Kansas City, MO, September 17-19, 2024.
- Registration will open in the coming weeks.
- Abstracts have been submitted for sessions, and the planning group is working on the agenda.

Data Flows 101

Dwane Young, EPA Office of Information Management, Office of Mission Support

- Dwane Young provided an overview of data flows 101.
- Back when data was originally exchanged, someone had to go on site to load STORET on the local servers and it just stored the data. Today we exchange data. Through standard data formats and standard terminologies, developers can build capacity to exchange, share and analyze data.
- Dwane asked participants what their experience is with using others' data? The following are the responses:
 - Trying to understand what field names mean so tables, datasets and boundary data can be joined.
 - Corraling all the data and harmonizing terms is hard.
 - Some data has “garbage” in it if the owner doesn't clean it.
 - A messy mess. 10 years of inconsistent methodologies, incomprehensive field names and values.
 - Different nutrient names.
- Dwane's presentation focused on the philosophies of how the Exchange Network works. Not everyone has been exchanging data for 20 years, so the presentation included the basic idea of how to set up data to exchange data with another entity. The presentation represented Dwane's philosophy and isn't the only way data can be exchanged.
- The EN exists to “...provide a way to share data in a standardized, open format, make data exchange more efficient and less costly. The Network empowers Partners to overcome the hurdles to effective data exchange so they can make better decisions—faster.” (<https://exchangenetwork.net/about/why-we-exist/>)
- EPA, states and Tribes often need to operate on common sets of data—a common operating framework and a way for us to understand the universe in the same way so we can share data with one another. The EN helps us build that common operating framework so we are working off the same data.

- There are four principles behind sharing data:
 - Data Model/Standards – common concepts that you share. This is the most important thing that needs to be done and it needs to be done first.
 - Common Terms – we need to use the same language. If we call the same thing different terms, it will be hard to pull the data together.
 - Communication Protocol – for computers to talk to each other, we need to agree on what the input parameters are.
 - Shared Authentication – a common way to authenticate you as a partner so you can access data that isn't accessible, and this goes the other way.
- The world is built on standards. When people agree to do things in a common way it opens up many opportunities for everybody to do things better.
 - Technology standards: like Bluetooth and Wi-Fi.
 - Data Standards: like financial exchanges.
 - The IRS has open standards for electronically filing your taxes and enabling a better user experience.
 - Banking: Open Financial Exchange standards enable you to access your banking information from many 3rd-party applications.
 - Government Standards: like the Exchange Network.
- When using a standards-based approach:
 - It is not dependent on a particular technology.
 - The data sharing model is based on the science instead of a technology.
 - Partners don't need to run EPA's software, but rather just need to map their data systems to the standard.
 - Partners can use any software they want, such as R, Python, SQL Server, etc.
 - Focus on the 'core' data elements needed for someone to communicate the data.
 - We don't need all the data in the universe. The WQX model is larger, but it could be larger if you consider all the data elements to describe water quality. WQX has distilled it down to the core data elements needed to communicate the data.
- A data standard or model is not talking about a database. Databases are the implementation of a data standard. Databases think about tables and the relationships of data. We are not talking about databases, but instead what are the concepts that describe the data you have.
 - We are looking at the business or science and the concepts of that business.
- Items to focus on:
 - Concepts: What are the concepts that define your data?
 - Attributes: What are the attributes of those concepts?
 - Relationships: How do those concepts relate with each other?

- Rules: What are the rules that define which attributes are required and under which conditions and which are not?
- Technical considerations:
 - Historically the EN has used xml and xml schemas to define our data standards (.xsd files)—don't feel constrained to do this. This is a little outdated.
 - We've used Excel, PowerPoint, Visio and XLM Spy to communicate around those standards.
 - Dwane has developed five data flows in his career, and he almost always starts with pencil and paper to draw the relationships.
- To create a data exchange with EPA, a state or Tribal partner, the following are the steps:
 - Step 1: What are the concepts?
 - Leverage subject matter experts—work with biologists, water quality monitoring experts, etc.
 - Look at governing regulations and guidance that will drive the standards.
 - Talk to practitioners.
 - Create a 'mental model' of how this 'idea' exists in the natural world.
 - Describe how to sample, the parameters and the concepts and metadata.
 - Step 2: Draw those concepts out on paper. And ask:
 - What are the attributes of these concepts? What attributes are common across concepts?
 - Step 3: "What are the relationships between concepts?"
 - This can at times be the most difficult. Is one concept the 'parent' of another concept. Are they 'siblings'? How many iterations of one concept can be related with one iteration of another? What's the 'root' concept?
 - Step 4: "What are the rules for concepts and attributes?"
 - Which ones are required?
 - Are there special conditions under which they would or wouldn't be required?
 - Step 5: Engage
 - Data standards require broad engagement. They cannot be created within a vacuum, nor by the few or the one. No one person knows everything. There are different perspectives.
 - Form a user group of actual practitioners and your user community.
 - Be open to modify/change.
 - This process will also help bring them along with the 'change' that you're proposing.
 - Don't underestimate the power of face-to-face interactions.
- Other Considerations

- Starting with Data Standards is a good thing, but...
 - Data standards are basically abstractions and/or a model of how something exists in the real world. It's really easy to get it wrong.
 - You may not find out you got it wrong until you actually go to use the standard.
 - When building your application, you may find requirements that the data model didn't consider.
- Once your community adopts your standard you may find yourself 'locked in' with little ability to change.
- Common Terminology is the twin to this principle. One without the other cannot survive.
- Within the EN, exchanges are documented as 'Data Exchange Templates (DETs)' – here's an example for RCRAInfo:
 - https://www.exchangenetwork.net/schema/RCRA/5/RCRA_DET_v5.13.xls
- Common Terms
 - What are the terms or domains that define your universe?
 - Once you start sharing data with others, it's important to have a universe of agreed upon terms or 'allowed values.'
 - Example: are these all the same thing?
 - Dissolved Oxygen
 - Oxygen, Dissolved
 - Dissolved Oxygen (DO)
 - DO
 - EPA's System of Registries can be a great place to start:
 - https://sor.epa.gov/sor_internet/registry/sysofreg/home/overview/home.do
- How do you interact with data?
 - API stands for 'Application Program Interface'. It's really just a fancy way of saying how does your application interact with the world.
 - APIs expose the functionality of your application so that others can build upon what you've already done.
 - We interact with APIs and Web Services all the time.
- Inputs and Outputs
 - Most APIs have some sort of 'Input' and produce some sort of 'Output.'
 - For example, a weather app has an 'Input' of a zip code and the 'Output' of weather data.
 - The App turns that 'data' into 'information.'
- EN Data Flow Perspective: Defined 'Standard Node' protocols.
 - Every single node on the EN can respond to a core suite of protocols.
 - These protocols provide the 'backbone' of an EN Data Flow.

- Authenticate
 - Submit
 - Query
 - Solicit
 - Get Status
 - Notify
 - Download
- Common Authentication—we can't all have our own authentication protocol. We must have a unified approach. This is currently done through the EN authentication. CDX is moving to Login.gov.
 - As we think about data exchange, we need to think about...
 - Data standards and common terminology.
 - Common APIs are still critical, but we won't be implementing these the same ways as in the past.
 - As we think about the future of the EN – which of these principles should still apply? If they do still apply, what do they look like in the future?
 - As we move to the E2i conference, what does the future of the EN look like?
 - We will do a similar session or workshop at the E2i conference and will be providing more training on this topic.