

# Water Data Infrastructure: A U.S. Perspective

Dwane Young,

U.S. Environmental Protection Agency

Office of Water:

Water Data Integration Branch

[Young.dwane@epa.gov](mailto:Young.dwane@epa.gov)



# Water Data Infrastructure

- Outline

- Big Picture: Telling a story about water
- Open Water Data: Concepts
- Integrating across EPA systems (internal integration)
- Enabling Integration beyond EPA
- System Design Principles
- Deep Dive – Water Monitoring: Samples and Sensors
- Deep Dive – Hydrologic Networks
- Use Case: Characterization of Pre-event conditions – example of where Open Linked Data could make a difference
- Pulling the Pieces Together: Telling the Water Story





# What can you tell me about my water?

Is it safe to drink?

Is there enough water?

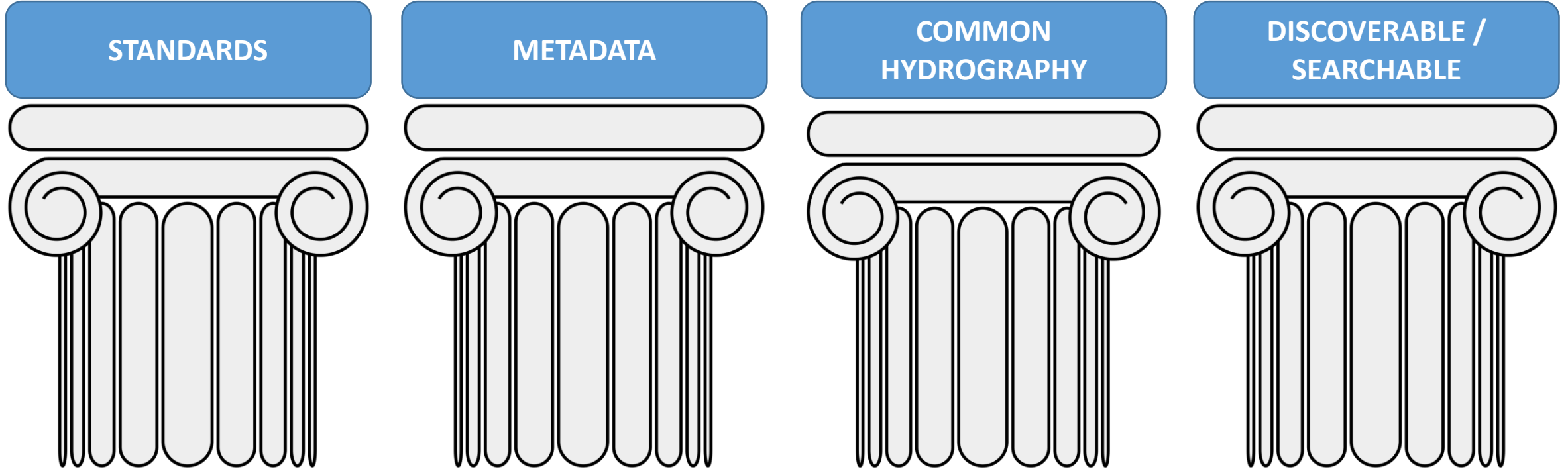
Can I swim in it?

Is it polluted?

If it is polluted, what are you doing about it?

If it isn't polluted, what are you doing to protect it?

What can I do to help?



# FOUR PILLARS OF OPEN WATER DATA

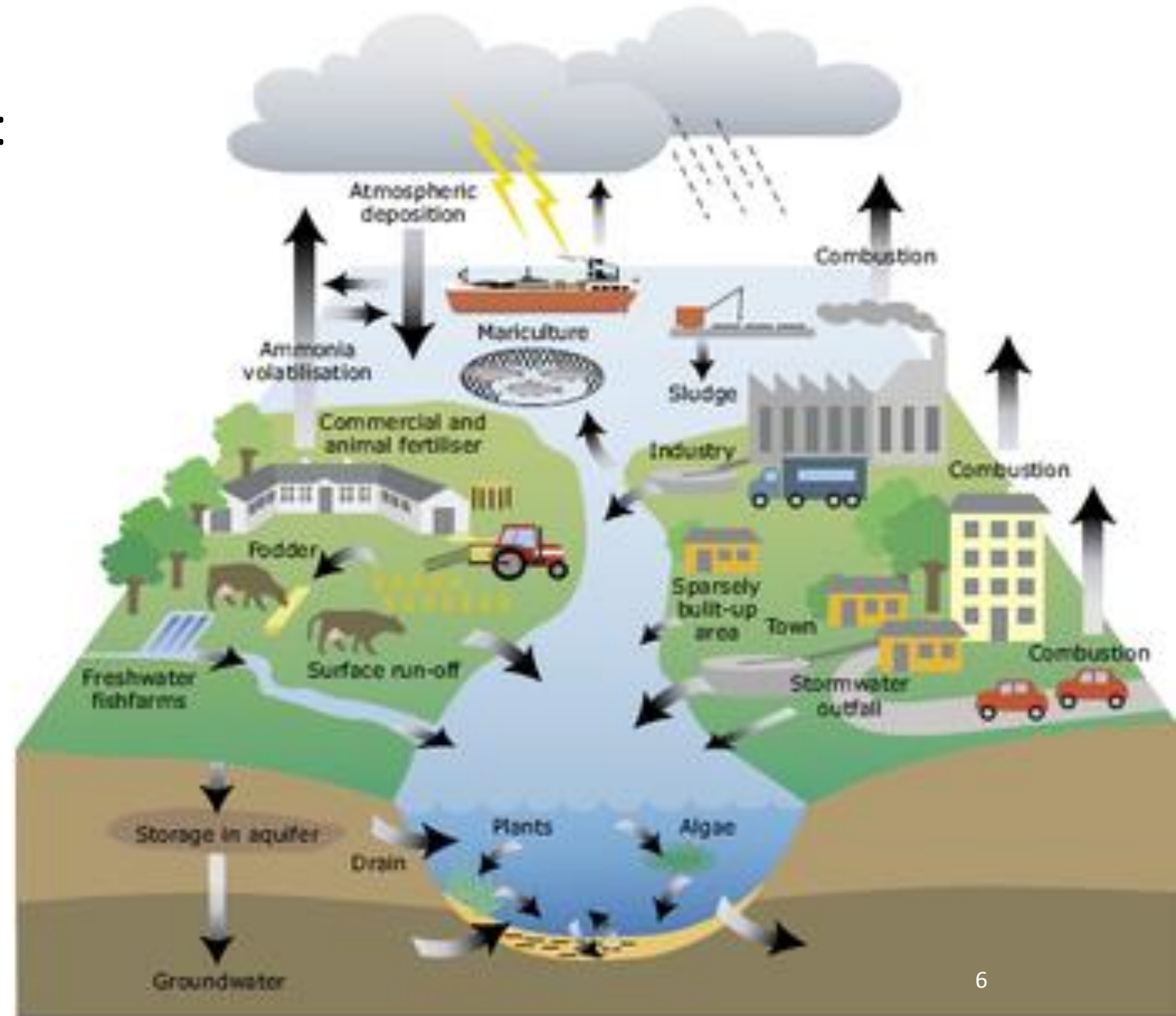


# Going Beyond Integrating Data



# Value of Open Water Data: Better Modeling

- Models need data, a lot of data
- Data can be used as input for models, and can provide necessary ground truth
- Imagine being able to run a model without having to first discover and then reformat the data needed to run the model; rather the model can grab the data as it needs it in real time





## Value of Open Water Data: Third-Party Applications

- Adopting and promoting standards sends a clear signal to the market that there is value in supporting those standards
- Enables the market to develop solutions for data storage, data sharing, and data visualization
- Open data allows those outside government to develop applications and tools that are valuable to them and meet their specific needs

# Value of Open Water Data: Quicker and Better Decisions



- Having open data decreases the amount of time needed to spend in discovering and standardizing data
- Reduces errors from hand-entering data values
- Allows for MORE data to be used (often can't use data that's hard to find or difficult to reformat)

**SAVED TIME = SAVED MONEY**



# How Do We Get There: Building Blocks

## Sampling Data

STANDARDS	✓
METADATA	✓
TIED TO HYDROGRAPHY	✓
DISCOVERABLE	✓

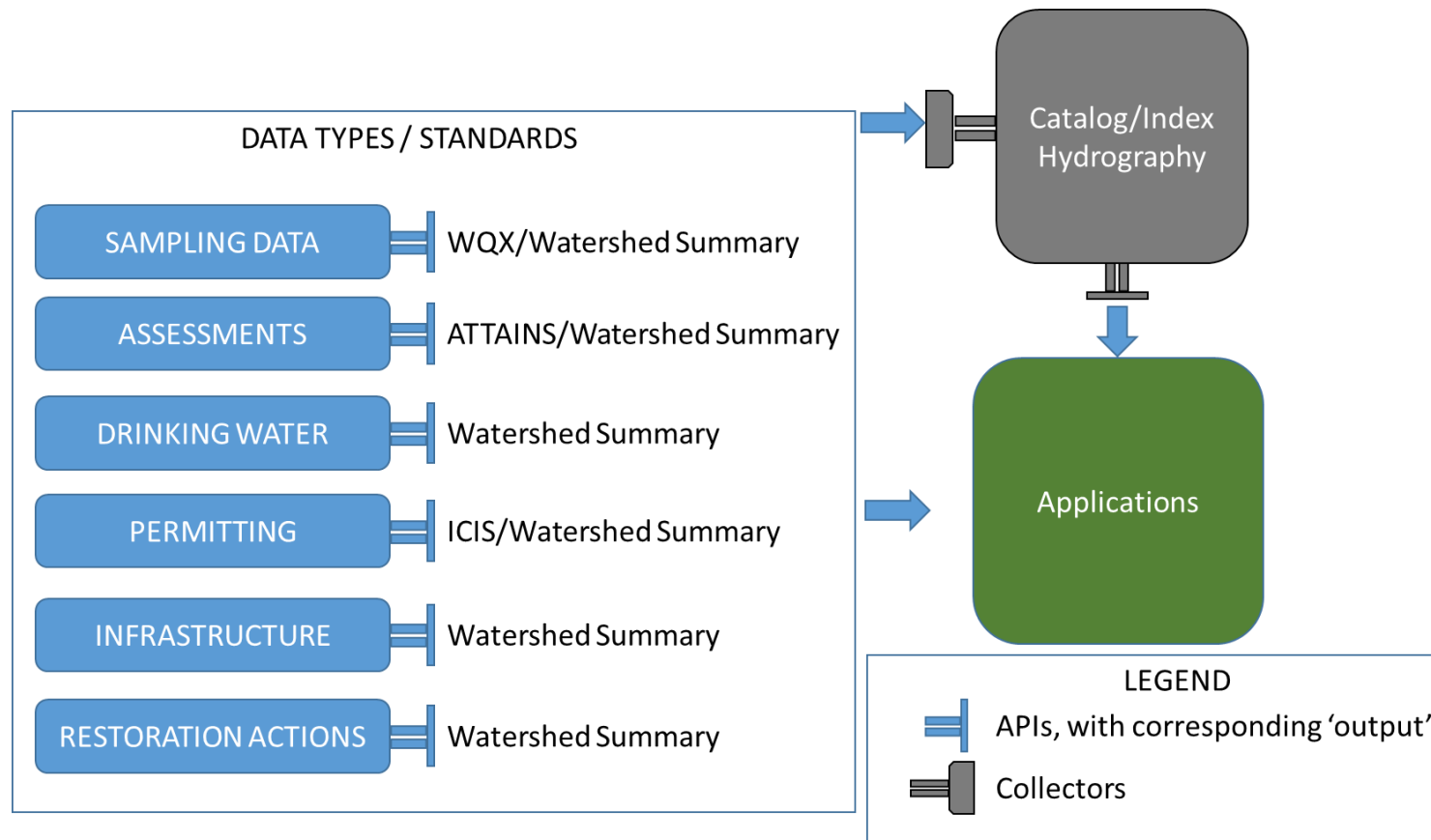
## Real Time Data

STANDARDS	✓
METADATA	GAP
TIED TO HYDROGRAPHY	SOME
DISCOVERABLE	SOME

## Other Attributes

STANDARDS	VARIES
METADATA	VARIES
TIED TO HYDROGRAPHY	VARIES
DISCOVERABLE	VARIES

COMMON HYDROGRAPHY

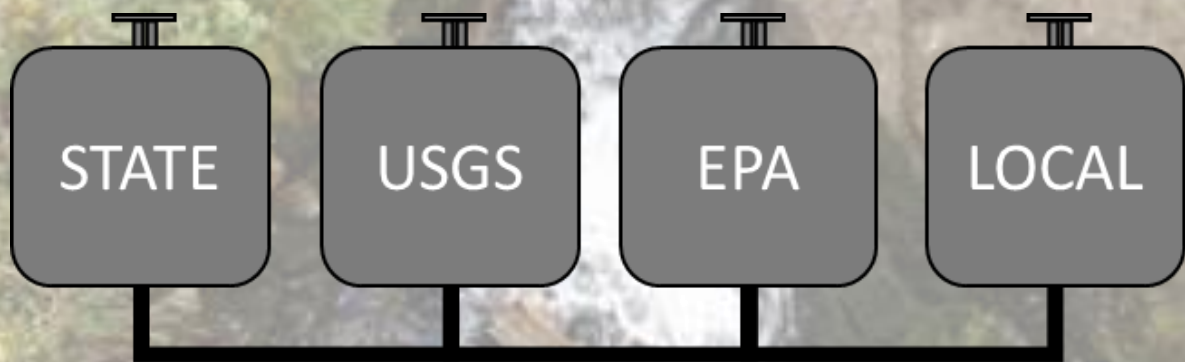


## Using Standards to Integrate

- Each system supports an API with a defined output
- Where common, generic outputs can be defined, those outputs are used
- Points of integration are also defined between systems
- All data are indexed for quick discovery as well as referenced to a common hydrography to enable advanced searches/discovery

# Connecting Outside Data

- Catalogs/Indexes allow for quick discovery of data across entities and programs
- A Common Hydrography enables search across entities
- Defined data standards enable data sharing across catalogs





# Principles of Integration

- No direct database access. All interaction is through an API (this is true for the system owners as well)
- Identify points of integration between systems to enable easy discovery and entry points across systems
- Data indexes (catalogs) allow quick discovery of data
- Data indexes can also facilitate common search functionality across systems
- Data are all connected to a common Hydrography



## Points of Integration:

- Assessed Water ⇔ Monitoring Locations
- Assessed Water ⇔ Permitted Facility
- Assessed Water ⇔ Restoration Actions
- Pollution Budgets ⇔ Permitted Facilities



# Deep Dive: Water Sampling – Water Quality Exchange (WQX)

---

- WQX is a ‘standards’ based approach for sharing water quality monitoring data
- WQX defines a common data model for communicating water quality data (sample data)
- Designed to be automated
- The structure of partner data systems don’t matter, so long as they can map to WQX
- WQX also provides a standard format for publishing data



# Some history on how EPA got here

---

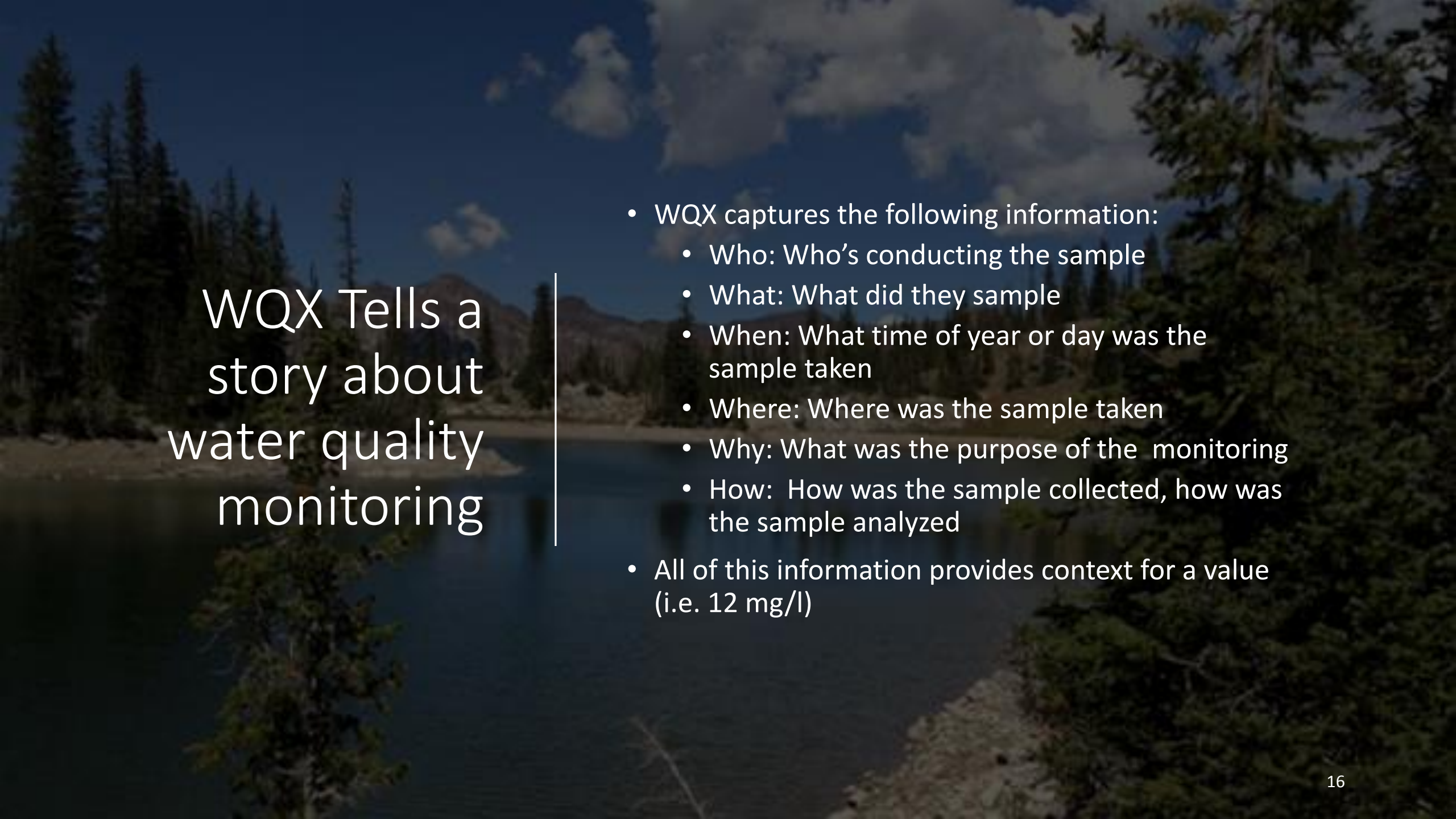
- From 1998-2007, EPA provided a distributed database (STORET) for partners to use to manage their data; partners could then send a 'copy' of their database to EPA
- Numerous issues:
  - Partners didn't buy-in to using the EPA database
  - "Keeping up with technology" was exceptionally difficult
  - Trying to build one system that met everybody's needs resulted in an overly complex system





# Why use a standards-based approach?

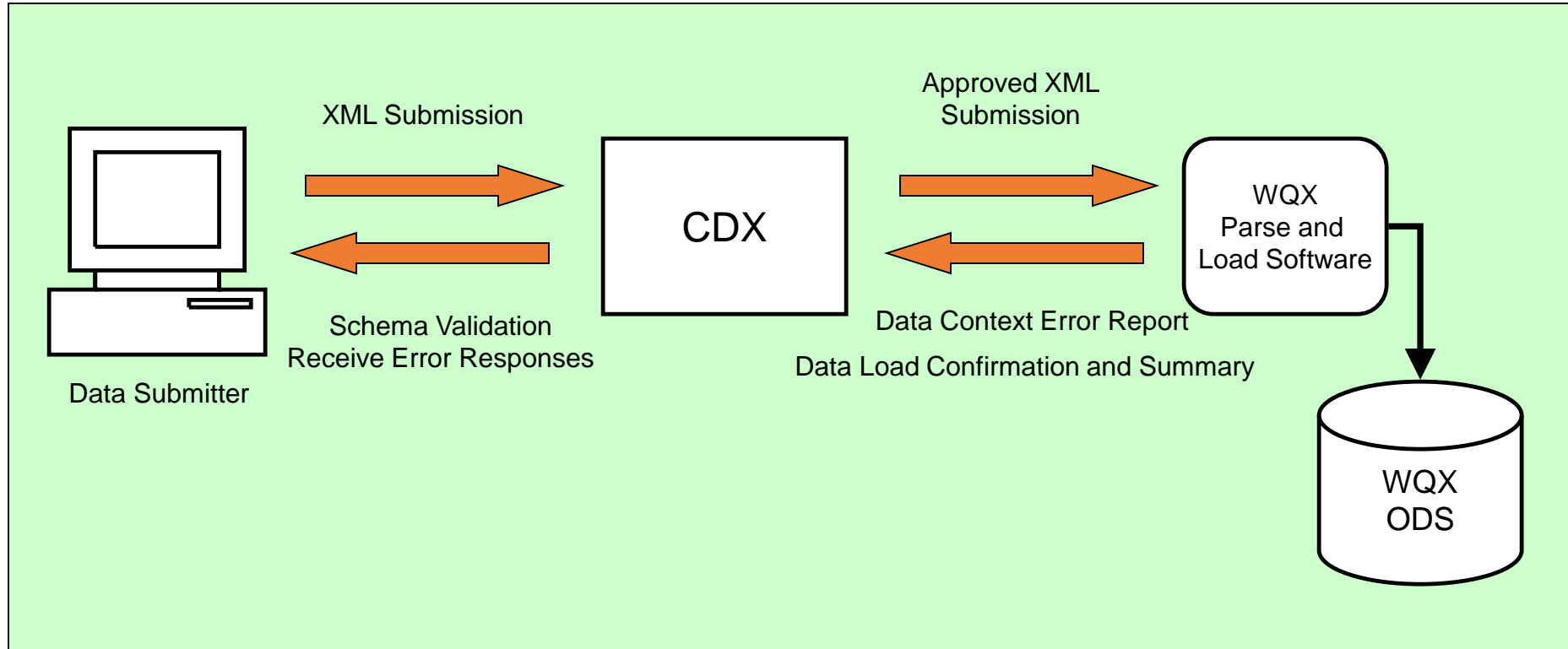
- WQX is not dependent on a particular technology
  - The data sharing model is based on the science of water quality monitoring, which doesn't change that much
- Partners don't need to run EPA's software, but rather just need to map their data systems to WQX
- EPA could focus on the 'core' data elements that were needed for someone to communicate water quality monitoring data



# WQX Tells a story about water quality monitoring

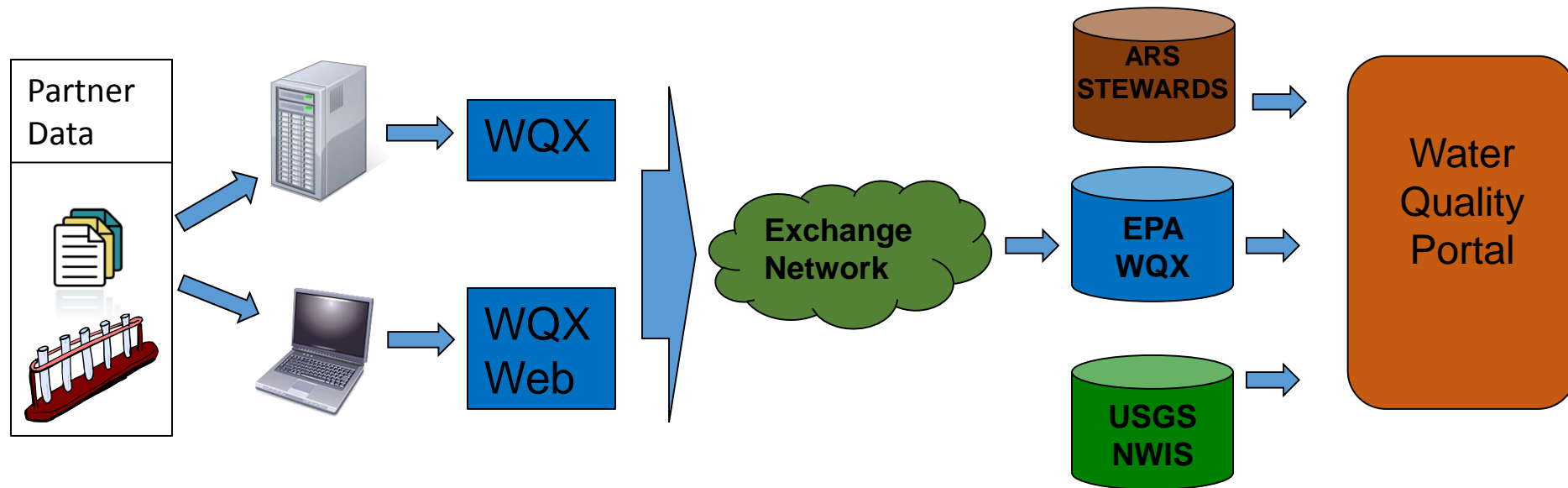
- WQX captures the following information:
  - Who: Who's conducting the sample
  - What: What did they sample
  - When: What time of year or day was the sample taken
  - Where: Where was the sample taken
  - Why: What was the purpose of the monitoring
  - How: How was the sample collected, how was the sample analyzed
- All of this information provides context for a value (i.e. 12 mg/l)

# HOW WQX WORKS





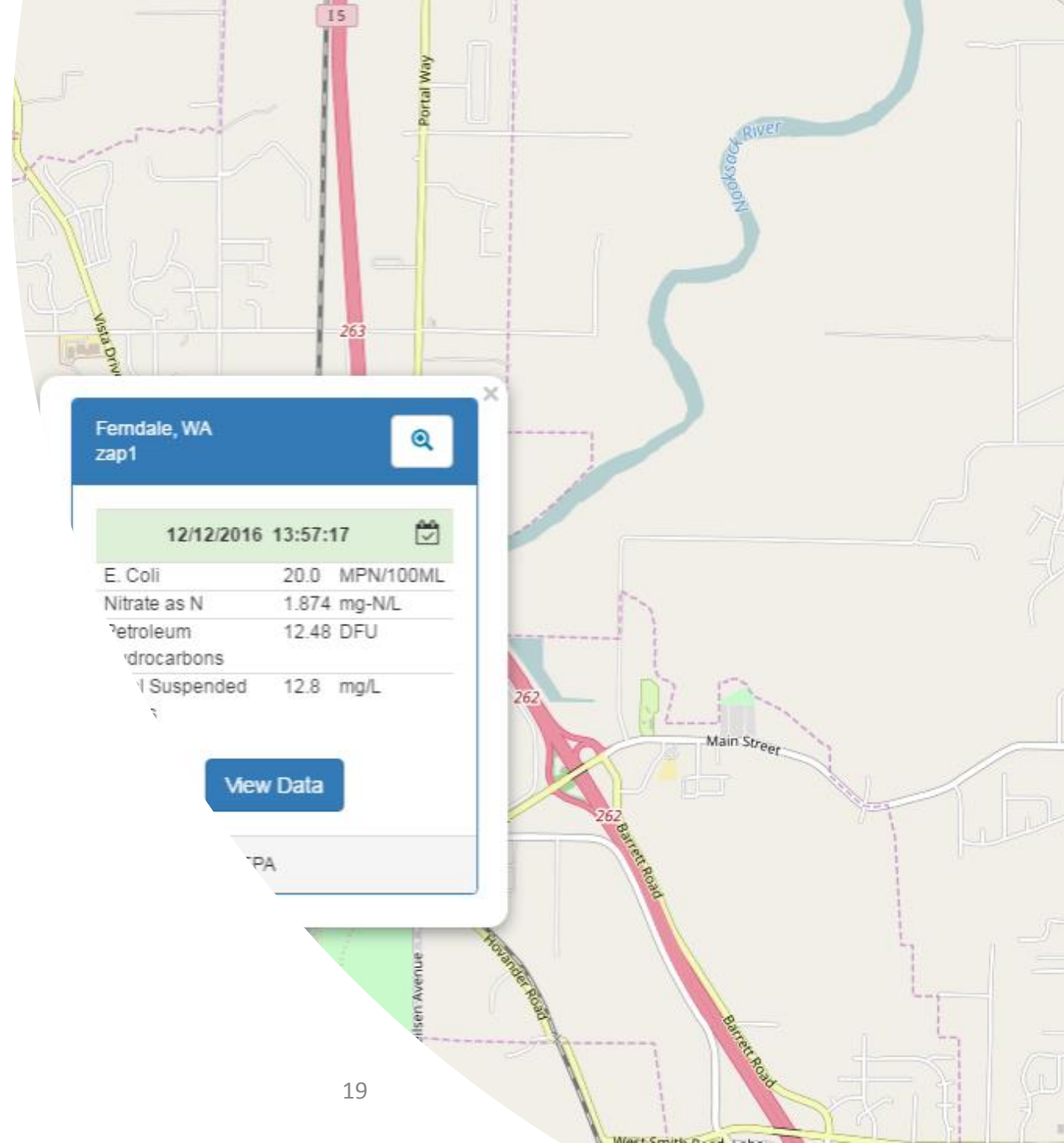
# The role of WQX in Data Sharing



For more information on the portal see: [www.waterqualitydata.us](http://www.waterqualitydata.us)

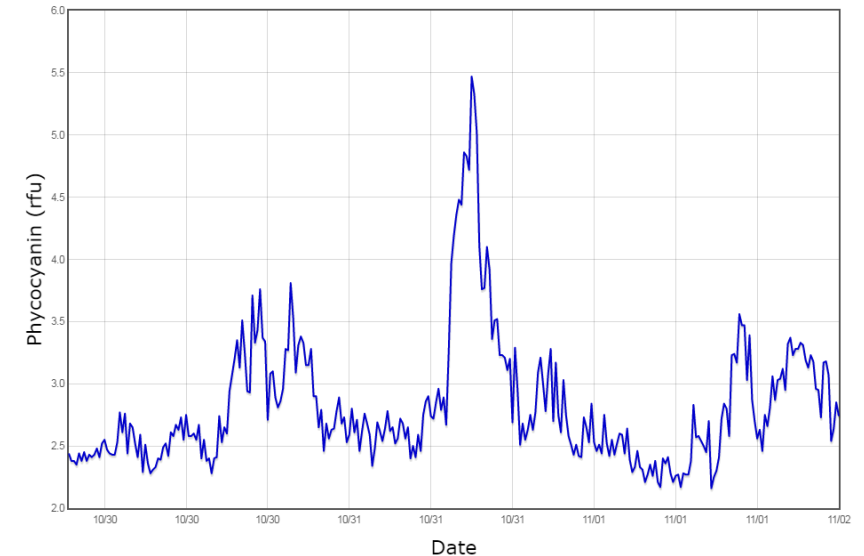
# Deep Dive: Real-time Data (Interoperable Watersheds Network)

- The Interoperable Watersheds Network was a demonstration project that focused on evaluating approaches to improve sensor data sharing
- It was based on knowledge gained from an recommendations report that EPA developed in 2014
- The project focused on addressing three major areas:
  - Data Standards
  - Metadata
  - System Architecture



# The Data Standards Problem

- We needed a common way to represent and communicate the data
- Standards for sensor data already exist, there was no need to create new standards
  - OGC Sensor Observation Service
  - OGC Water ML 2 and Sensor ML
- The Open Geospatial Consortium is an open-source, international standards setting body





# The Metadata Problem

- Needed a standard way to answer the following questions:
  - What data are available and for what parameters?
  - What data can I use?
  - What's the quality of the data?
- IWN had to develop standard ways to do this (no existing standard existed)
- Further work needs to be done in this area

zap1

- [nitrate\\*](#) (11/10/2016 - 02/13/2017)
- [oil\\*](#) (11/10/2016 - 02/13/2017)
- [total\\_suspended\\_solids\\*](#) (11/10/2016 - 02/13/2017)
- [e\\_coli\\*](#) (11/10/2016 - 02/13/2017)

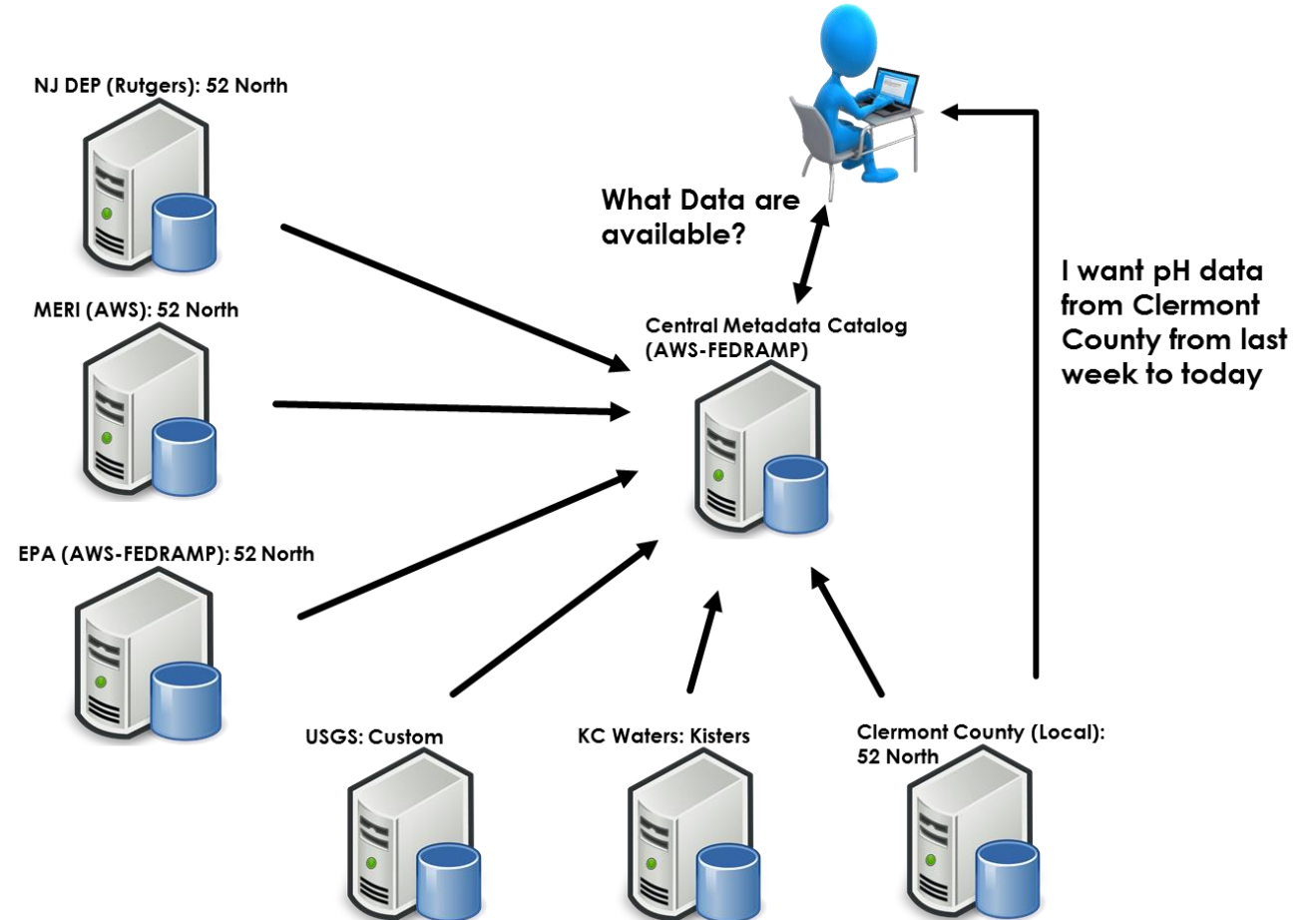




# The Architecture Problem

How do you solve the problem of multiple data providers with large amounts of data that have the potential to change every 3-15 minutes?

- Used a central catalog/index that references every data owner's assets with the corresponding metadata for each sensor
- Allowed for quick searching and discovery of available data
- This approach is similar to how Google allows you to search the internet
- Actual data comes from the partners systems in real-time



# Underlying Catalog Services IWN Defined

- **GetOrganizations:** Returns who is providing data with their endpoints
- **AvailableParameters:** What parameters can be queried
- **GetSensors:** Gets the sensor information and provides different methods for querying sensors (i.e. by county, by HUC, by buffer, by a bounding box, upstream, and downstream)
- **GetSensorParameters:** Gets parameters for a sensor, including the period of record
- **GetOrganizationParameters:** Gets the parameters for an entire Organization

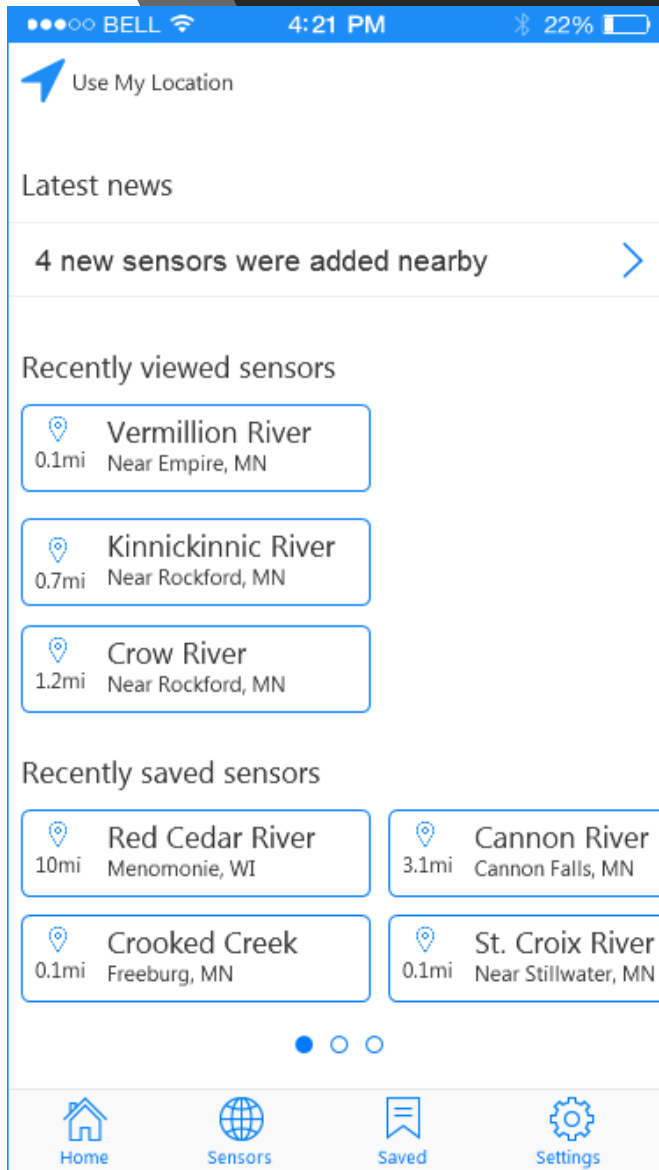


## OGC Defined Services (SOS 2)

**Each endpoint supports Sensor Observation Service in XML format (WaterML2)**

- **GetObservation**: Gets the most recent data or retrieves a collection of data
- **GetCapabilities**: Getting all the metadata from the endpoint
- **DescribeSensor**: Describes the sensor

# IWN's Open Architecture Allows Other Possibilities



- IWN is built using an open architecture, meaning that all the functionality you see in the demonstration tool is also available as a corresponding **Web Service** or **Application Program Interface (API)**
- Enables for other apps to be developed (like mobile apps)
- Also allows for other third-party applications (like Excel) to be able to directly interact with the data without having to go to a website and 'download' the data

A map of the Northeast United States, including parts of New York, New Jersey, Pennsylvania, and Delaware. The map is overlaid with numerous red location pins, each containing a white star icon, representing sensor locations. The pins are densely clustered in the New York City area and more sparsely distributed in the surrounding regions. Major cities like New York, Philadelphia, and Washington D.C. are labeled. The map also shows state boundaries and major highways.

# Outcomes

- CURRENTS is a demonstration tool
- It contains data from:
  - 8 Partners:
    - EPA Region 1: 2 Sensors
    - EPA Region 7: 18 Sensors
    - EPA Region 10: 1 Sensor
    - EPA ORD: 3 Sensors
    - NJ DEP: 106 Sensors
    - NJ Meadowlands: 3 Sensors
    - Clermont County, OH: 4 Sensors
    - USGS: 15,541 Sensors (nationwide coverage)

Organizations ▼

New Jersey DEP ✕

Results

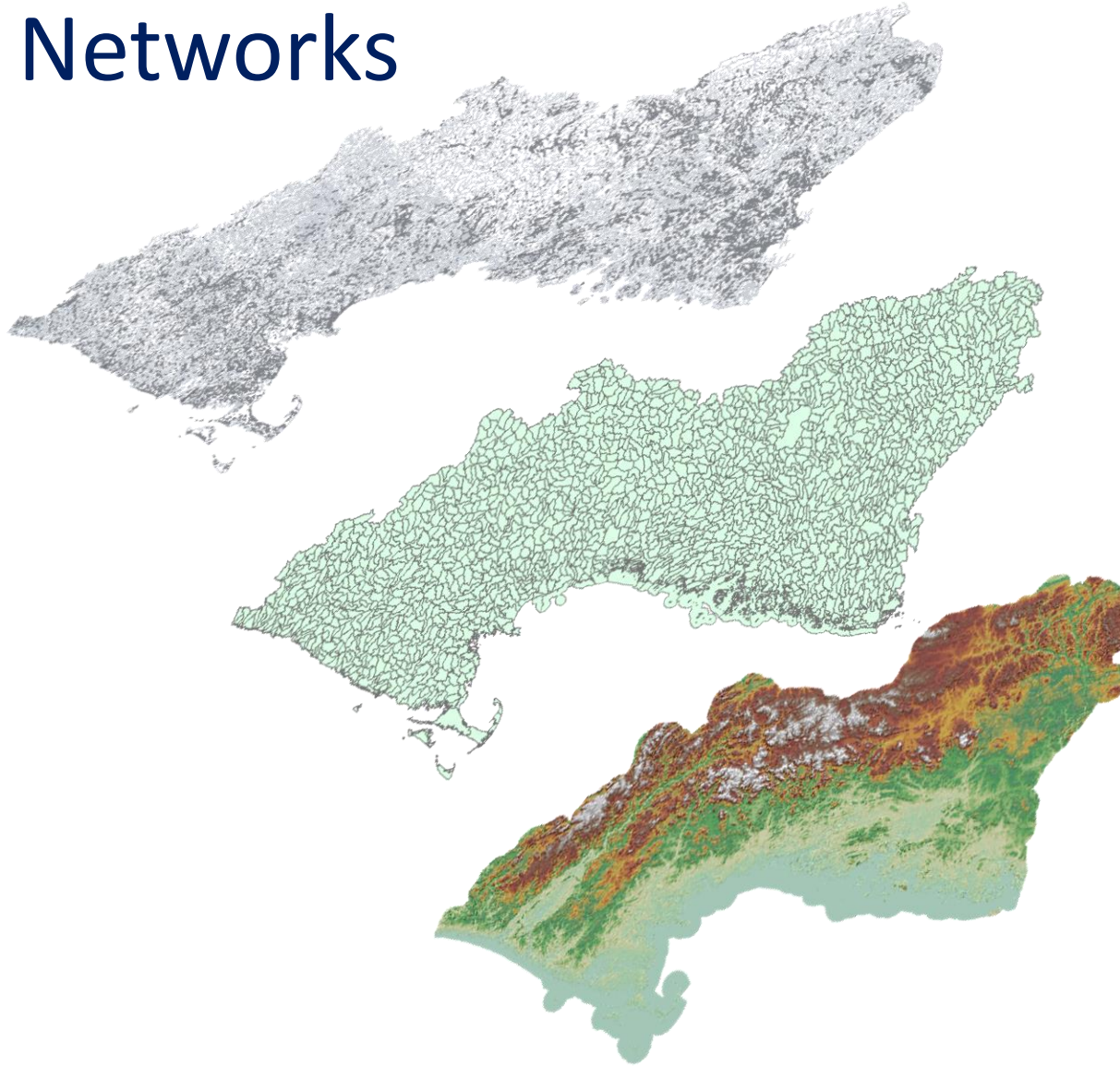
bfbm000237 ✕

an0605 ✕

an0742 ✕



# Deep Dive: Hydrologic Networks



National  
Hydrography  
Dataset (NHD)  
(1:100K and better)

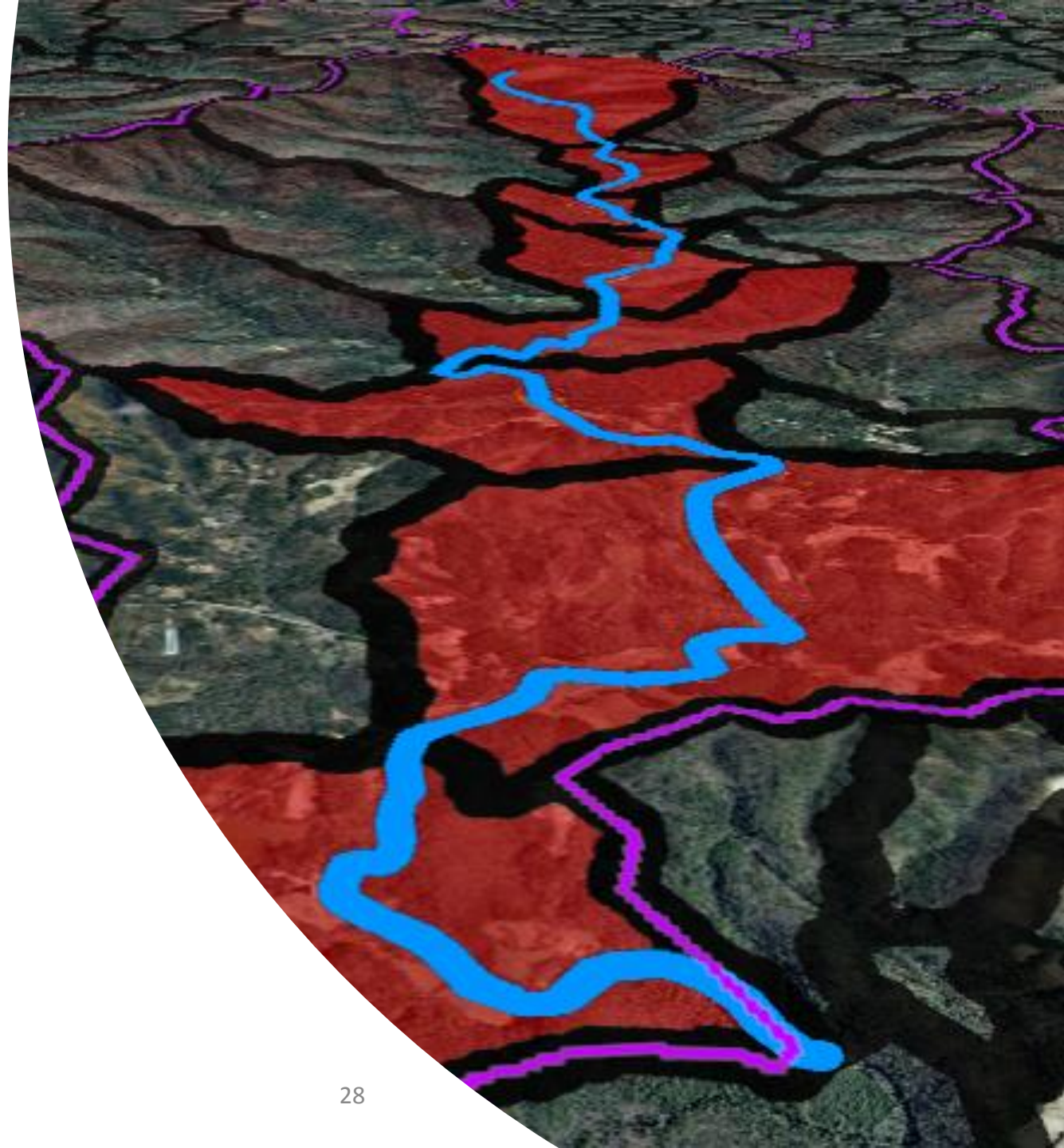
Watershed  
Boundary  
Dataset (WBD)  
(1:24K)

National  
Elevation  
Dataset (NED)  
(10 meter and better  
source resampled to 30  
meter)

# Catchment-based Indexing Approach

---

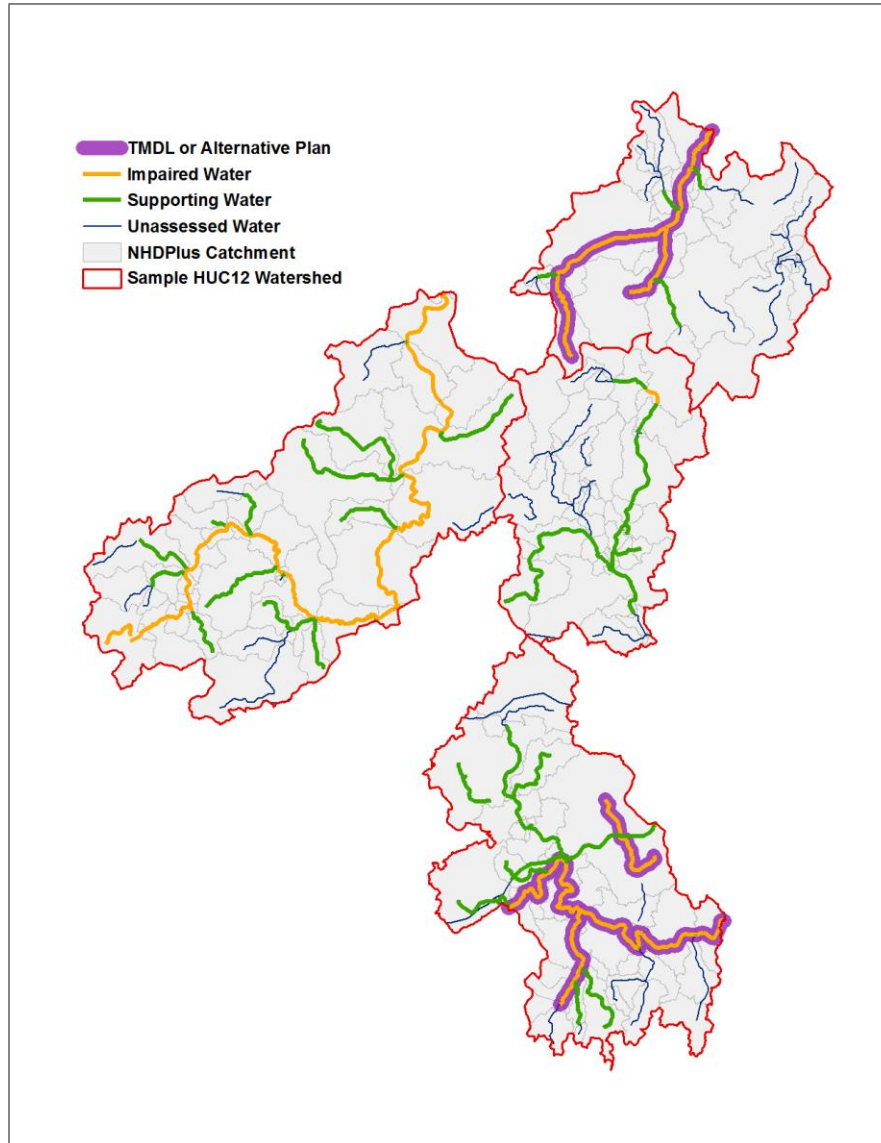
- Use the NHDPlus catchments as an abstraction for higher-resolution data and as a common unit for measures
- Use an automated process to develop correspondence between state-provided data and the NHDPlus catchments
- Resulted in a significant cost-savings for EPA and opened up a whole new world of communication opportunities – PARADIGM SHIFT



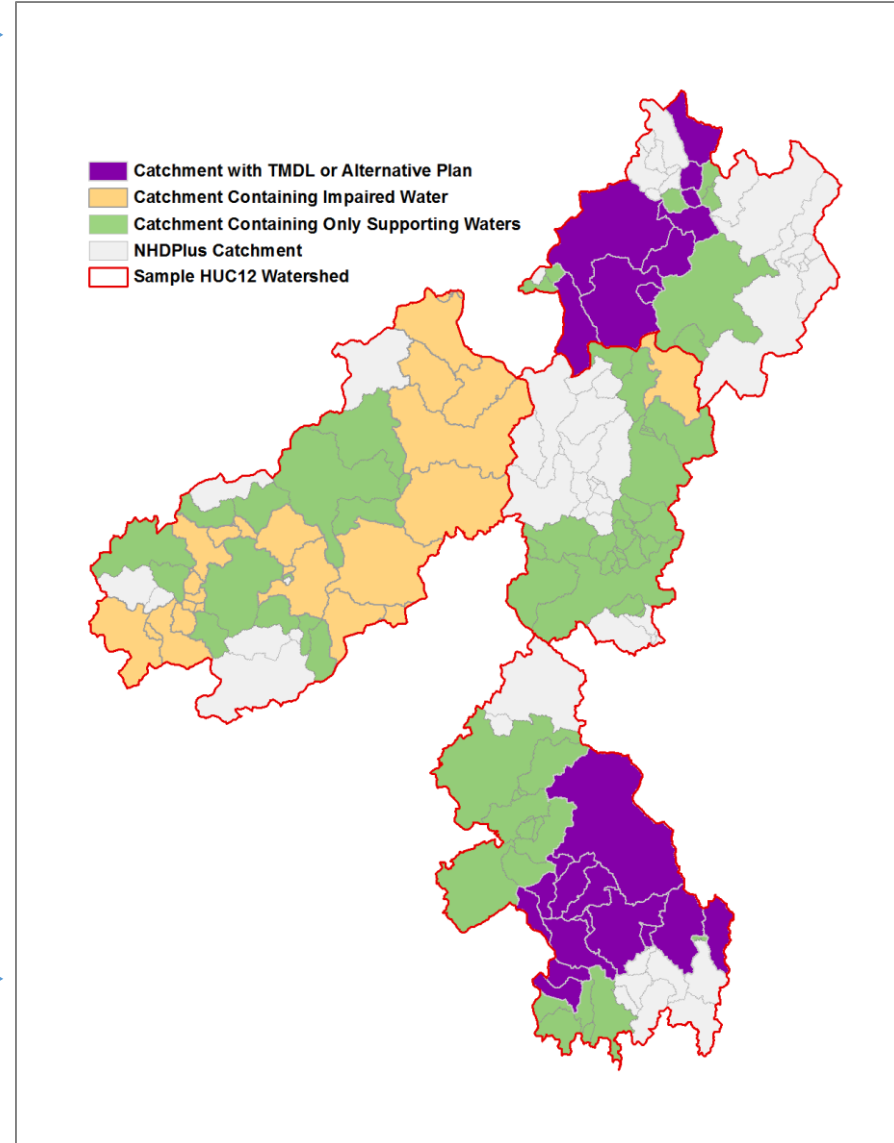


# Mapping State GIS Data to the Catchments

Receive GIS data from States

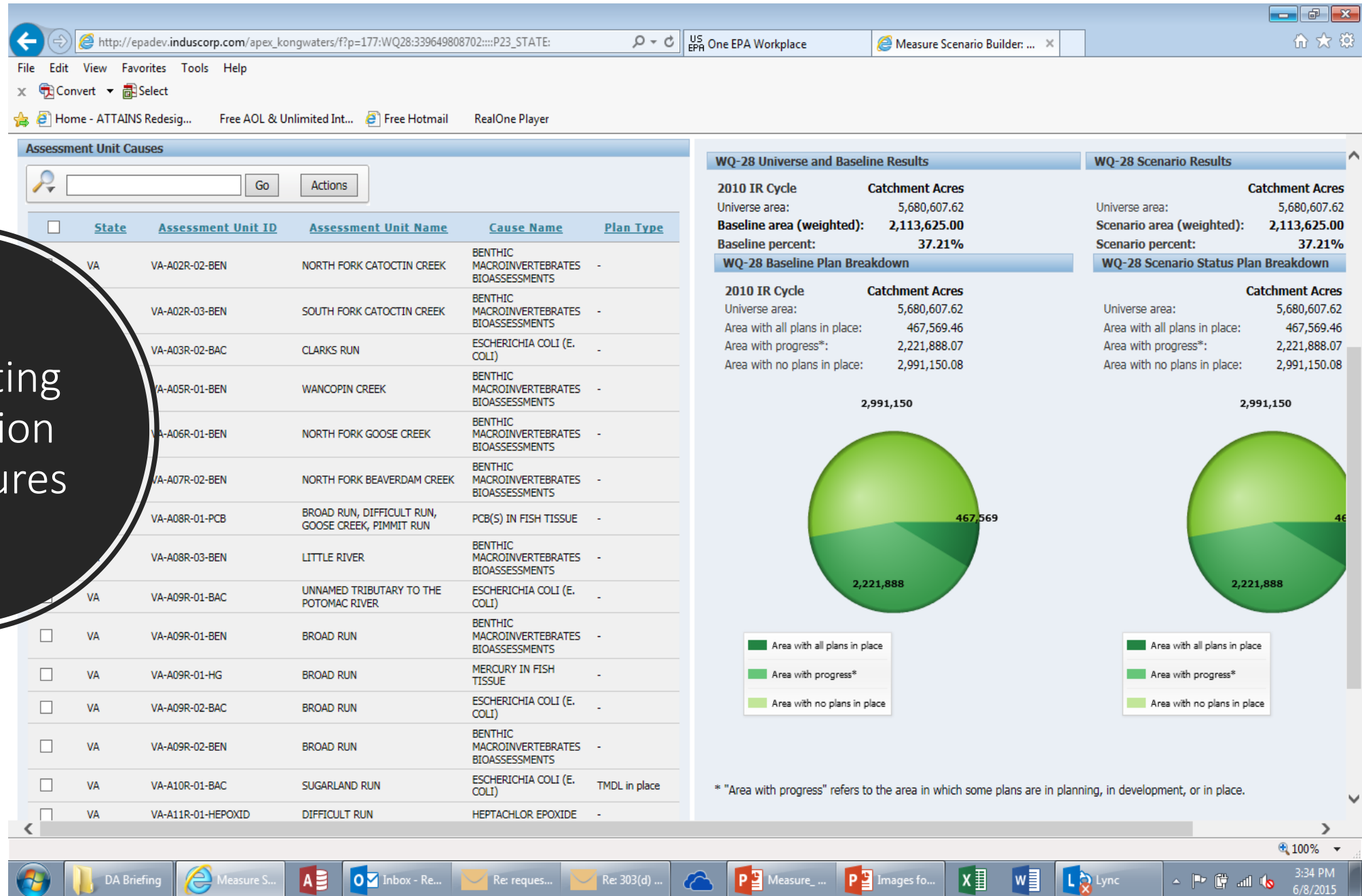


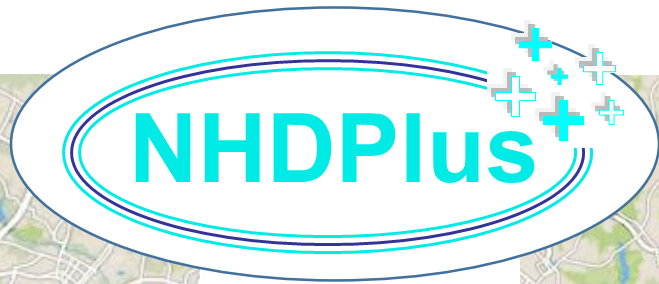
Translate to Catchments



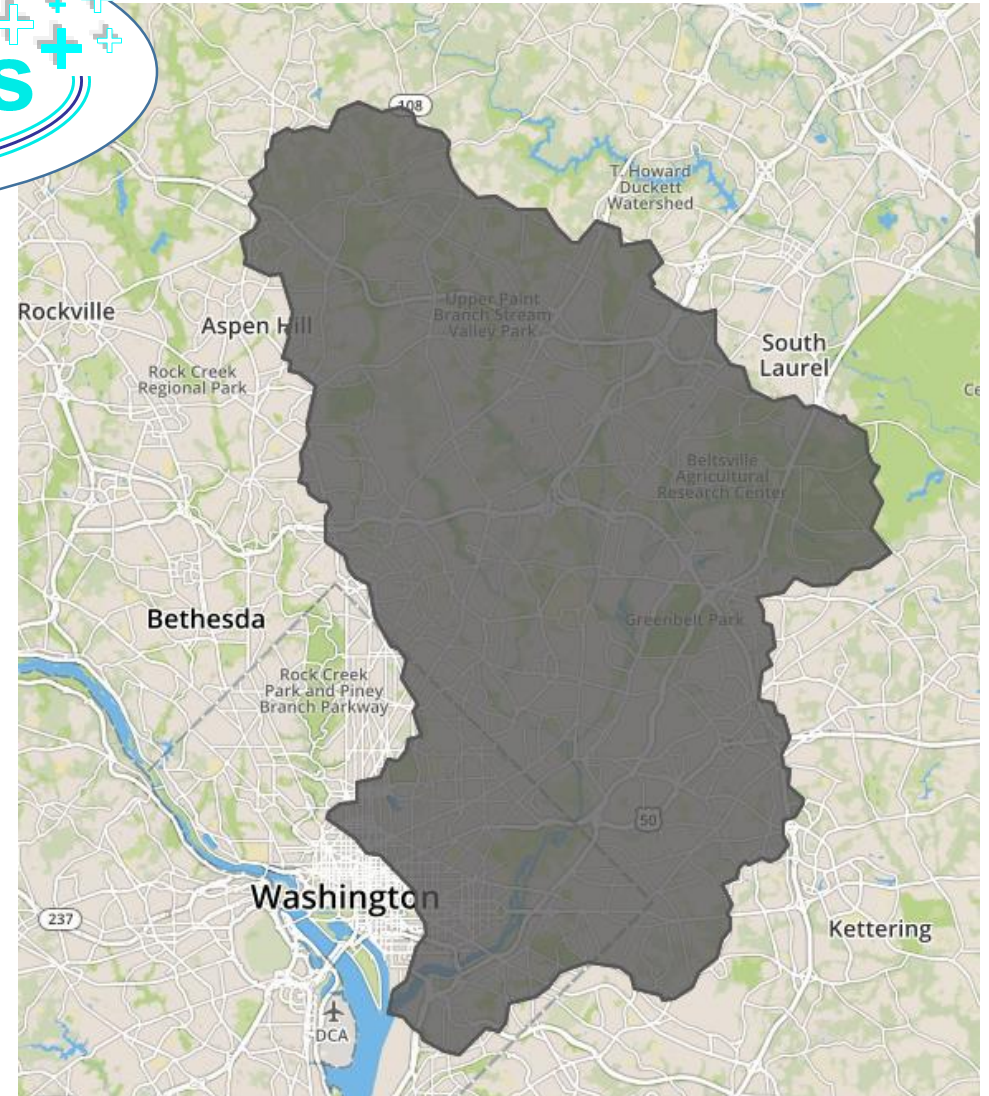


# Automating Calculation of Measures





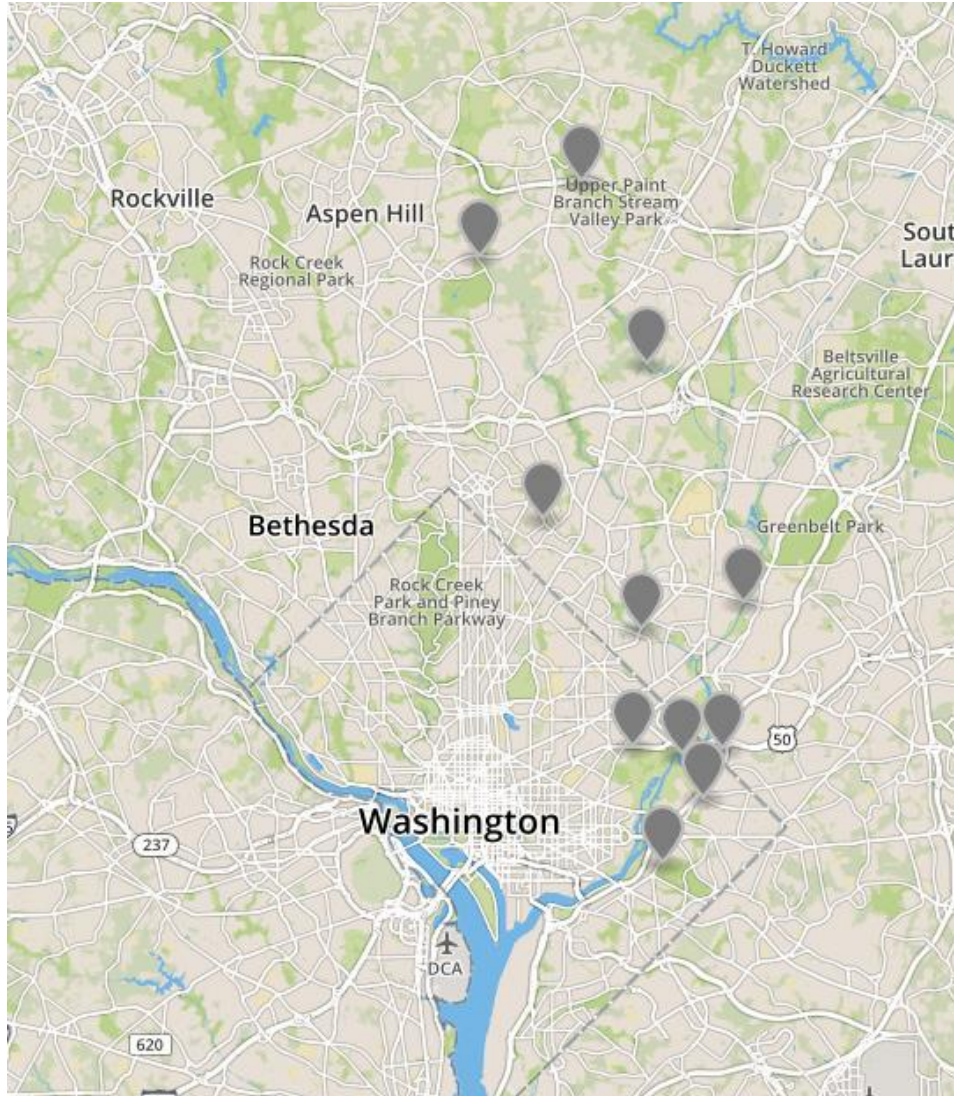
Stream Network



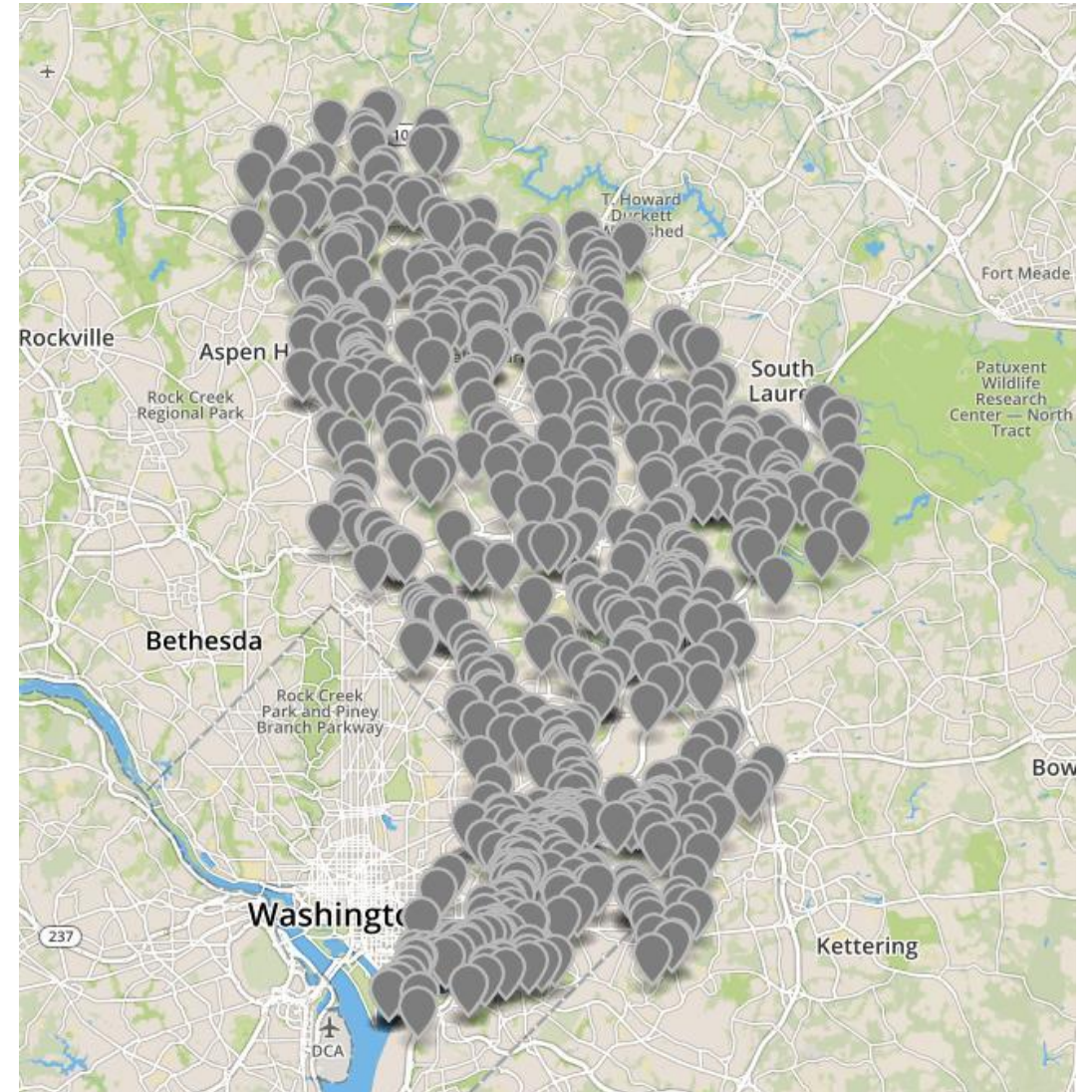
Anacostia Watershed



# National Linked Data Index (NLDI)



Sensors



Sampling Points



# Characterizing Pre-Event Conditions – A Use Case

- When an event occurs, a common question that is asked: “What did it look like before the event?”
- In August 2015, there was only one API available for characterizing pre-event conditions (Water Quality Portal API)
- Synthesizing data took a fair amount of effort from several people, and was not easily updatable as new data came in





# With integrated data there is much broader capability

- Downstream services:
  - Would allow immediate discovery of monitoring locations, and immediate data download
  - Quick discovery of impaired waters downstream
  - Quick discovery of potentially impacted drinking water facilities
- Sensor Network Services:
  - Allow access to real-time data as the event occurs
- Watershed Characterization:
  - Non-point source activities in the watershed
  - Pollution Budgets in place
- Water Use:
  - Potential irrigation withdrawal points to identify potential impacts to agriculture





# Pulling the Pieces Together: Telling the Water Story

Provide the public with usable, meaningful information

Communicate progress states, tribes, and EPA are making towards restoring or protecting water quality

Engage the public in understanding impacts on water and issues related to water





# Design Concepts

- All information is based on a service
- Tell the story at multiple levels (National, State, County, Local)
- Integrate data across systems
- Allow for in-depth stories in addition to interactive content

**Community Overview**

**Drinking Water**  
How is my community's drinking water?

**Recreation**  
How can I stay healthy fishing & swimming?

**Ecological Life**  
How does our water impact ecological life?

**Potential Issues**  
What puts clean water at risk?

**Detect**  
Is the water being tested?

**Restore**  
What is being done to restore our water?

**Protect**  
How can I help?


## Community


Water Connects us. Thanks to the Clean Water Act, the EPA, state tribes, volunteer groups, and organizations work together to monitor water quality. This information is shared with the states and the public to help protect and restore our Nations Waters.


### Your Waters: What We Know

The outline on the map shows your local sub-watershed or HUC 12.

**1**  
Drinking Water Provider

**26**  
Monitoring Stations  
 Show on map

**174**  
Permitted Dischargers  
 Show on map

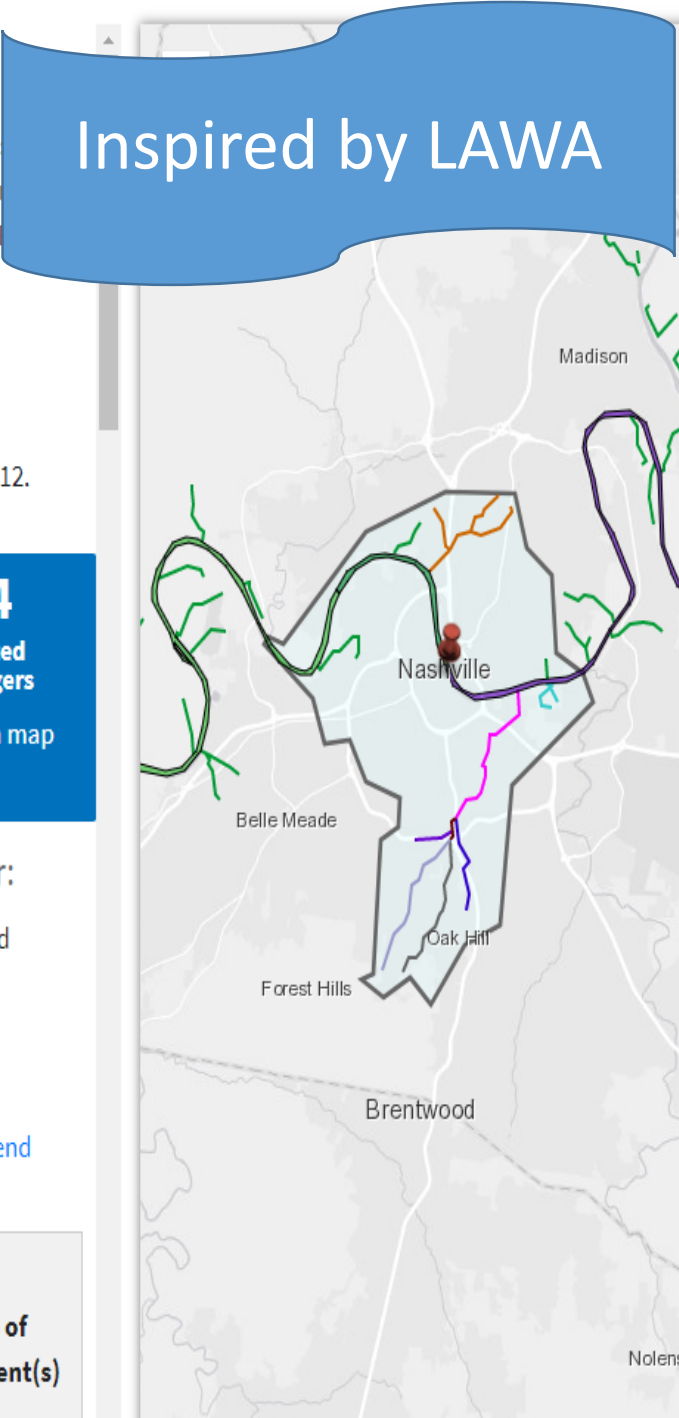
 Water Quality Assessments were conducted for:

- ☒ fish, shellfish and wildlife protection
- ☒ fishing, swimming and boating
- ☒ agricultural use
- ☒ drinking water
- ☒ industrial use

☒ Show waterbodies on map

[Hide Full Legend](#)

Waterbody Name <small>Click to view on map</small>	Color in Map	Condition	Cause of Impairment(s)



# APIs and Integrated Data Open Up Broader Usage Outside of Your Application

Retrieve Stations

Use this form to download your monitoring stations from the Water Quality Portal (<http://waterqualitydata.us>). Any data submitted to WQX is published through the Portal within one week of being submitted to WQX. If you have not submitted data to WQX, you will not be able to retrieve any data. This will retrieve all of the stations for the organization identified below.

**Current Organization (click the dropdown to change):**

CHEROKEE\_WQX

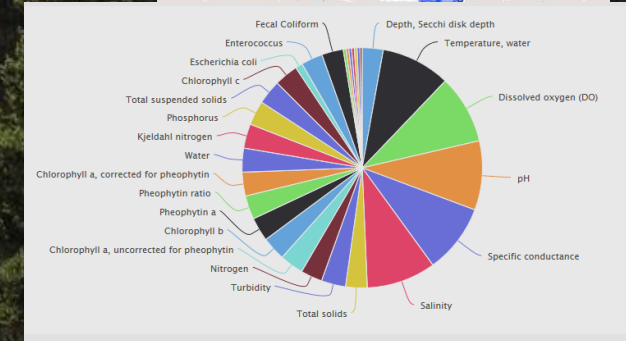
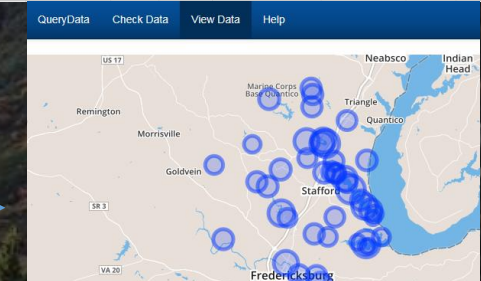
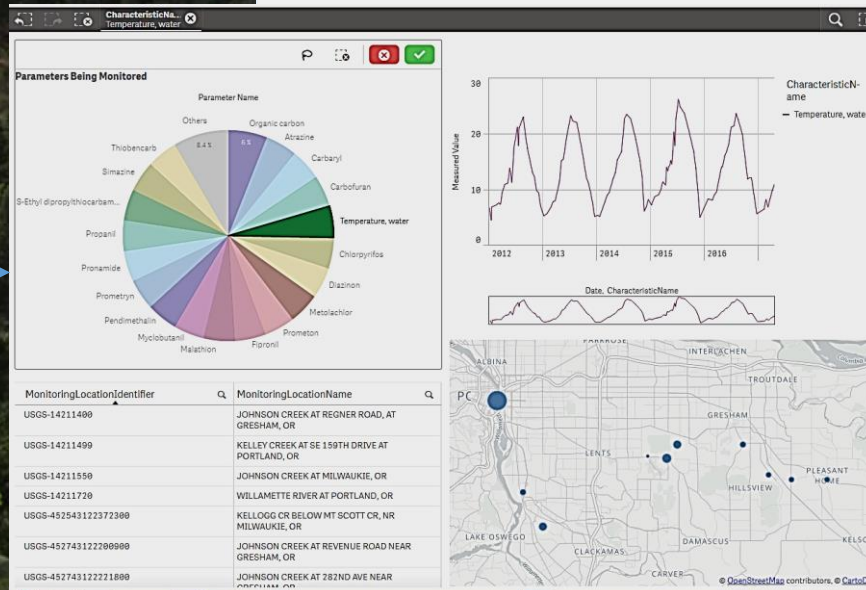
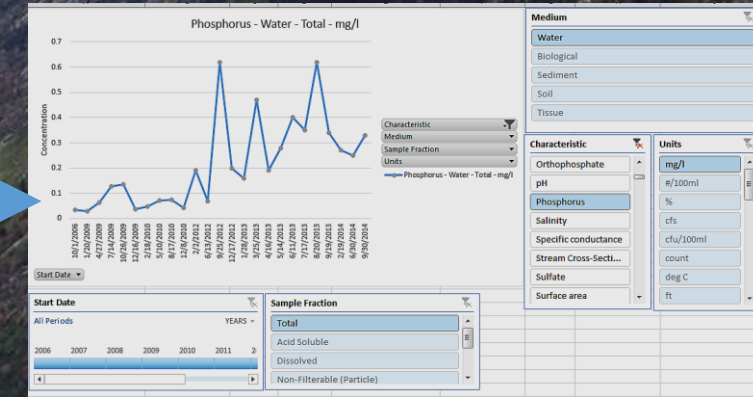
Retrieve Stations

Powered by the Portal

WQX NWQMC

Enabled through WQX

CLOSE





**Dwane Young**

[Young.dwane@epa.gov](mailto:Young.dwane@epa.gov)

**Other Links:**

Water Quality Portal:

<https://waterqualitydata.us>

National Hydrography Dataset Plus:

<https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus>

National Linked Data Index:

<https://owi.usgs.gov/blog/nldi-intro/>

Sensors Demonstration Site:

<http://54.210.62.171/>

# QUESTIONS?