

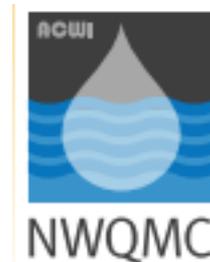
Water Quality Portal

JIM KREFT,

USGS CENTER FOR INTEGRATED DATA ANALYTICS

NATIONAL WATER QUALITY MONITORING COUNCIL
WEBINAR

FEBRUARY 18, 2015



Purpose

- Describe the Water Quality Portal
- Explain the benefits of the Water Quality Portal
- Conduct a Demonstration of the Portal
- Demonstrate a series of different uses of the portal

Portal Background

The “New Jersey Problem”

- New Jersey data was collected in a cooperative agreement with USGS
- Data Collected by USGS was in NWIS
- Data Collected by NJ DEP was submitted to STORET
- Combining data was laborious and error-prone

National Water Quality Monitoring Council took on this challenge

Portal Background- 2003 Memo of Understanding

USGS and EPA will deliver data from USGS-NWIS and EPA-STORET in a common format to:

- 1. Analyze and report on the state of the nation's water environment
- 2. Provide a common basis for integrated water-quality analysis and protection
- 3. Provide an information base for scientific inquiry about water quality

An underlying goal is to ensure that the data from these important government databases are documented to describe their quality so that users can establish the utility and comparability of the data.

Benefits of the Portal

Reduces the effort to use other data sources

- Collecting data from multiple sources
- Combining into common format
- Delivering in a single file

Leverages and protects investments in monitoring data

- Common data elements emerging from monitoring community
- Marketplace of what, when and where for monitoring

Supports water quality based decision making

- Comparing to water quality standards
- Identifying hotspots
- Developing protection and restoration plans
- Modeling expected changes

Data are valuable, plan for re-use

Electronic data are more valuable than data in file cabinets

The more data are re-used, the more valuable they become

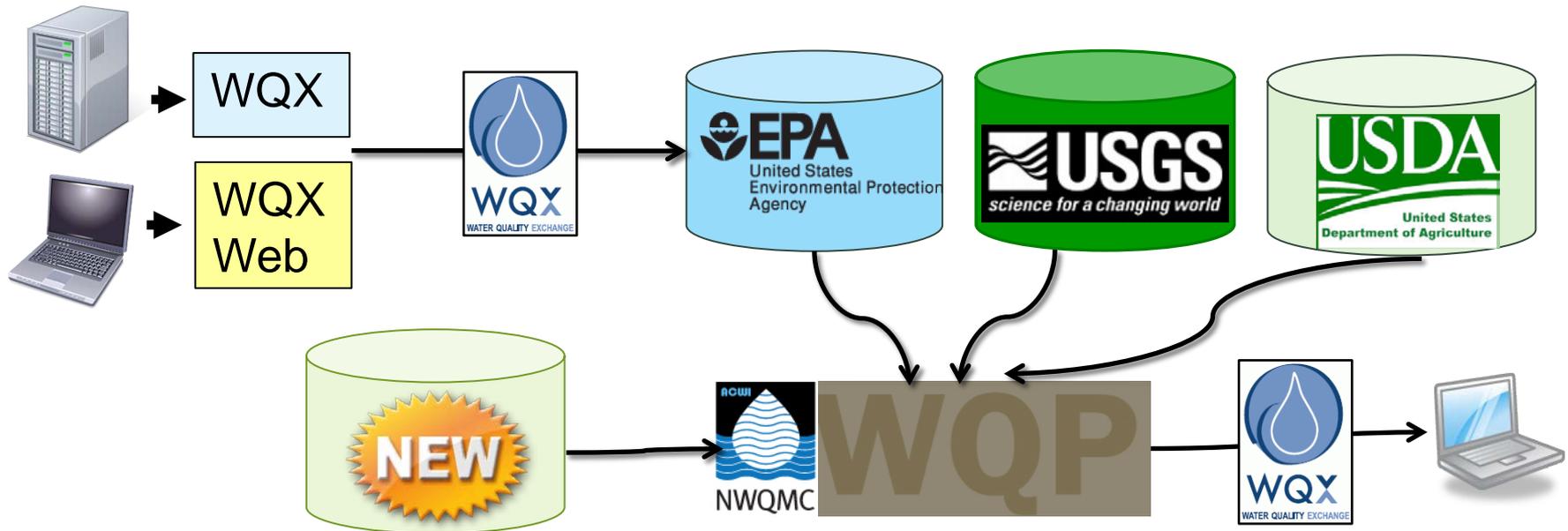
- Collect once – use multiple times

Shared data are of even higher value

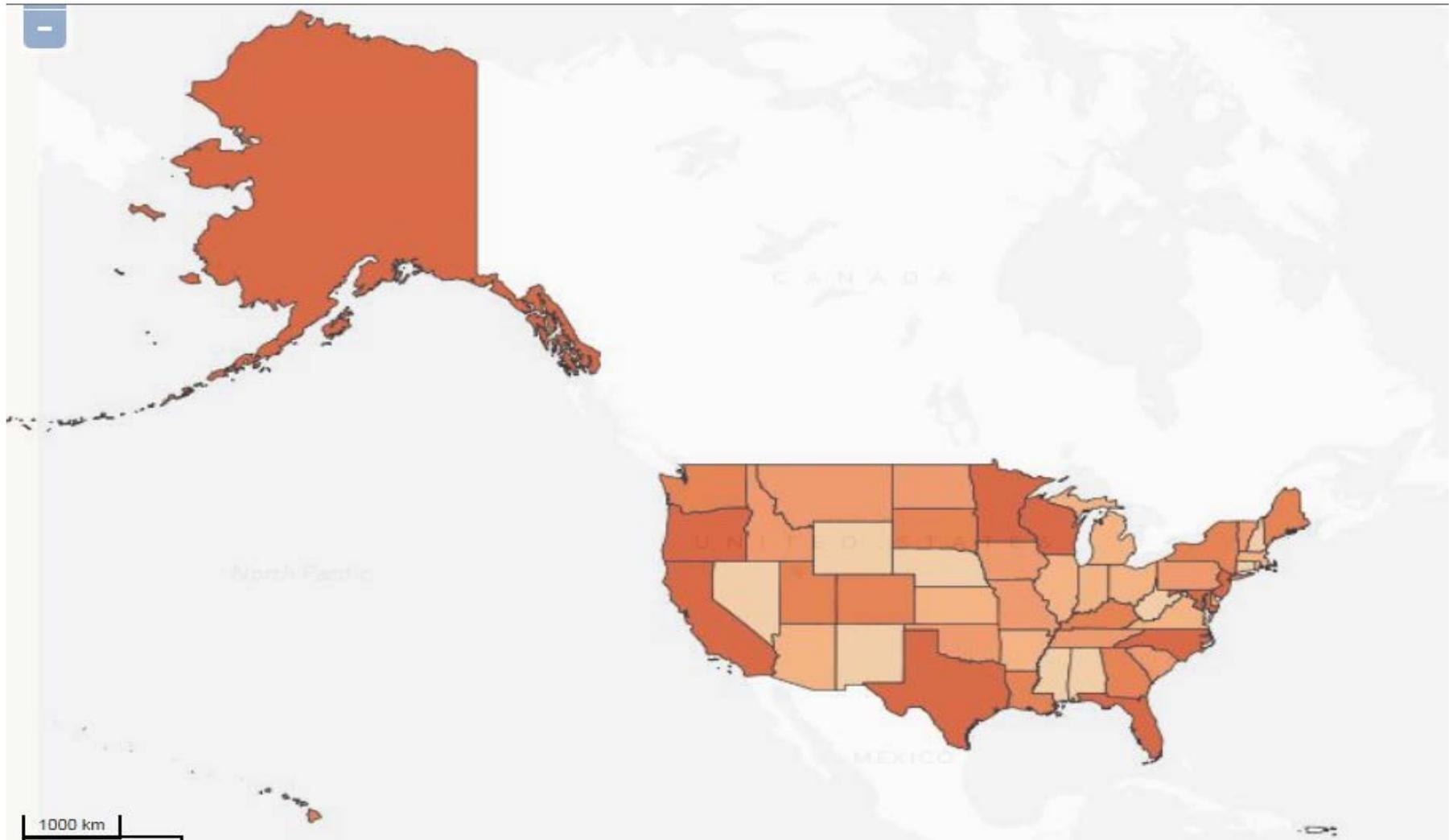
- Provide for better planning and management decisions
- Incentivize collaborative efforts
- Make the most use of the data collection resources being invested

Increases Access to Data

With the National Water Quality Monitoring Council (**NWQMC**), the Water Quality Portal (**WQP**) integrates publicly available water-quality data, through use of the Water Quality eXchange (**WQX**), from the USGS **NWIS**, EPA **STORET**, and USDA **ARS STEWARDS**.



Access to Water Quality Data



Over 250 million discrete water data records and 2.2 million stations

Access to Multiple Data Types

Portal Data Records (252m total)

- USGS NWIS – 90m records
- USDA STEWARDS – 1m records
- USEPA STORET – 162m records

Portal Data Contributors

- Federal – EPA, USGS, USACOE, NPS, USBR
- States and territories – 50 with 5 more in progress
- Tribes – 130 agencies
- Other organizations – county, watershed groups, academic

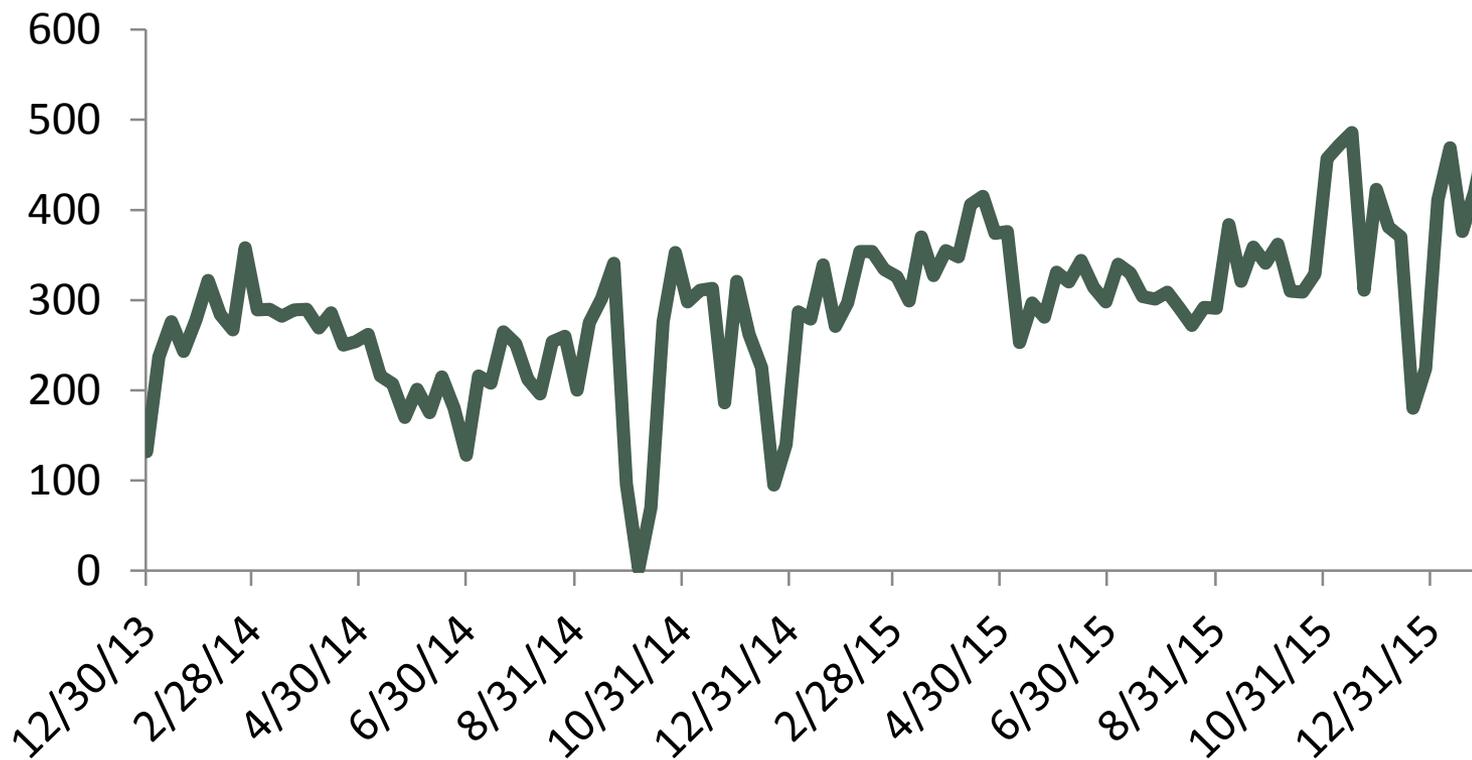
Portal Data types

- Physical/Chemical
- Biological (Coming soon)
- Habitat, Metrics, Indexes (Coming Soon)

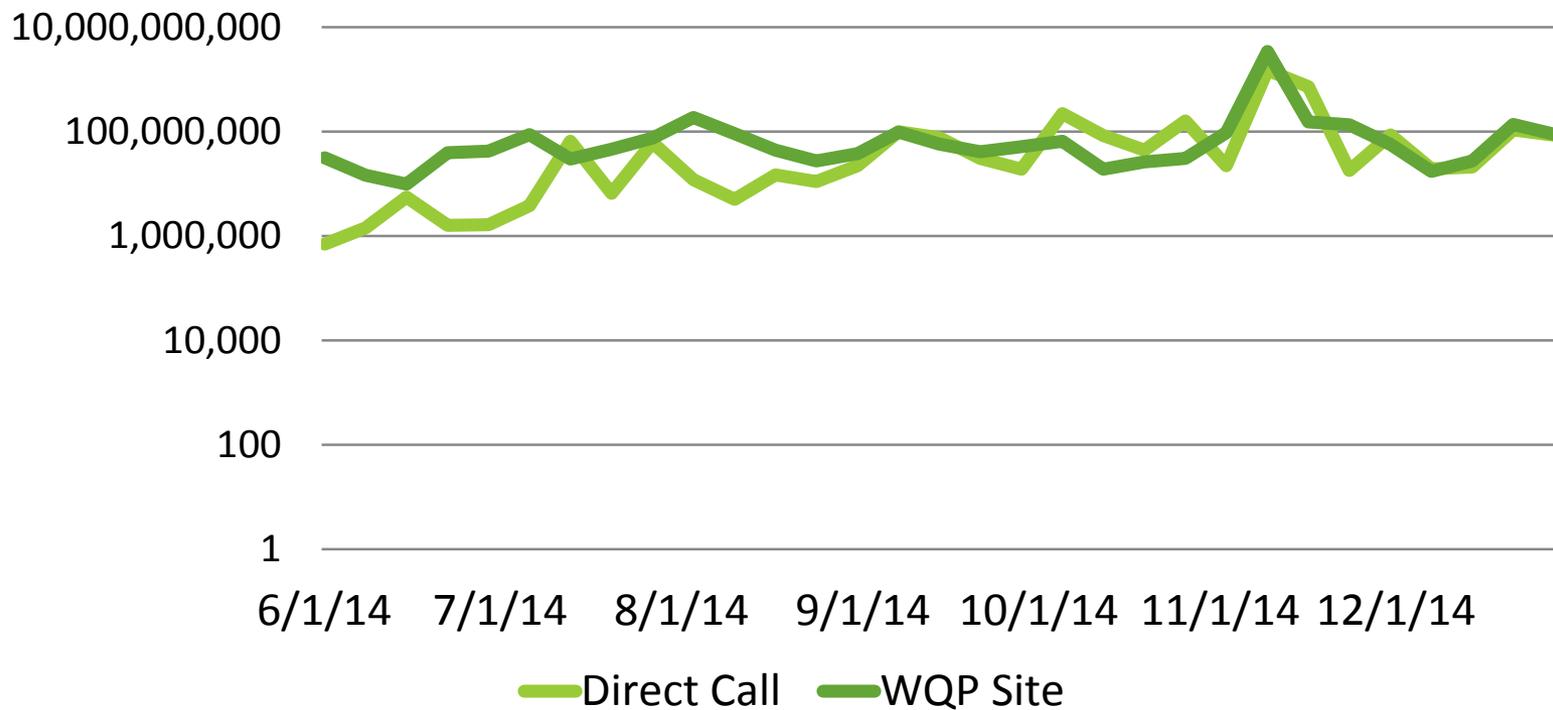


Portal Usage continues to rise

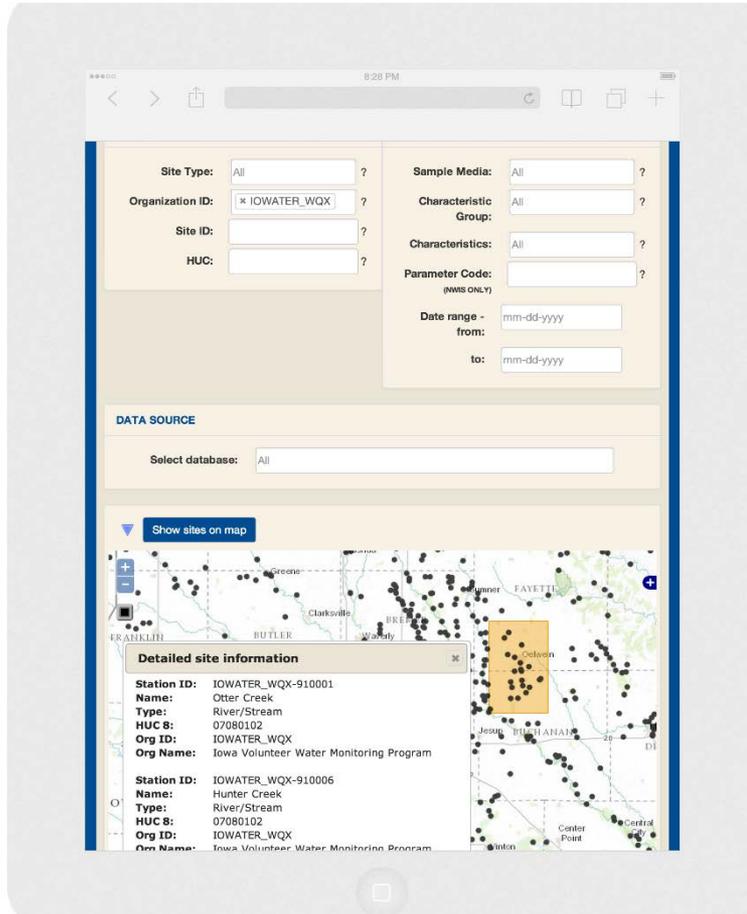
Weekly Visits



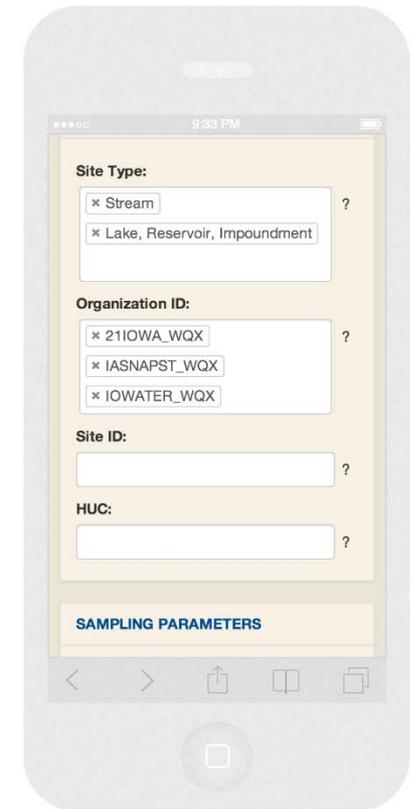
Result Records Downloaded



Built using Responsive Design



Water Quality Portal works on any device with a modern web browser



Water Quality Portal Demo

Portal data holdings

- <http://waterqualitydata.us>
- Characteristics and query capabilities
- Display stations on a map
- Web Service usage example

The screenshot shows the homepage of the National Water Quality Monitoring Council (NWQMC) Water Quality Portal. The header features the NWQMC logo and the text "NATIONAL WATER QUALITY MONITORING COUNCIL" and "Working Together for Clean Water". The main content area includes a large "WQP" logo, the title "Water Quality Portal", and a brief description: "The Water Quality Portal (WQP) is a cooperative service sponsored by the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and the National Water Quality Monitoring Council (NWQMC)." Below this, there are three main sections: "DOWNLOAD DATA" (with a link to download data in Excel, CSV, TSV, and KML formats), "HOW TO USE THE WQP" (with links for User Guide, Web Services Guide, FAQs, and Upload Data), and "NATIONAL RESULTS COVERAGE" (with a link to view water-quality data in your state). At the bottom, there are logos for USGS and EPA.

Demo- Query Options

- Geospatial
 - Countries
 - States
 - Counties
 - Bounding box
 - Radius around location
- Site Metadata
 - Organization
 - Site type
 - Site ID
- Sampling information
 - Sample type
 - Characteristic Type
 - Characteristic Name

different file types

what does the “RESTlike Query” button do for you?

Demo- Data display and download options

Map Display

- Coming soon-
 - More informative styling
 - Web Map Service
 - Web Feature Service

Download options

- Tabular
- KML for sites

RESTLike Query

The key to advanced usage of the portal.

Every combination of the

Show RESTlike queries ?

Sites

```
http://www.waterqualitydata.us/Station/search?statecode=US%3A55&siteType=Stream&characteristicType=Nutrient&mimeType=csv&zip=yes
```

Sample results

```
http://www.waterqualitydata.us/Result/search?statecode=US%3A55&siteType=Stream&characteristicType=Nutrient&mimeType=csv&zip=yes
```

What is a client?

Bridge between the an API and an external tool.

Allows people familiar with a given software tool to easily get started without having to first figure out the WQP API

Water Quality Portal has two clients:

- DataRetrieval for R
- pywqp for Python

Water Quality Portal R Client

<https://github.com/USGS-R/dataRetrieval>



The screenshot shows the GitHub repository page for 'USGS-R / USGSHydroTools'. The repository is described as a 'Compilation of tools to aid in hydrologic and water quality analysis.' It has 71 commits, 3 branches, 0 releases, and 3 contributors. The current branch is 'master'. The commit history shows a recent update to the README.md file by 'Idecico-USGS' 26 days ago. The commit message is 'Update README.md'. The commit details show changes to the 'R' directory (updated help files), 'data' directory (clipped the big shapefile), 'man' directory (updated help files), 'DESCRIPTION' file (changed name to USGS), 'GSHydroTools.Rproj' file (forgot to right Rd files), 'LICENSE' file (updated license text), 'NAMESPACE' file (added summarize data function), and 'README.md' file (update README.md).

File	Commit Message	Time Ago
R	Updated help files to reflect current function.	29 days ago
data	Clipped the big shapefile, added some defaults to new arguments.	a month ago
man	Updated help files to reflect current function.	29 days ago
DESCRIPTION	Changed name to USGS	a month ago
GSHydroTools.Rproj	forgot to right Rd files with roxygen...had to rebuild the package so...	2 months ago
LICENSE	Updated license text.	8 months ago
NAMESPACE	Added summarize data function, started cleaning up mapSizeColor. Stil...	a month ago
README.md	Update README.md	26 days ago

<https://github.com/USGS-R/USGSHydroTools>



Using R for data exploration

DataRetrieval

- Officially supported R client for WQP and NWIS data. On CRAN and Github

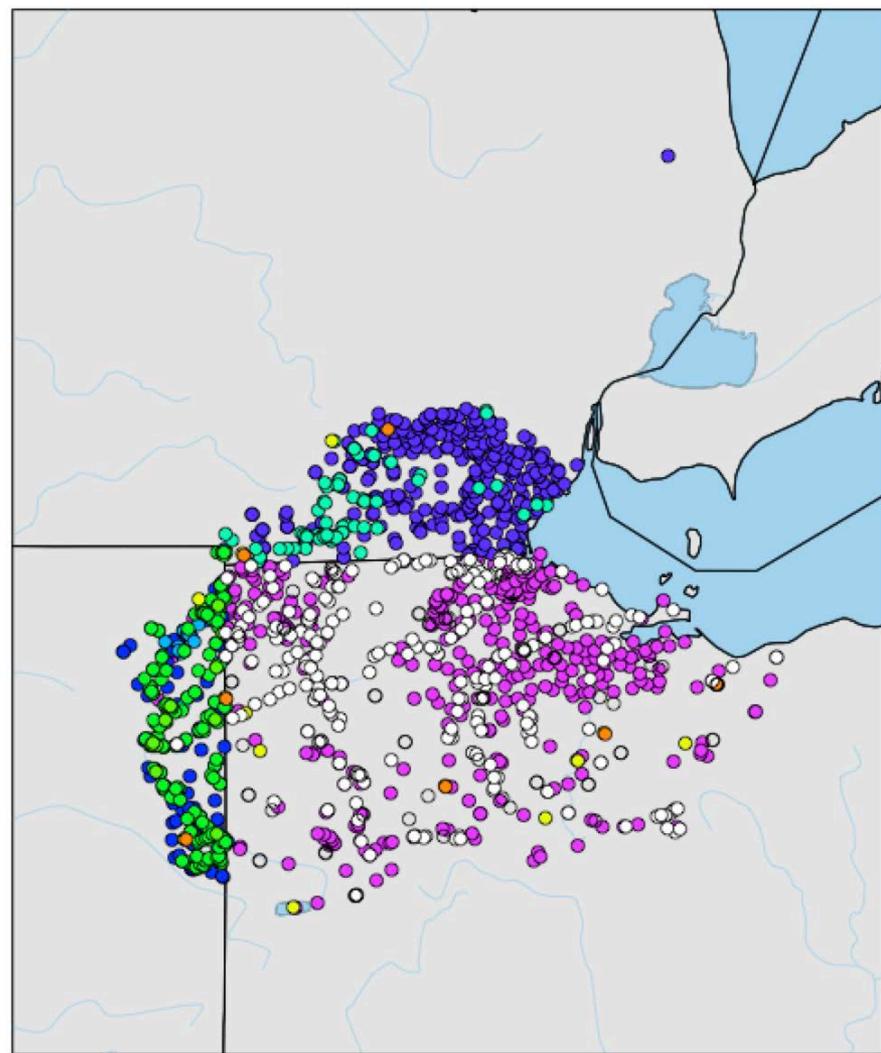
WQP_Data_Exploration

- suite of R-based data exploration examples for the WQP, available to all on Github
- https://github.com/USGS-CIDA/WQP_Data_Exploration

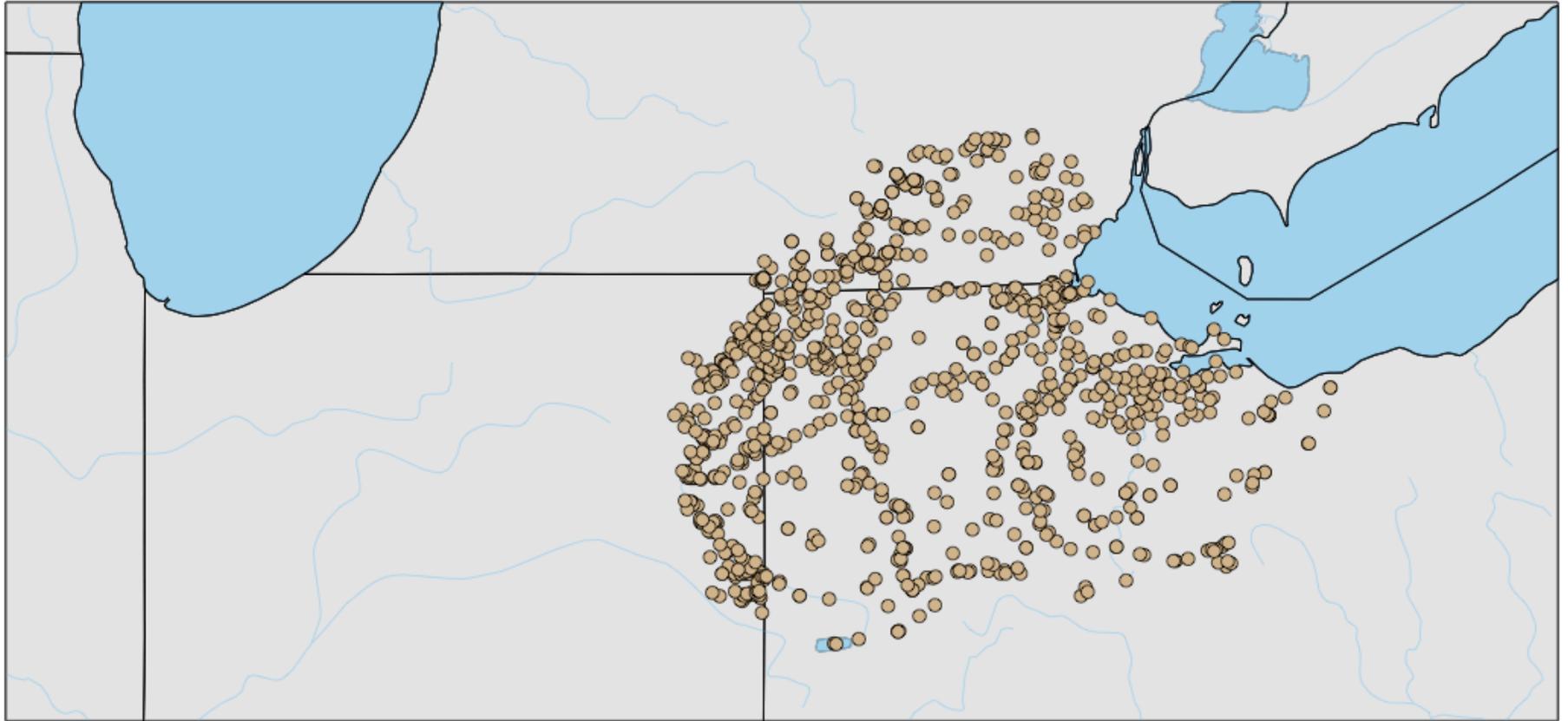


What agencies have sampled for nutrients in the western Lake Erie basin?

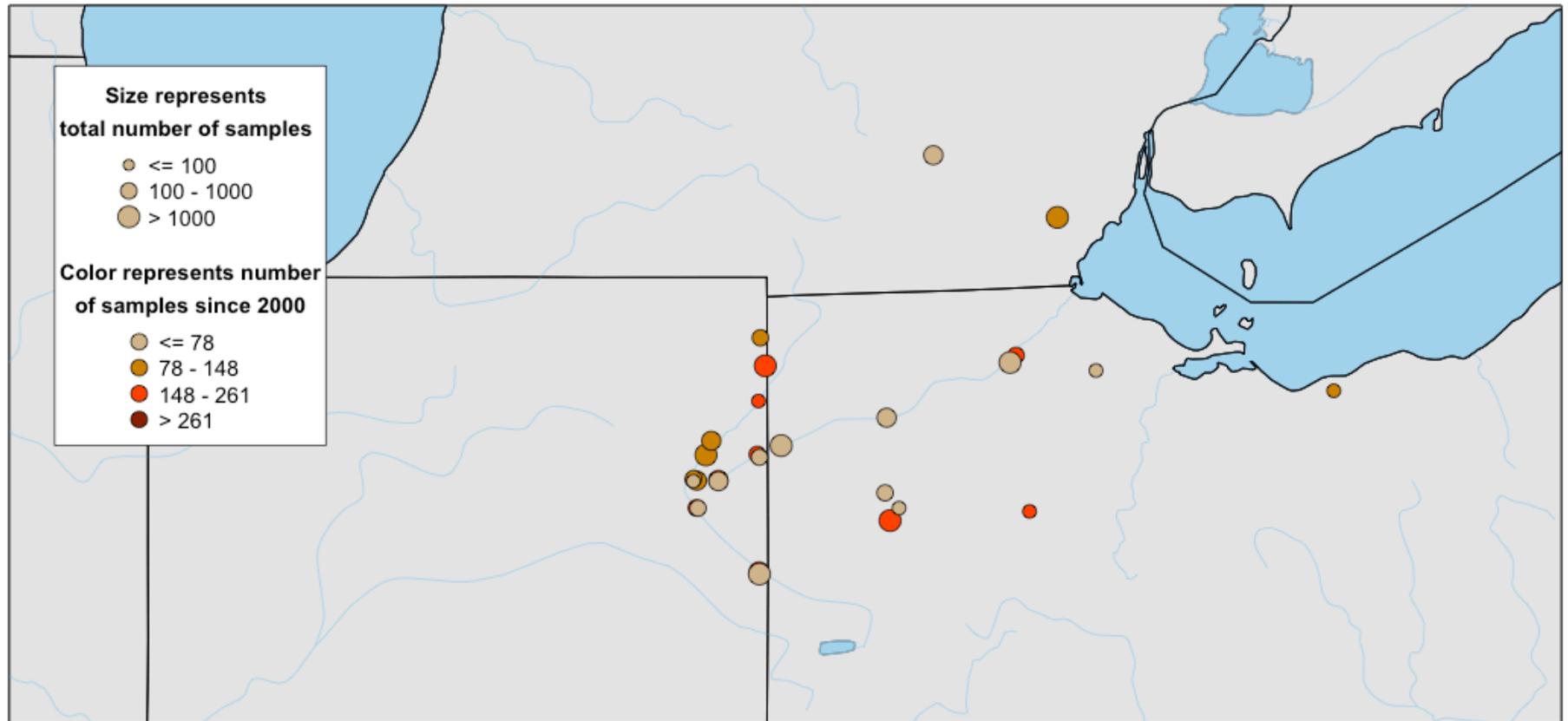
	OrganizationFormalName	count
1	USGS Ohio Water Science Center	427
2	USGS Michigan Water Science Center	351
3	Division of Surface water (Ohio)	221
4	Indiana STORET	201
5	USGS Indiana Water Science Center	107
6	Michigan Department of Environmental Quality	74
7	Division of Drinking and Ground Water (Ohio)	63
8	EPA National Aquatic Resources Survey	16
9	USDA Agricultural Research Service	16
10	IDEM	12
11	EPA National Aquatic Resource Survey Data	7



How many sites in the Western Lake Erie basin have been sampled for phosphorus?



What sites have more than 50 phosphorus samples, with at least 10 of those samples after 2000?



Change Two lines of Code:

```
data <- readWQPdata(huc="0410*",  
characteristicName="Phosphorus")
```

is changed to

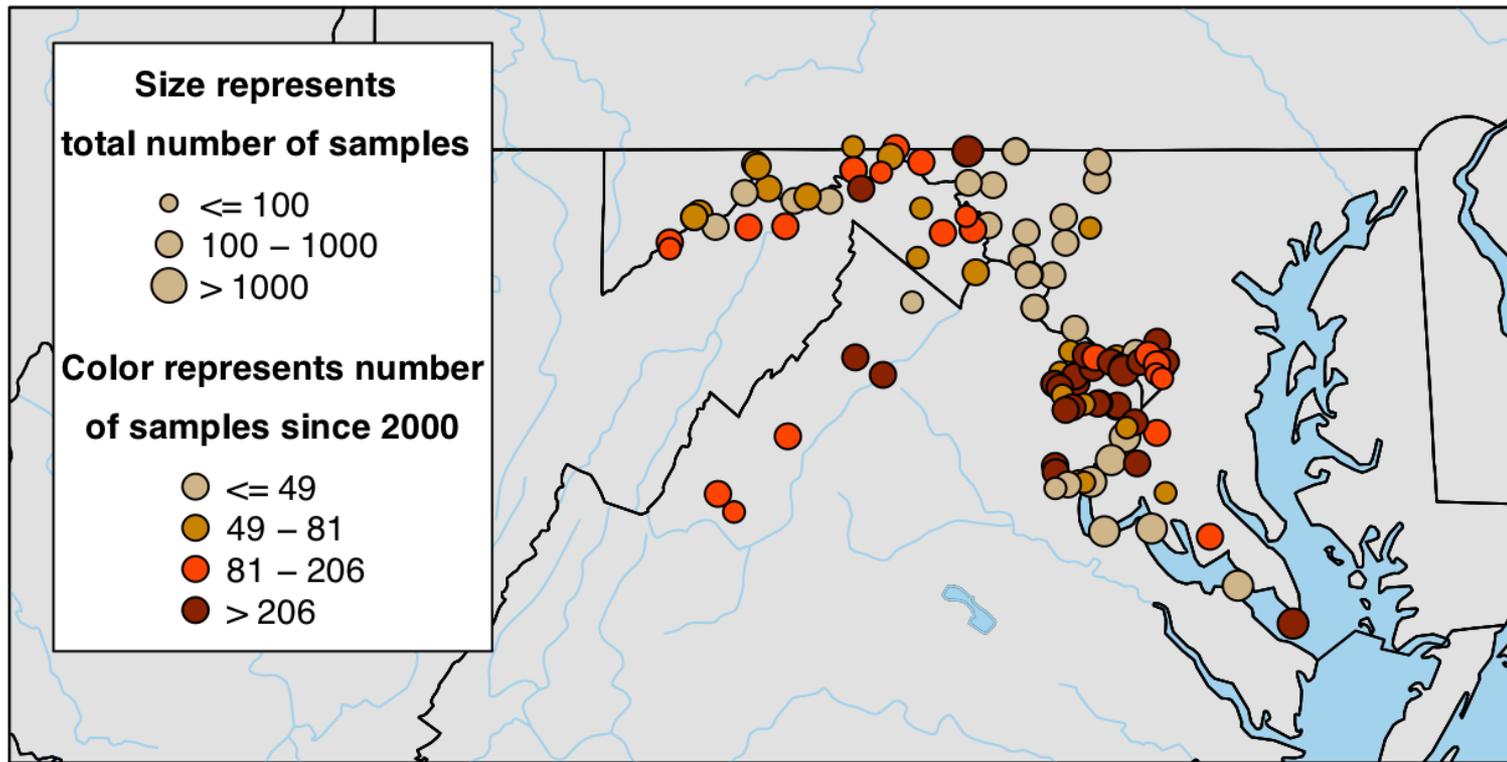
```
data <- readWQPdata(huc="0207*",  
characteristicName="Phosphorus")
```

```
#sites <- whatWQPsites(huc="0410*",  
characteristicName="Phosphorus")
```

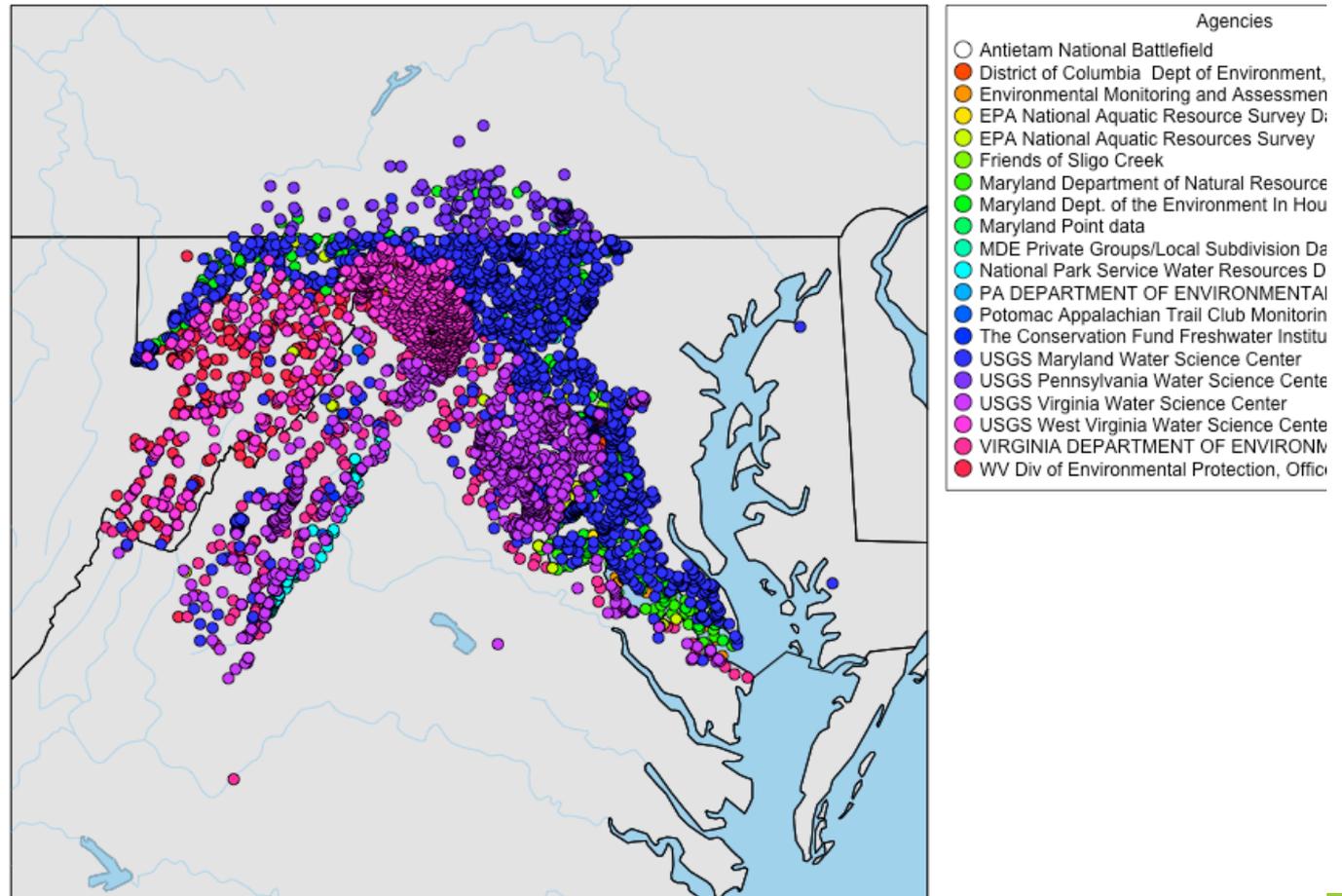
is changed to

```
sites <- whatWQPsites(huc="0207*",  
characteristicName="Phosphorus")
```

And we have changed to the Potomac



Potomac Organizations and sites that collect nutrient data

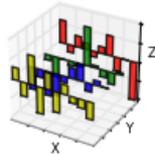
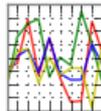
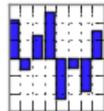


Water Quality Portal Python Client

<https://github.com/USGS-CIDA/pywqp>



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



This repository Search Explore Gist Blog Help jkreft-

USGS-CIDA / pywqp Unwatch 13

A generic scriptable Python client for downloading datasets from the USGS/EPA Water Quality Portal: an alternative to manual use of the website at <http://www.waterqualitydata.us>.

83 commits 1 branch 0 releases 5 contributors

branch: master pywqp / +

add documentation for wxq_mappings

whblondeau authored 16 days ago latest commit e5307c4cfc

.idea	fixed references in the fileformat section, cleaned up formatting to ...	2 months ago
documentation	mass restructuring following lettuce BDD implementation	2 months ago
pywqp	manual merge with origin master	26 days ago
tests	steps to support dataframe.features	22 days ago
.gitignore	Add LICENSE, edit README and .gitignore	22 days ago
LICENSE	add hard line breaks	22 days ago
README.md	add documentation for wxq_mappings	16 days ago
dev_setup.sh	Added virtualenv, pip install setup script.	2 months ago
requirements.txt	Added virtualenv, pip install setup script.	2 months ago
setup.py	manual merge with origin master	26 days ago

README.md

pywqp

A generic scriptable Python client for downloading datasets from the Web Services offered by the USGS/EPA Water Quality Portal: an alternative to manual use of the [WQP website](http://www.waterqualitydata.us).



Portal FY15 Activities

Development

- Identify additional data partners and sources
- Examine opportunities to improve data quality
- Serve biological data out of the Portal
- Improve mapping interface and possibly NHD based searches
- Discuss continuous data solution
- Increase visibility of data in the portal

Outreach

- Highlight new community tools
- Promote data sharing
- Attend meetings
- Conduct webinars

Questions

Charles Kovatch

Kovatch.Charles@epa.gov

202-566-0399

Jim Kreft

jkreft@usgs.gov

608-821-3919

