Water Quality Portal

JIM KREFT,

USGS OFFICE OF WATER INFORMATION

IOOS DMAC MEETING

JUNE 2, 2016

Purpose

- Describe the Water Quality Portal
- Explain the benefits of the Water Quality Portal
- Where are we at now?
- What is the future?
- How can the Water Quality Portal Work with IOOS?

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.

Mission and Vision

Vision

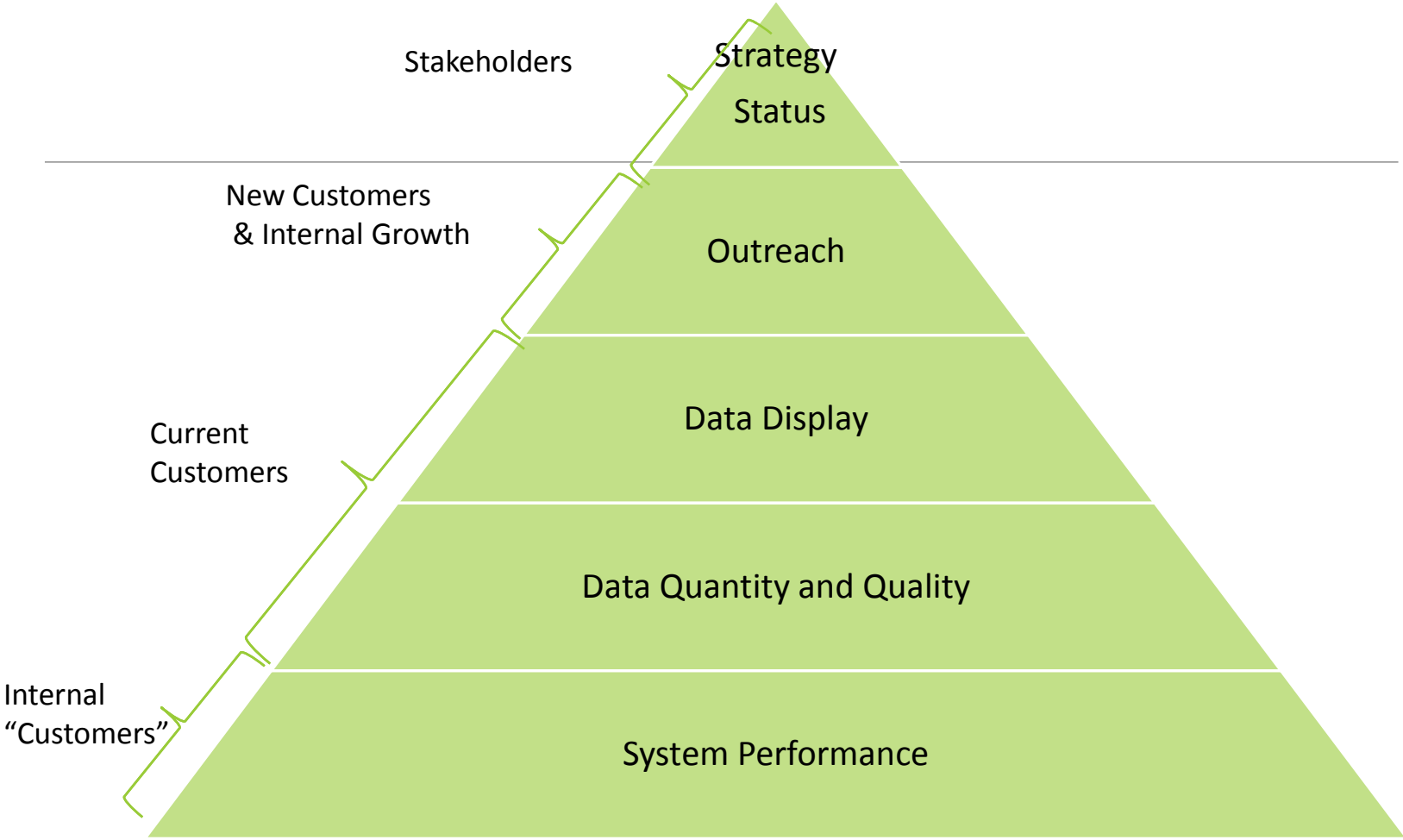
Be the premiere source for water quality data for everyone, everywhere.

Mission

Provide easy access to all water quality data, facilitate improvements in data quality, and enhance data discovery and data summaries to inform sound water-quality decision making at local, state, regional, and national scales.

Scope

Water quality data collected from discrete samples of ambient surface and groundwater in the United States.



Relationship of 5 Portal Strategy Sections to Customer Groups

Data are valuable, don't just use them once!

Electronic data are more valuable than data in file cabinets

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

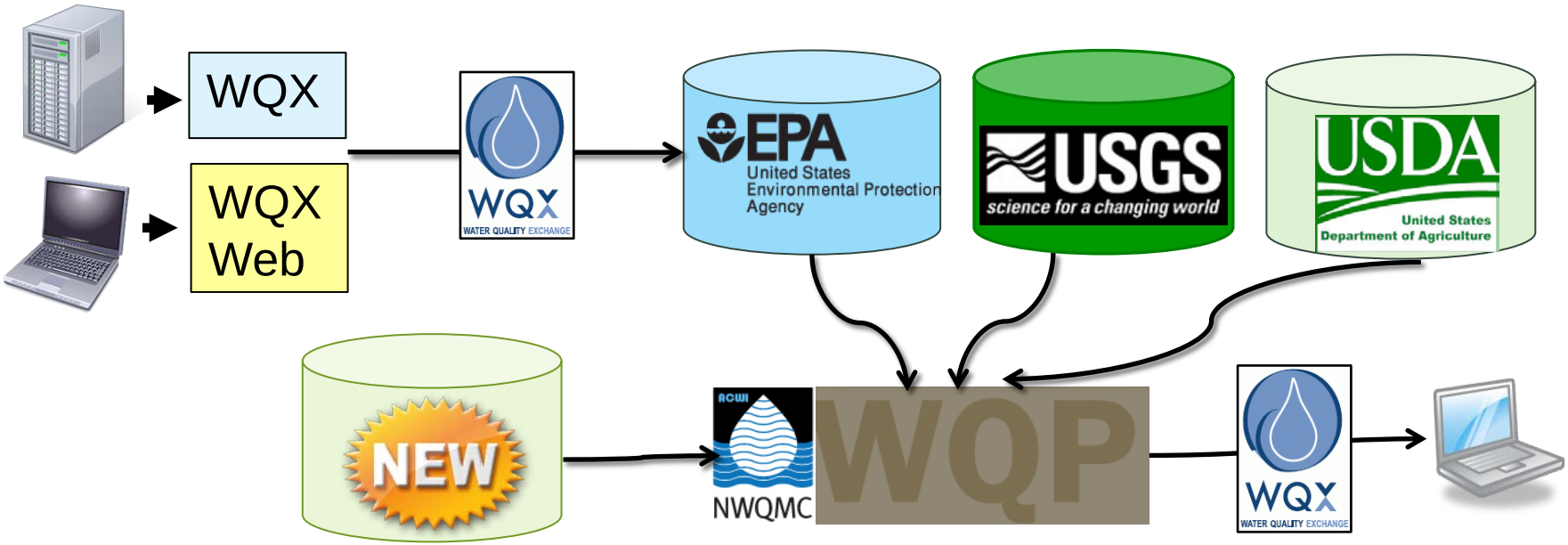
- Collect once – use multiple times

Shared data are of even higher value

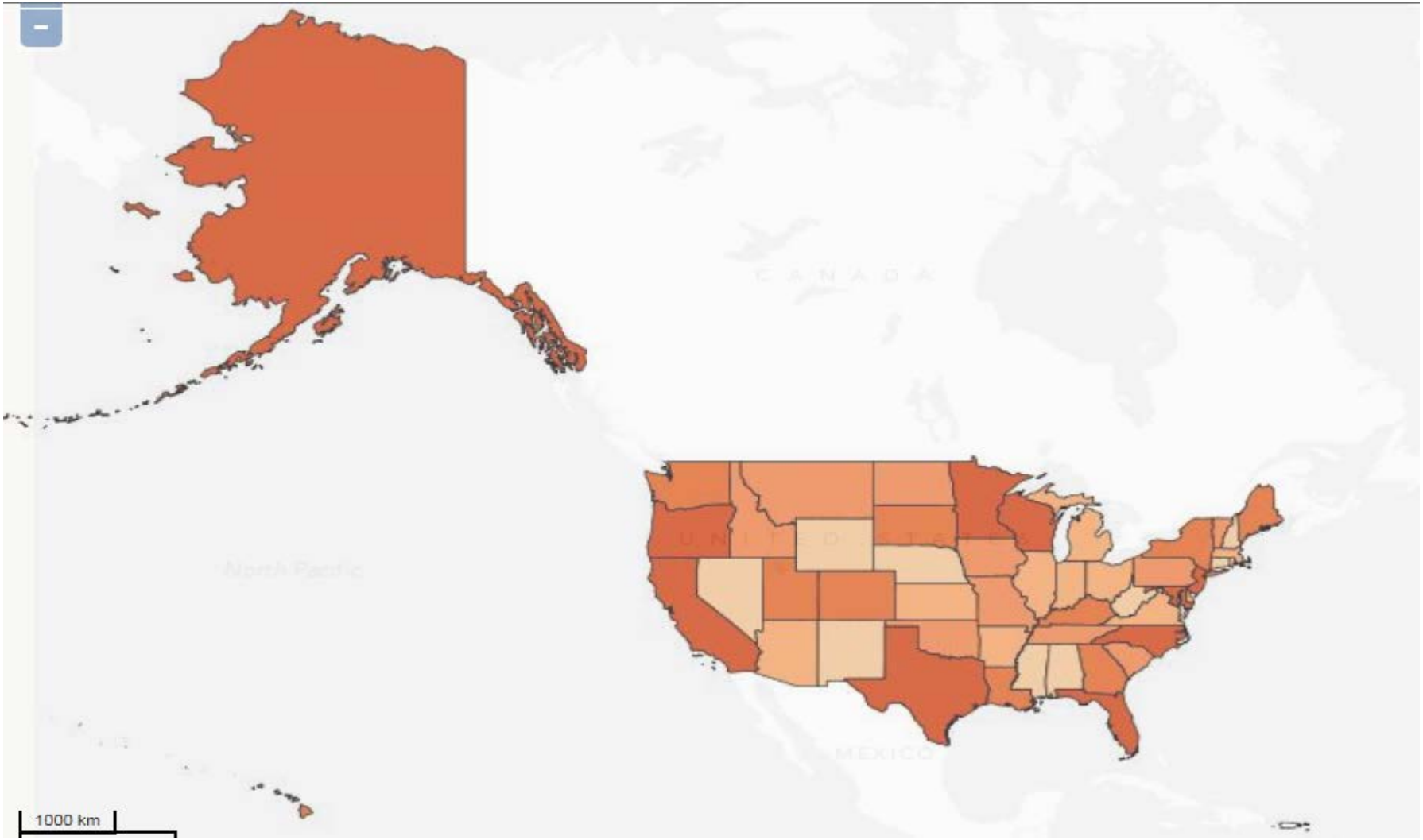
- Provide for better planning and management decisions
- Incentivize collaborative efforts
- Sharing data makes the most use of the data collection resources being invested

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 275 million discrete water data records and 2.5 million stations

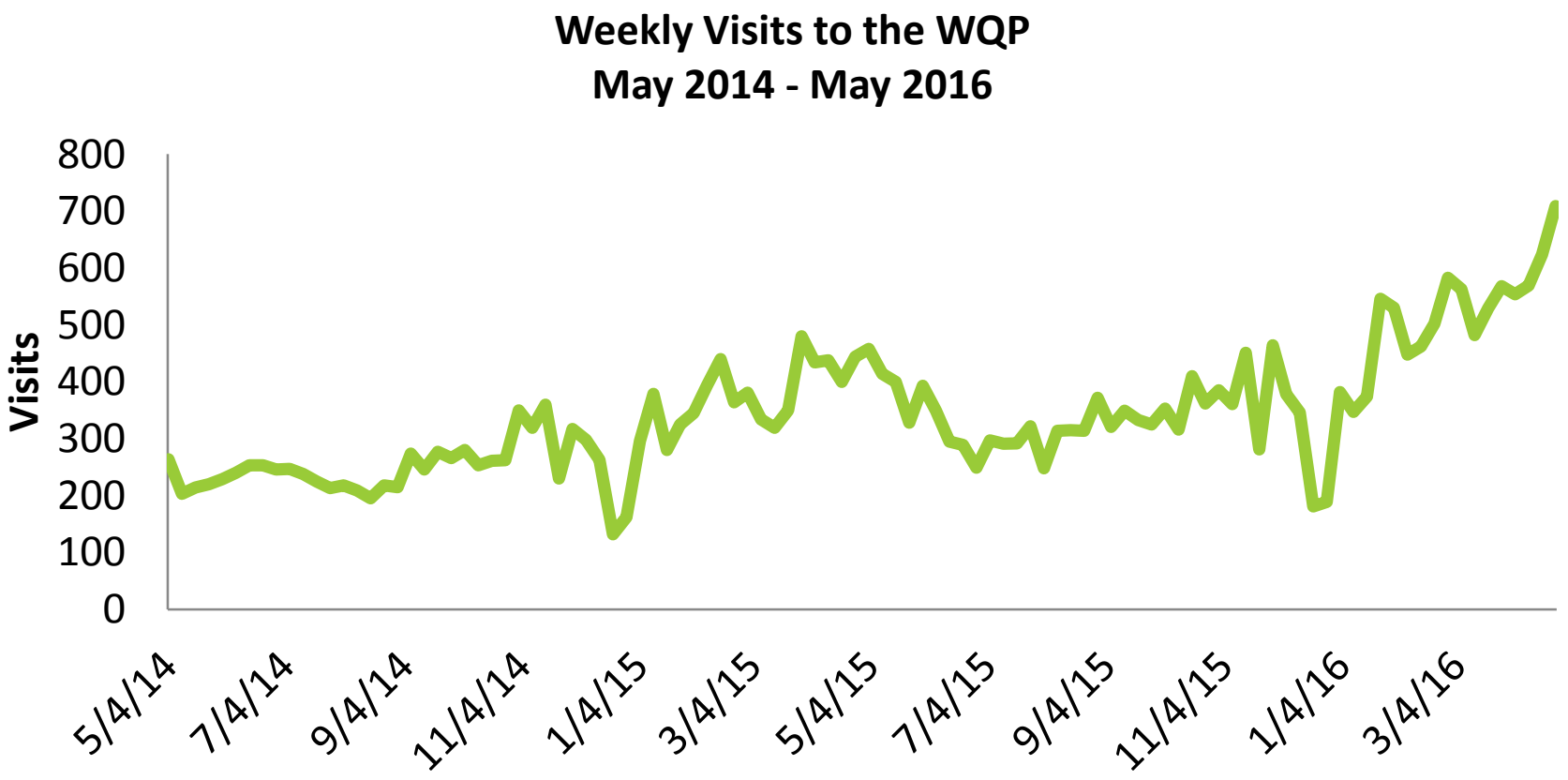
Access to Multiple Data Types

Portal Data Records (272m total)
- USGS NWIS – 90m records
- USDA STEWARDS – 1m records
- USEPA STORET – 162m records

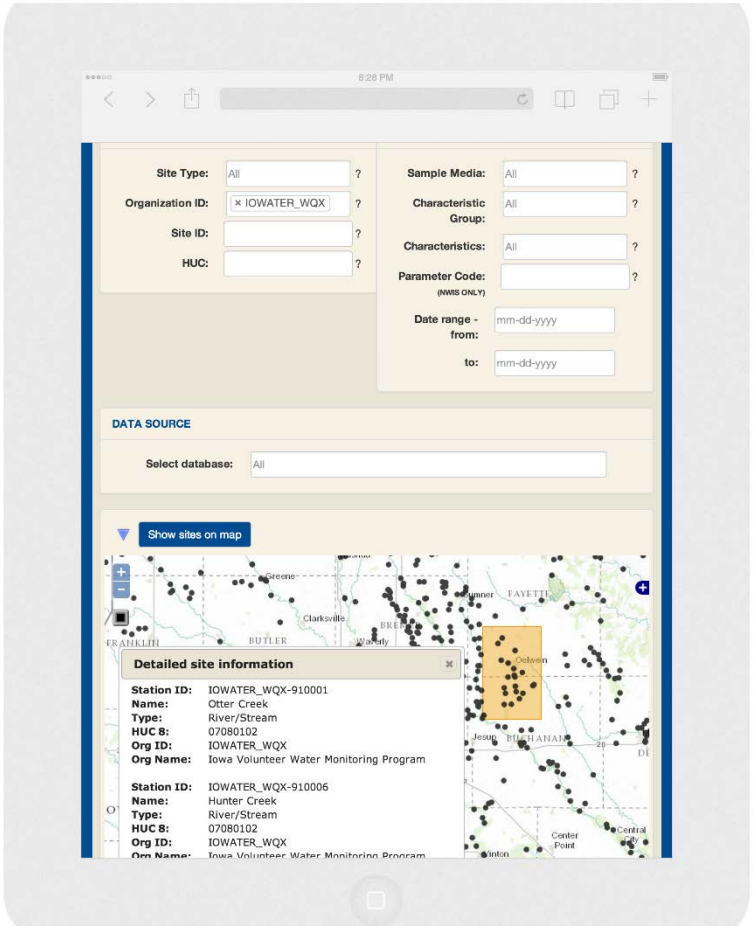
Portal Data Contributers
- Federal – EPA, USGS, USACOE, NPS, USBR
- States and territories – 50 with 5 more in progress
- Tribes – 130 agencies
- Other organizations – county, watershed groups, academic

Portal Data types
- Physical/Chemical
- Biological Collections
- Habitat, Metrics, Indexes (Coming Soon)

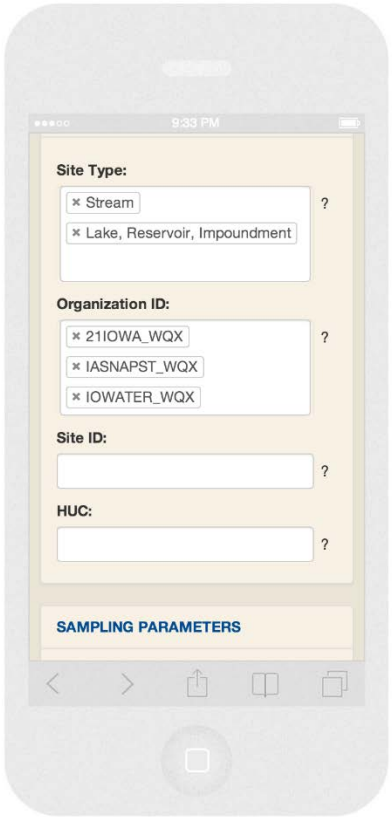
Portal Usage continues to rise



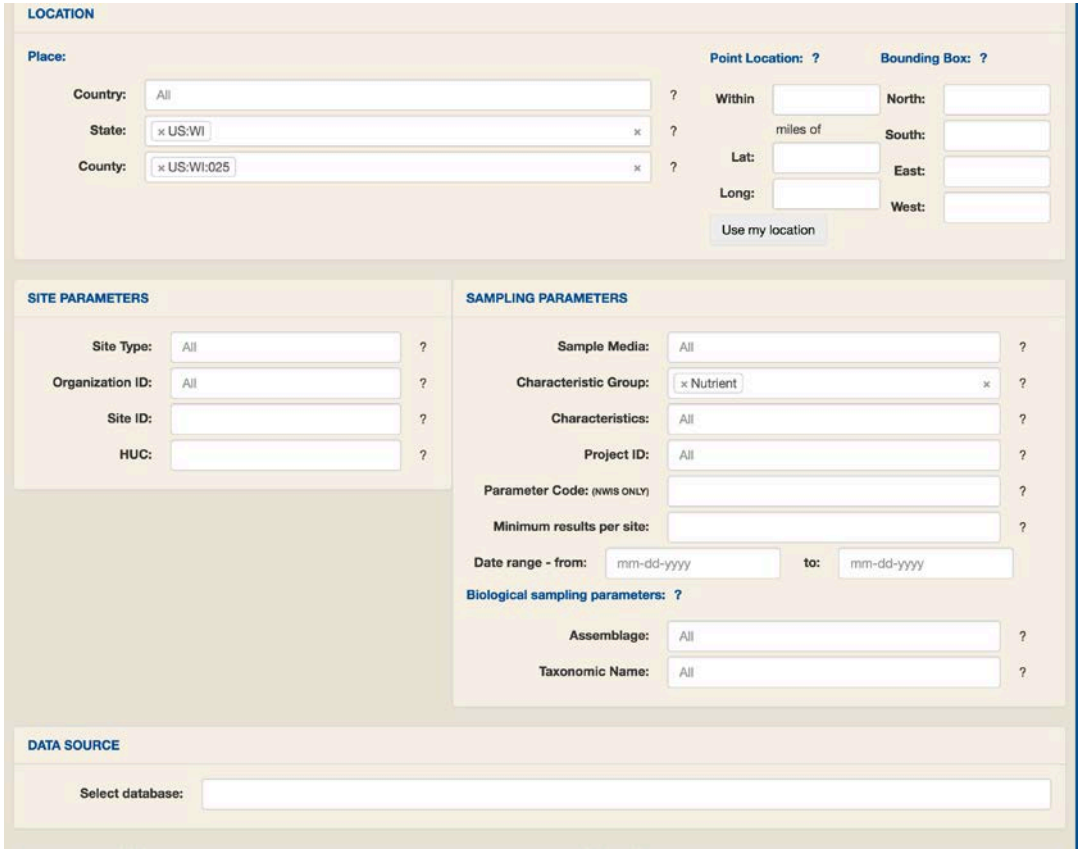
Built using Responsive Design



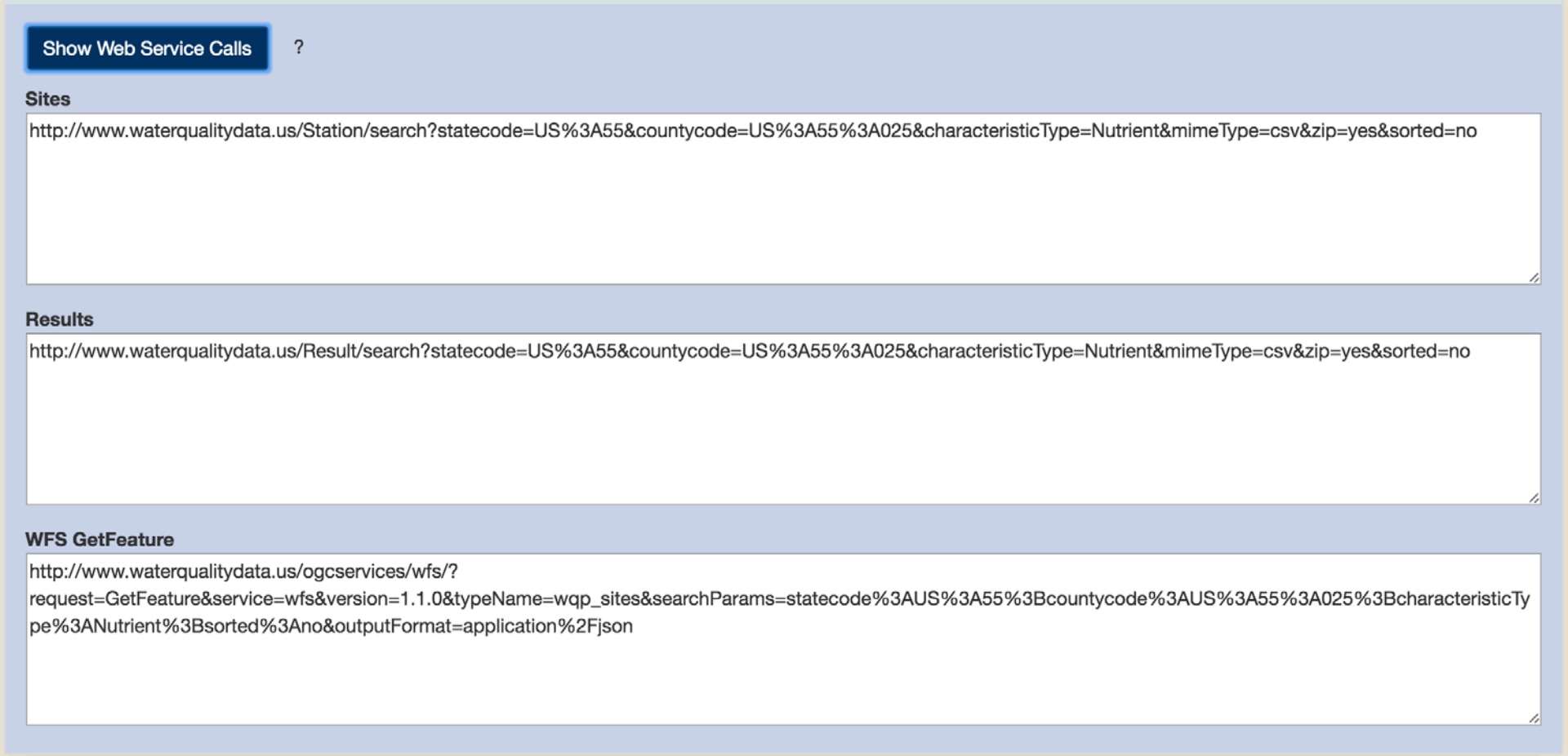
Water Quality Portal works on any device with a modern web browser



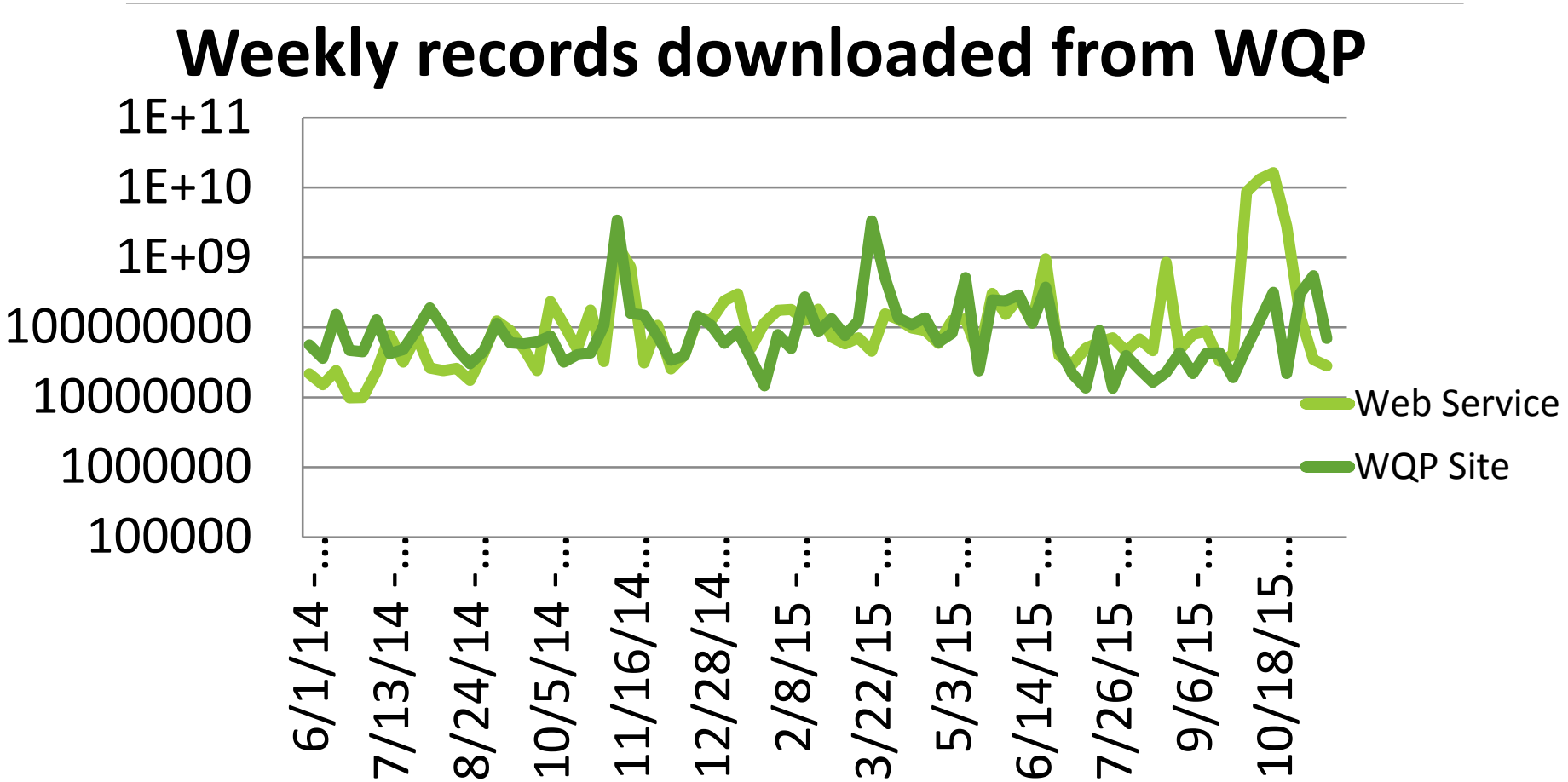
WQP Portal Page



Web Service Calls



Web services are key



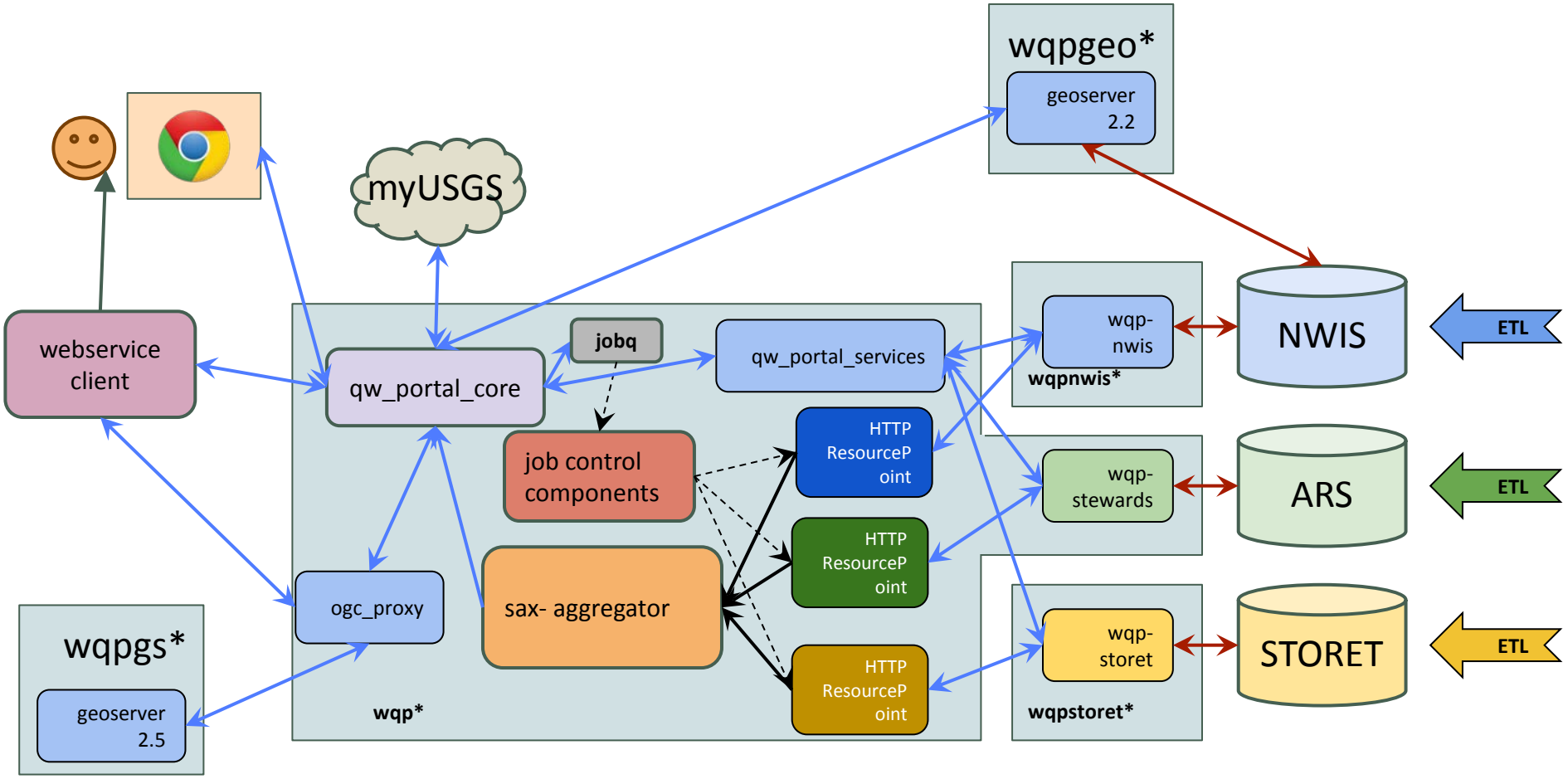
Web services drive everything

- Focus on high performance
- Do not limit query size or complexity
- Build clients to support core use cases
  - Javascript- Web Portal
  - Scientists
    - R
    - Python

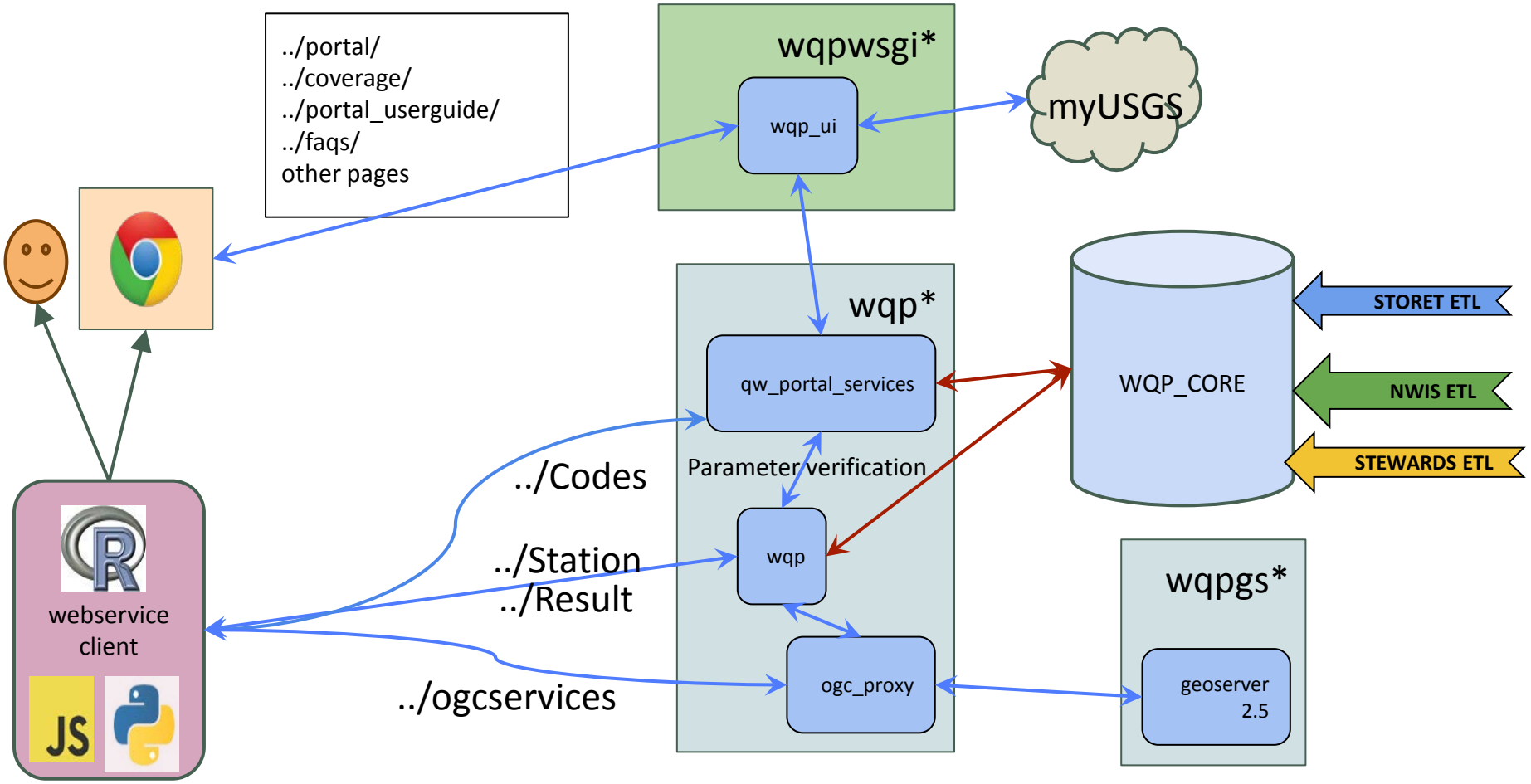
What have we done recently?

- Completely revamped underlying services
- Added Biological data
- Added new query parameters
- Searchable site pages
- Upstream-downstream- coming soon!

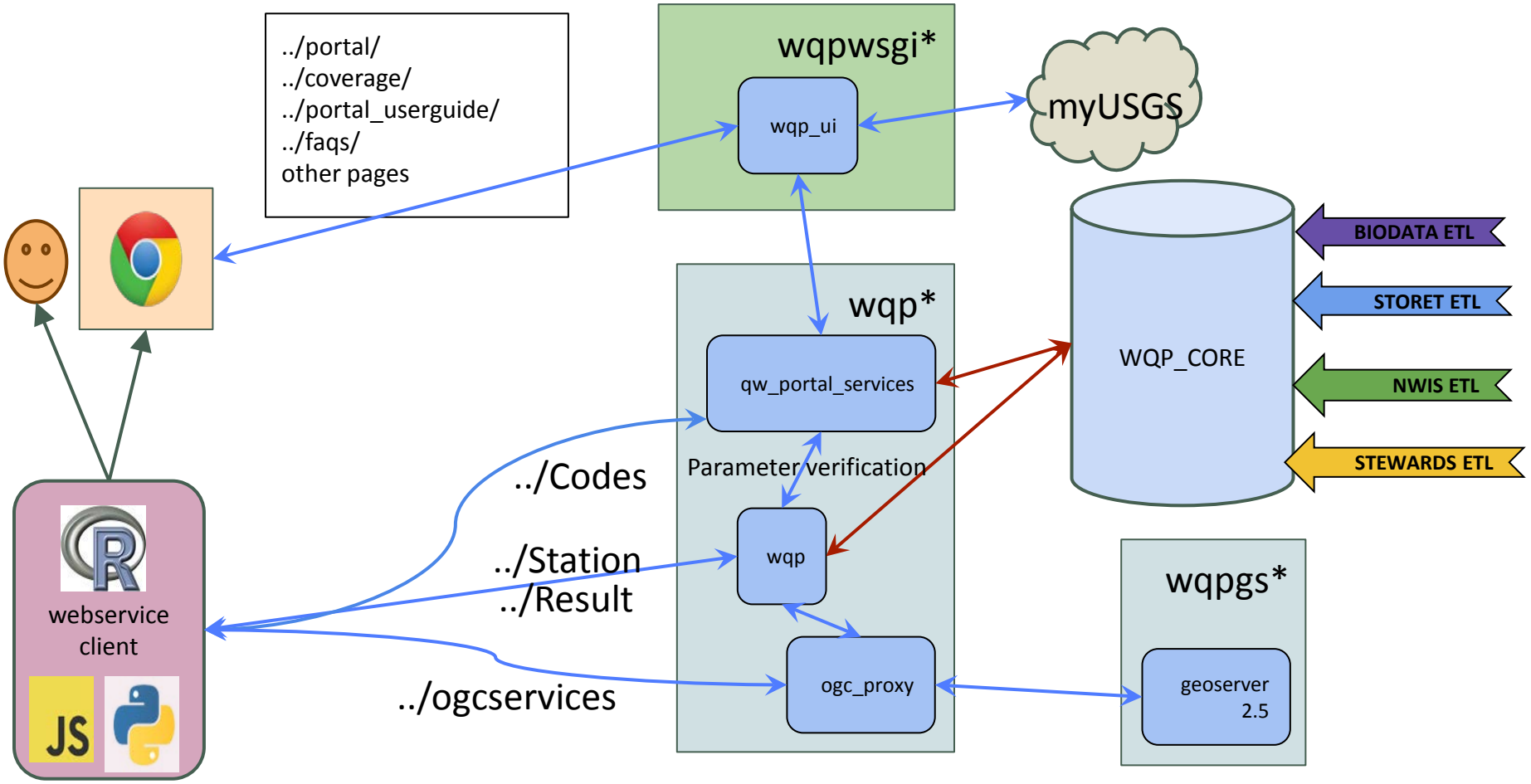
Architecture Before



Architecture After



Adding Biodata was trivial

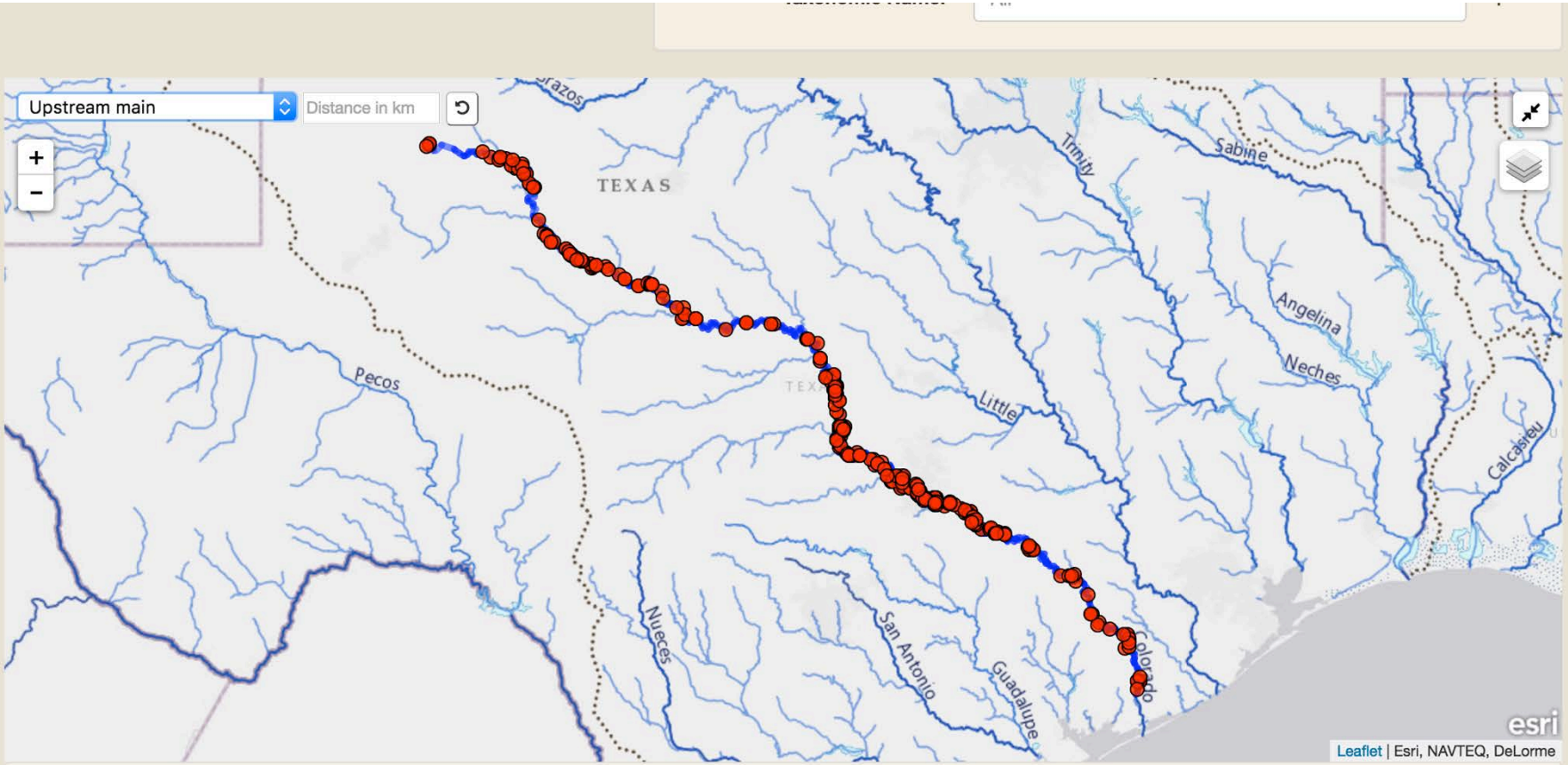


Additional Query Parameters

- Sampling Parameters
  - Project ID
  - Biological Parameters
    - Assemblage
    - Taxonomic Name
- Site Parameters
  - Coming soon!
    - Upstream-downstream queries

ProjectID: EPABEACH

Upstream-downstream queries

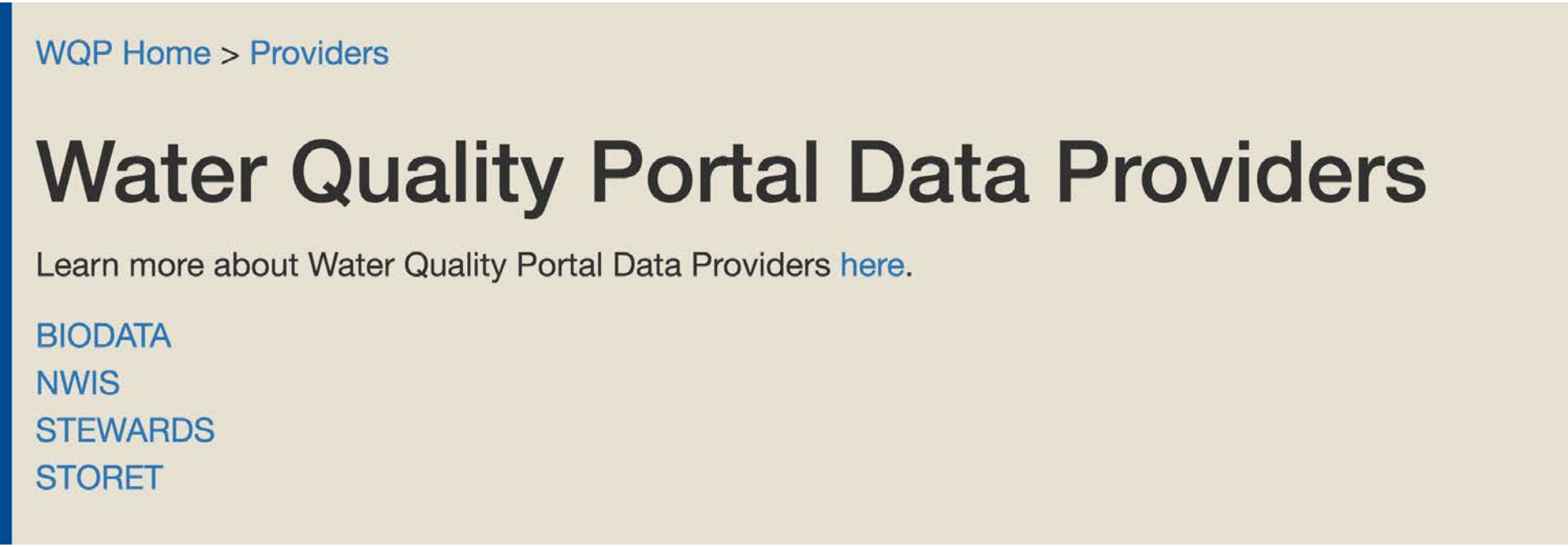


Bringing WQP Content into the light

- Water Quality Portal is part of the “Deep Web”
- Only can get WQP data if you know it is there
- Goal:
  - Expose Water Quality Portal content to search engines
  - Make data findable to both humans and robots

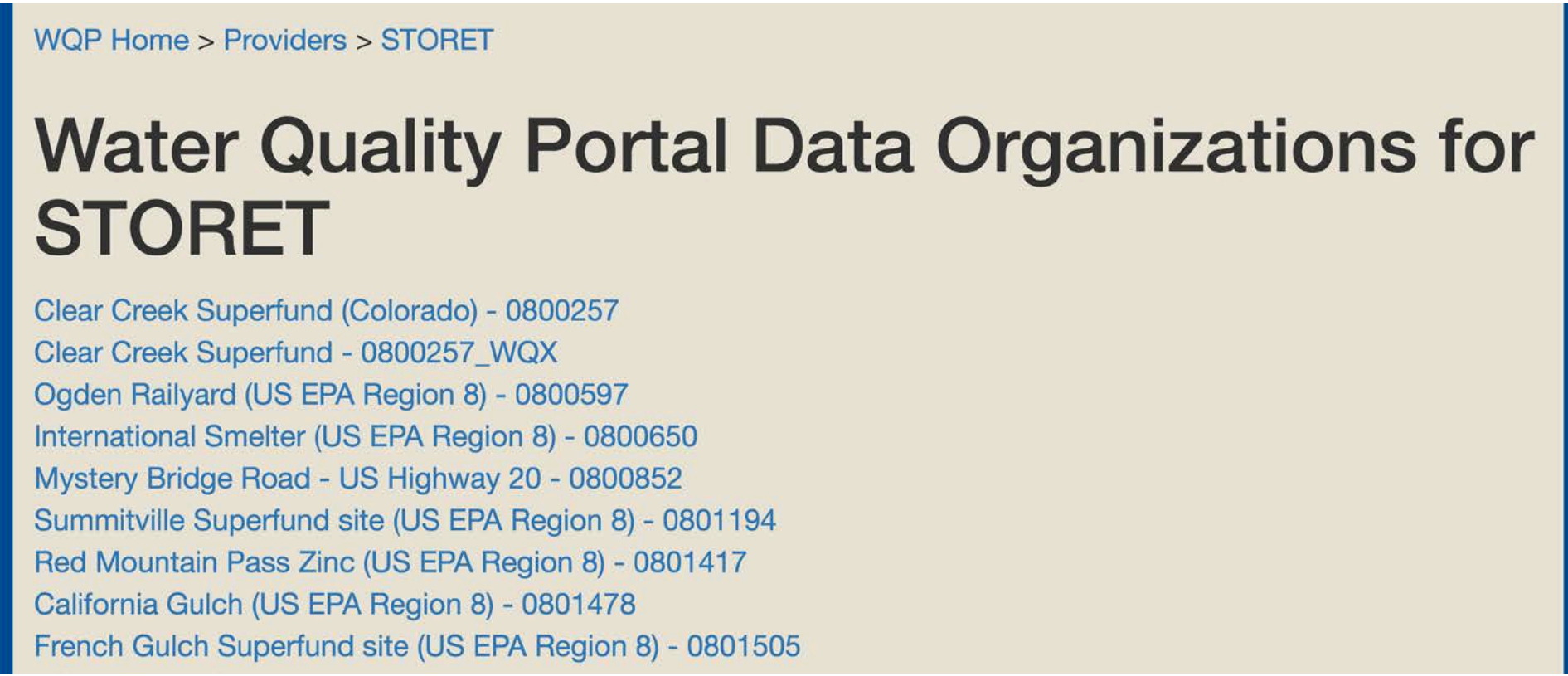
http://www.waterqualitydata.us/provider/

Bringing WQP Content into the light



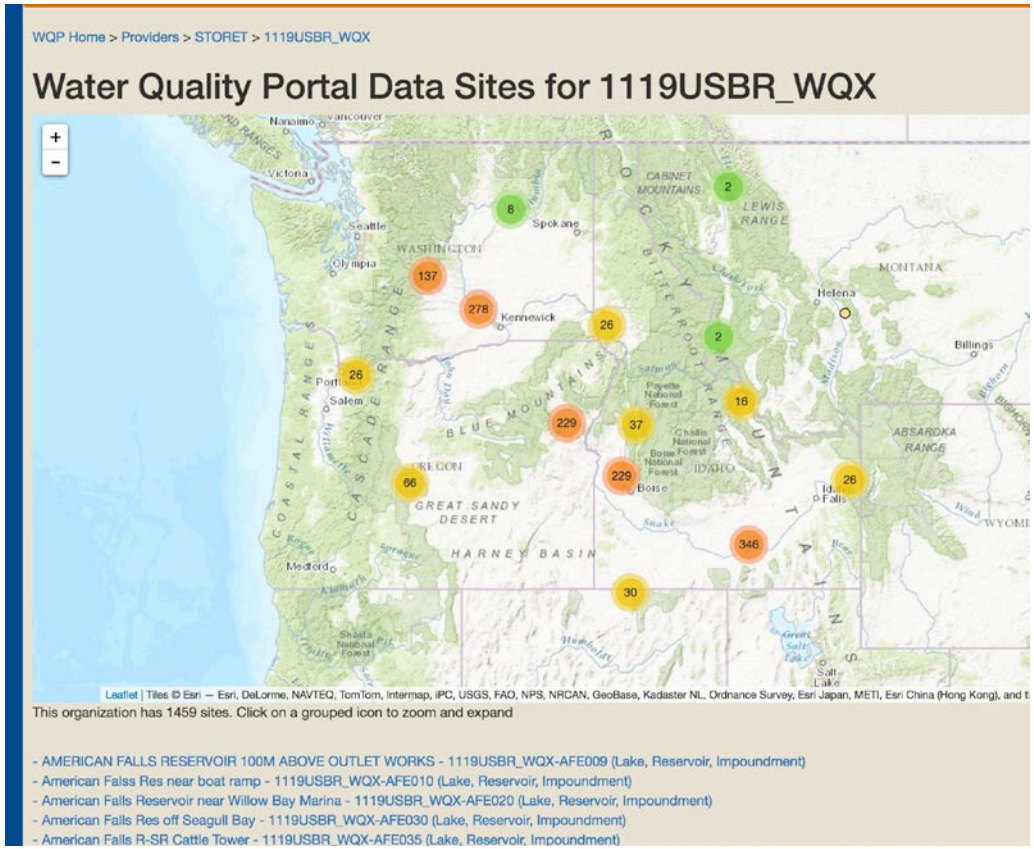
http://www.waterqualitydata.us/provider/

Bringing WQP Content into the light



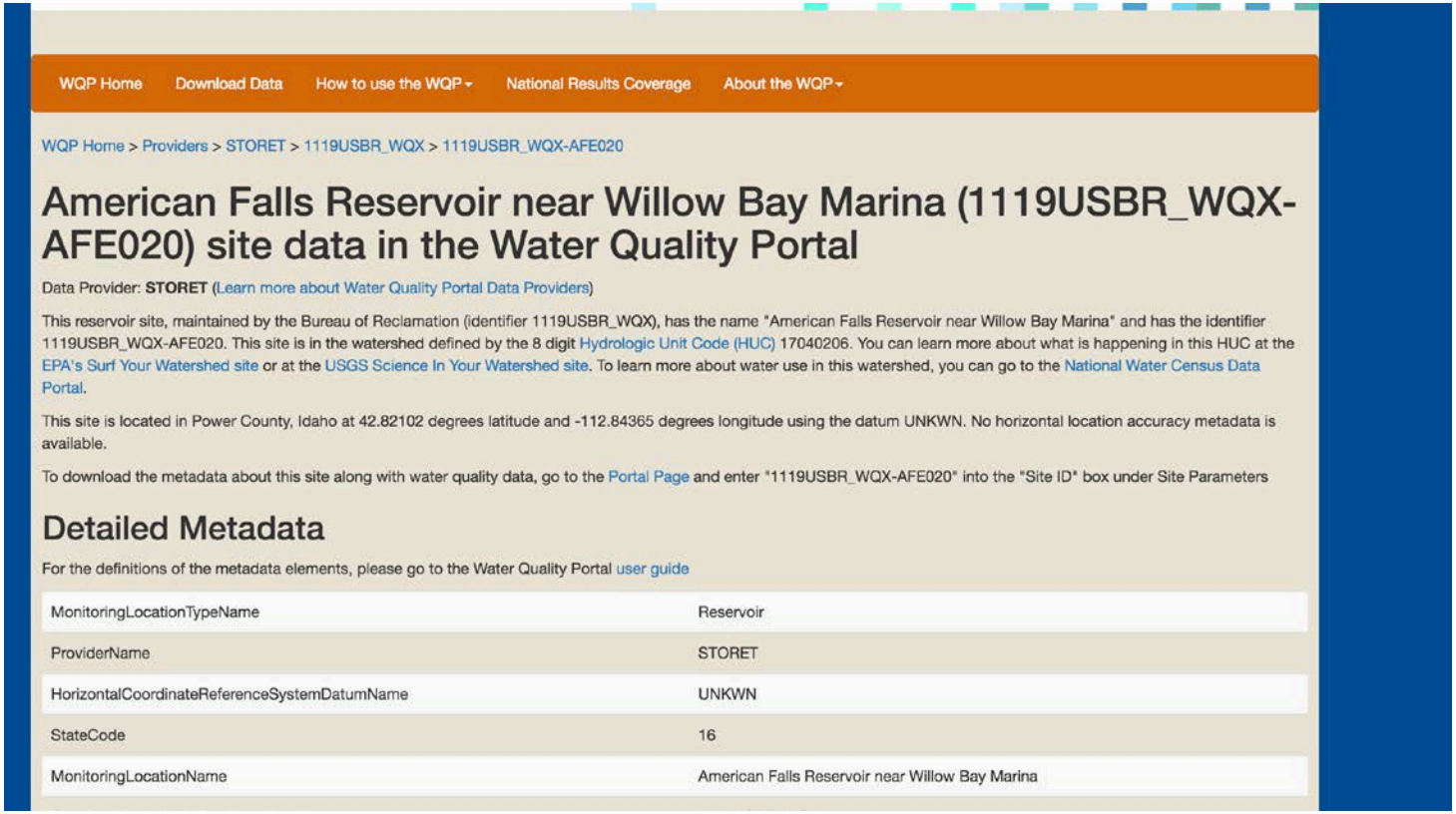
http://www.waterqualitydata.us/provider/STORET

Bringing WQP Content into the light



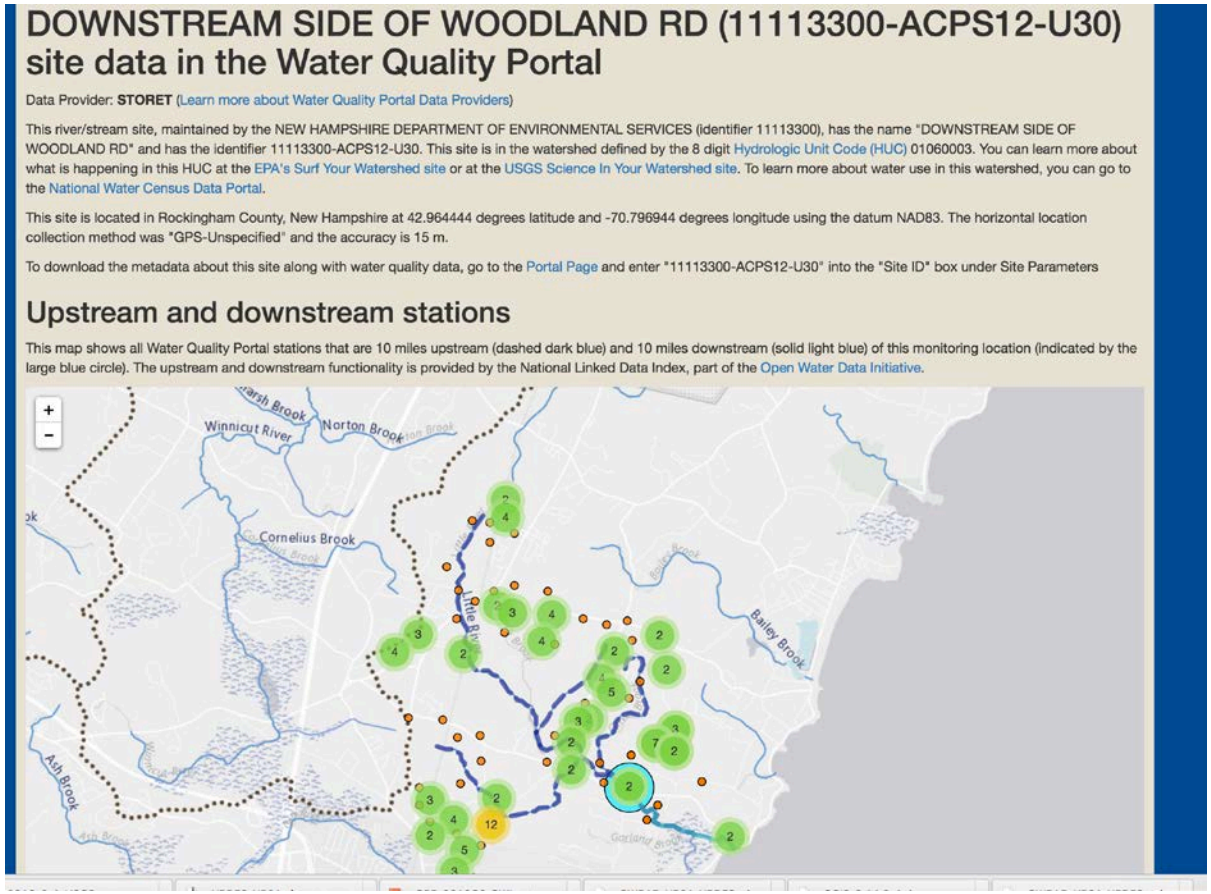
http://www.waterqualitydata.us/provider/STORET/1119USBR_WQX/

Bringing WQP Content into the light



http://www.waterqualitydata.us/provider/STORET/1119USBR_WQX/1119USBR_WQX-AFE020/

Bringing WQP Content into the light



DOWNSTREAM SIDE OF WOODLAND RD (11113300-ACPS12-U30) site data in the Water Quality Portal

Data Provider: **STORET** (Learn more about Water Quality Portal Data Providers)

This river/stream site, maintained by the NEW HAMPSHIRE DEPARTMENT OF ENVIRONMENTAL SERVICES (identifier 11113300), has the name "DOWNSTREAM SIDE OF WOODLAND RD" and has the identifier 11113300-ACPS12-U30. This site is in the watershed defined by the 8 digit Hydrologic Unit Code (HUC) 01060003. You can learn more about what is happening in this HUC at the EPA's Surf Your Watershed site or at the USGS Science In Your Watershed site. To learn more about water use in this watershed, you can go to the National Water Census Data Portal.

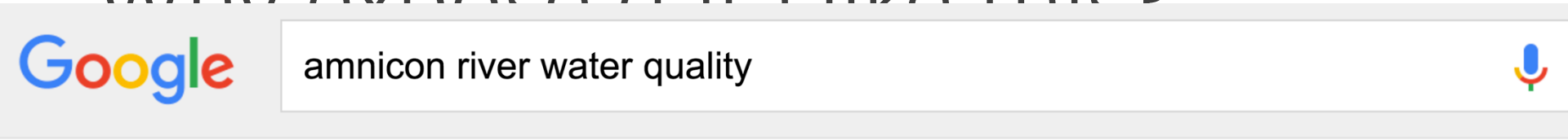
This site is located in Rockingham County, New Hampshire at 42.964444 degrees latitude and -70.796944 degrees longitude using the datum NAD83. The horizontal location collection method was "GPS-Unspecified" and the accuracy is 15 m.

To download the metadata about this site along with water quality data, go to the Portal Page and enter "11113300-ACPS12-U30" into the "Site ID" box under Site Parameters

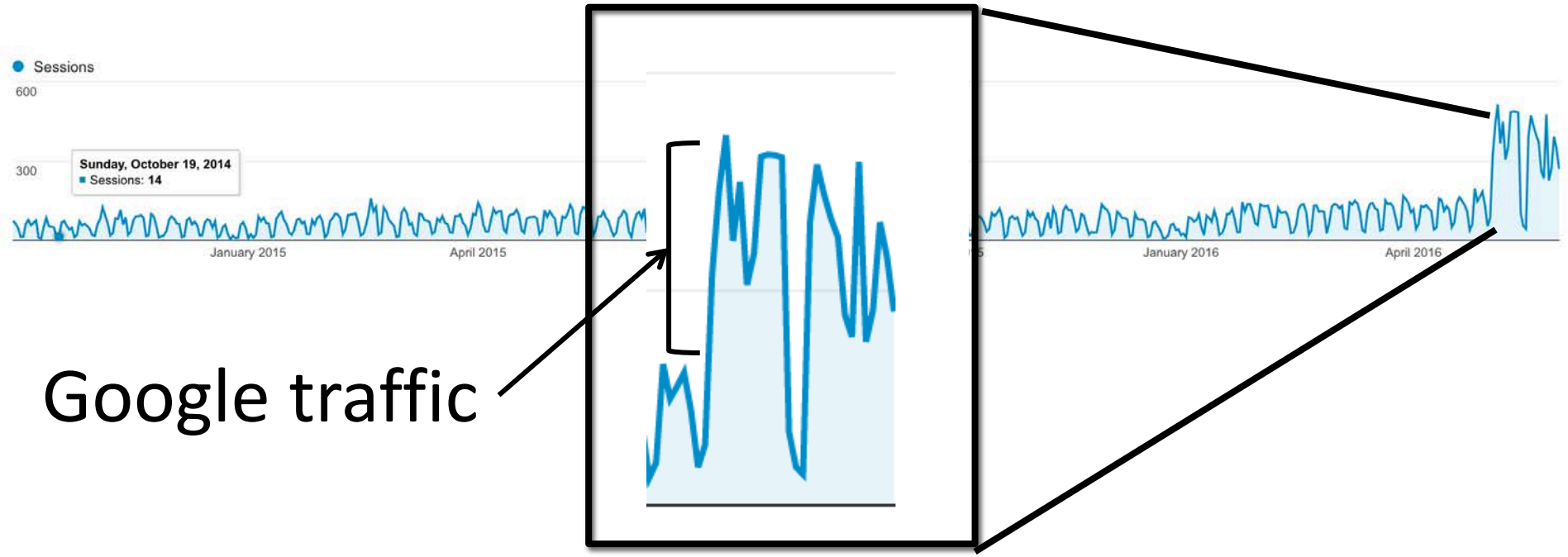
Upstream and downstream stations

This map shows all Water Quality Portal stations that are 10 miles upstream (dashed dark blue) and 10 miles downstream (solid light blue) of this monitoring location (indicated by the large blue circle). The upstream and downstream functionality is provided by the National Linked Data Index, part of the Open Water Data Initiative.

Why expose data like this?



Indexing works!



Google traffic

Portal FY16 Activities

Development

- Identify additional data partners and sources
- Examine opportunities to improve data quality
- Discuss continuous data solution

Outreach and planning

- 5 year Strategic Planning exercise
- Highlight new community tools
- Promote data sharing
- Attend meetings
- Conduct webinars

Specific Activities

- *Education and outreach*:
  - Promote the use of WQP and web services
  - Help advertise user-developed tools which utilize the Portal web services

- *Expand data types and formats*:
  - Serve out biological, habitat, metrics, index
  - Explore continuous monitoring or sensor data

- *Enhance geospatial capabilities*:
  - Develop an OGC-compliant web map service
  - Integrate the mapping interface with the National Hydrography Dataset (NHD)*Plus* flow volume and direction attributes
  - Take advantage of in-development NWIS mapping tools for Gage information.

FY16 Activities

Examine opportunities to improve data quality

- Identify common issues in data quality
- Develop dynamic reporting mechanism to increase visibility of data issues so that they can be fixed

- *Add new data sources*:

- Identify new data partners
- Increase state data contributions

- *Enhance Discoverability*

- Build summary pages for collections of data
  - HUC
  - Organization
  - State
- Make WQP data visible to search engines

Long term

- Support EPA and USGS enterprise data access
- Follow strategic planning
- More data providers and partners
- Deep linkage with the Open Water Data Initiative
  - WQ Sensor data
  - Flow data
  - Remote Sensing?

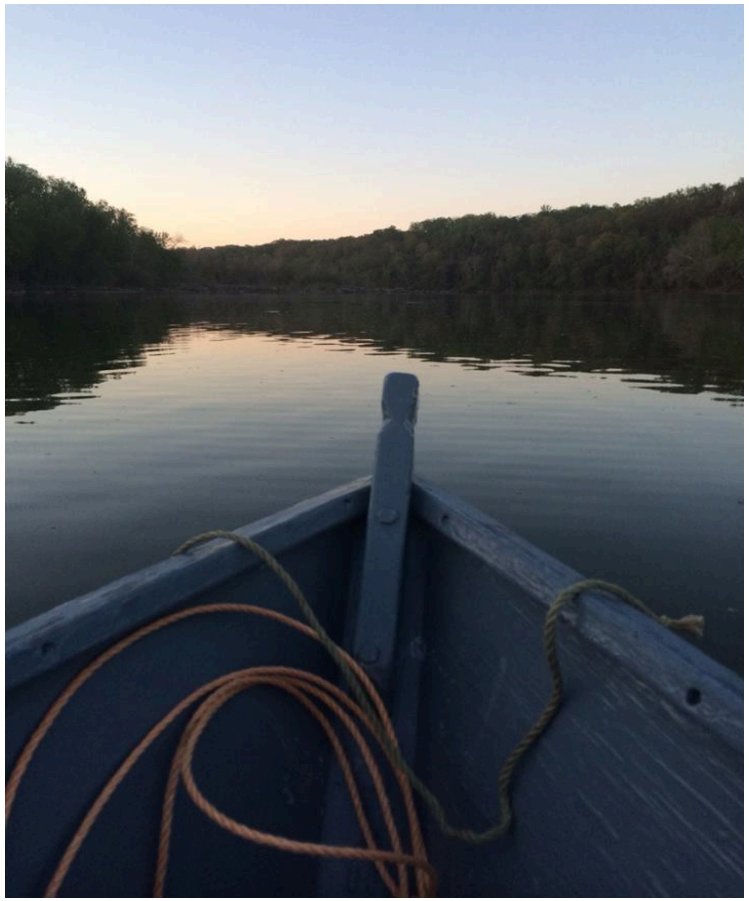
How Can WQP and IOOS start to work together?

- WQP has 42,400,133 sample results from 107,952 Estuary and Ocean sites
  - Link WPP to IOOS?
- WQP focus is on discrete data, and discrete metadata
  - IOOS discrete chemical/physical data?
  - IOOS biological data?
- IOOS integration with Open Water Data Initiative?
-

Questions

Jim Kreft
jkreft@usgs.gov
608-821-3919

Dwane Young



Environmental Chemistry Names

USGS Parameter Codes are mapped to SRS Names

http://www.waterqualitydata.us/public_srsnames/

NWIS Public SRS Names: October 2015

parm_cd	description	characteris...	measureun...	resultsamp...	resulttemp...	resultstatis...	resulttimeb...	resultweigh...	resultpartic...	last_rev_dt
00004	Stream widt...	Instream fea...	ft							2008-02-21
00010	Temperature...	Temperature...	deg C							2008-02-21
00011	Temperature...	Temperature...	deg F							2010-06-28
00020	Temperature...	Temperature...	deg C							2010-06-28
00021	Temperature...	Temperature...	deg F							2010-06-28
00025	Barometric ...	Barometric ...	mm/Hg							2008-02-21
00030	Incident sol...	Solar irradia...	cal/cm2/d							2012-07-02
00032	Cloud cover...	Cloud cover...	%							2008-02-21
00034	Depth to 1 p...	Light attenu...	ft							2012-07-02
00036	Wind directi...	Wind directi...	Deg							2009-09-24
00042	Altitude, fee...	Altitude	ft abv MSL							2009-09-24
00045	Precipitation...	Precipitation	in	Total						2009-09-24

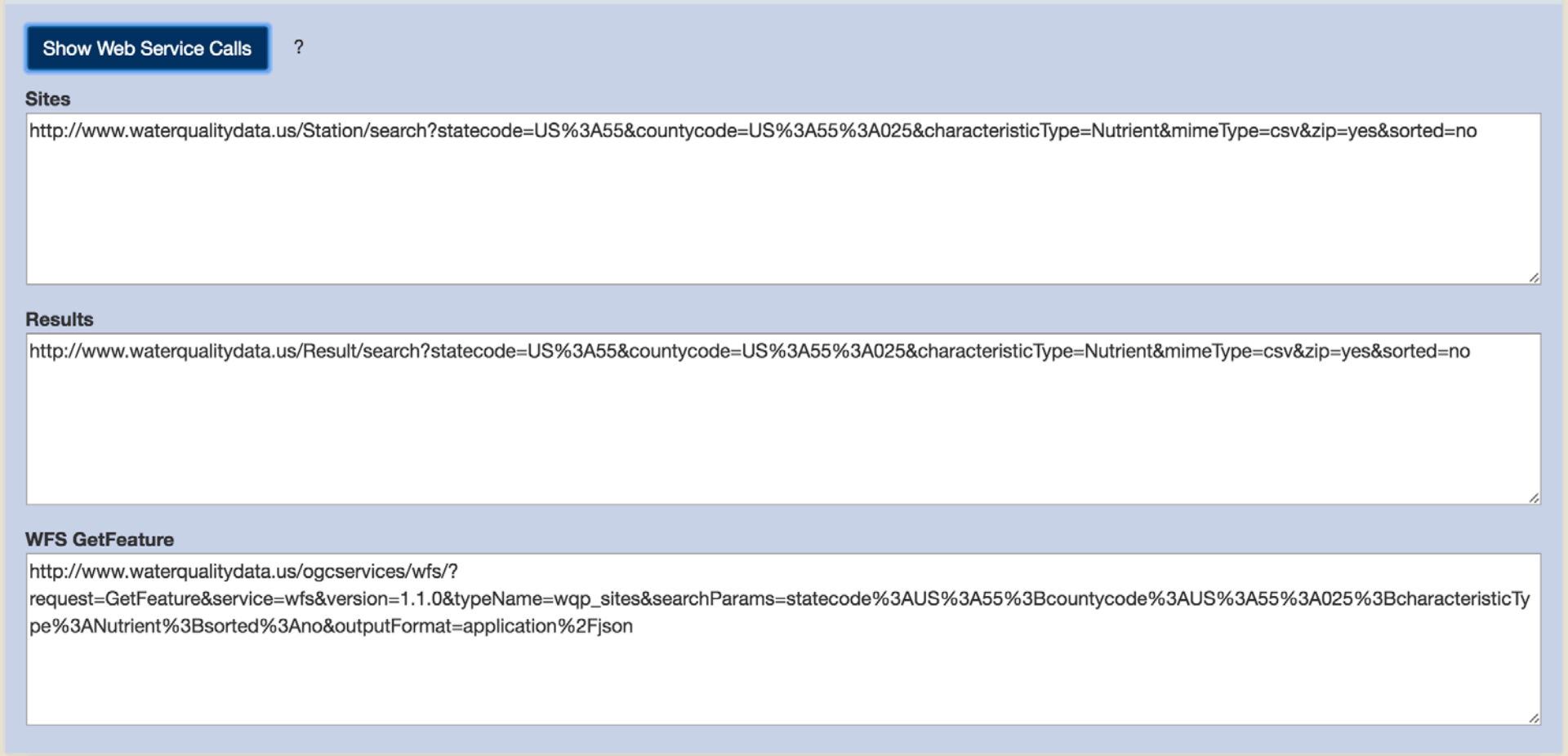
Issues with NWIS-SRS Mapping

- Parameters are not available in SRS
- Sometimes take a while to get new parameters into the mapping- takes a human
- Parameter code control is not absolute- plenty of overlap in parameter codes

Just a name is often not enough.

- Method
  - Collection
  - Analytical
  - Quality Assurance
- Fraction
- Medium
- Units
- Biological data
  - Name
  - Indexes
- Categorization
  - Chemical → Nutrient → Nitrogen →Inorganic nitrogen (nitrate and nitrite)

Web Service Calls



What is a client?

Bridge between the an API and an external tool.

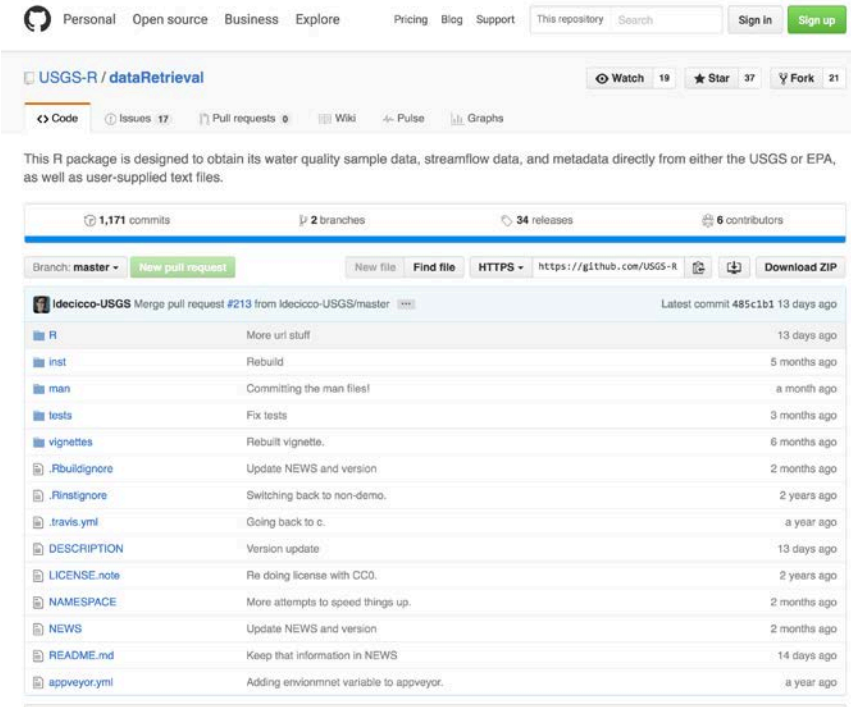
Allows people familiar with a given software tool to easily get started without having to first figure out the WQP API

Water Quality Portal has two clients:

- DataRetrieval for R
- pywqp for Python

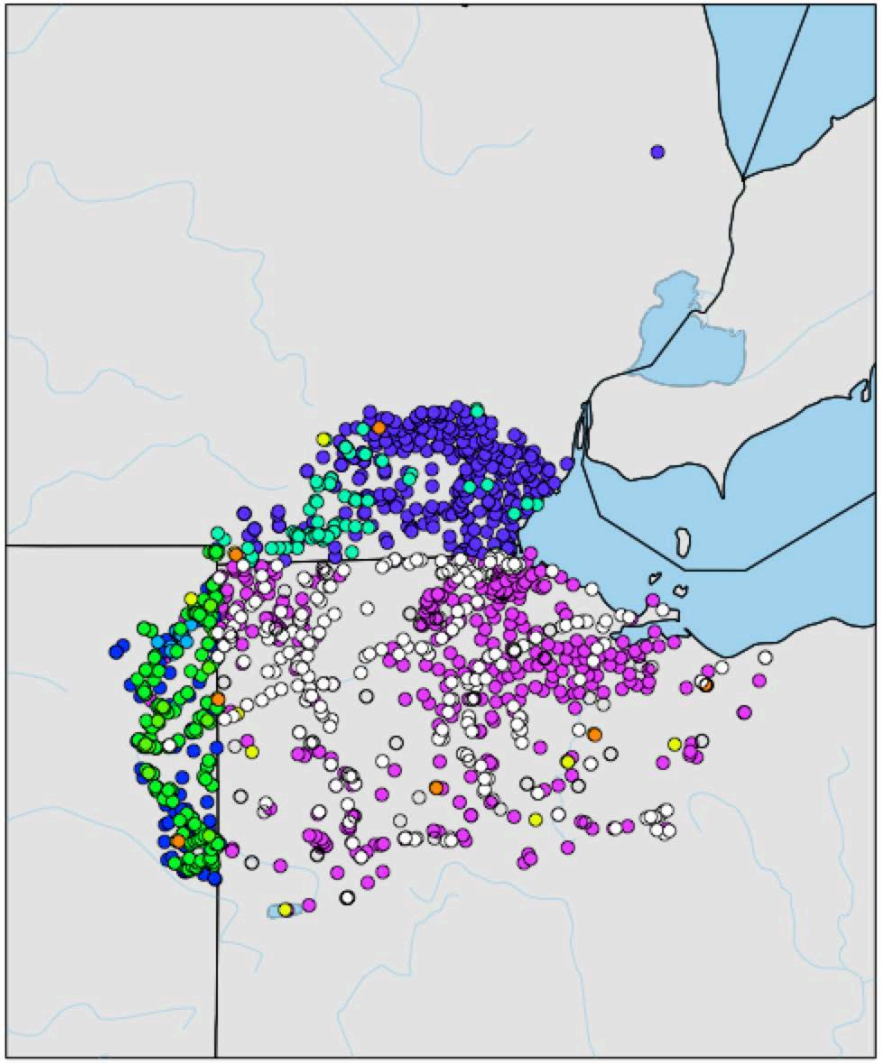
Water Quality Portal R Client

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

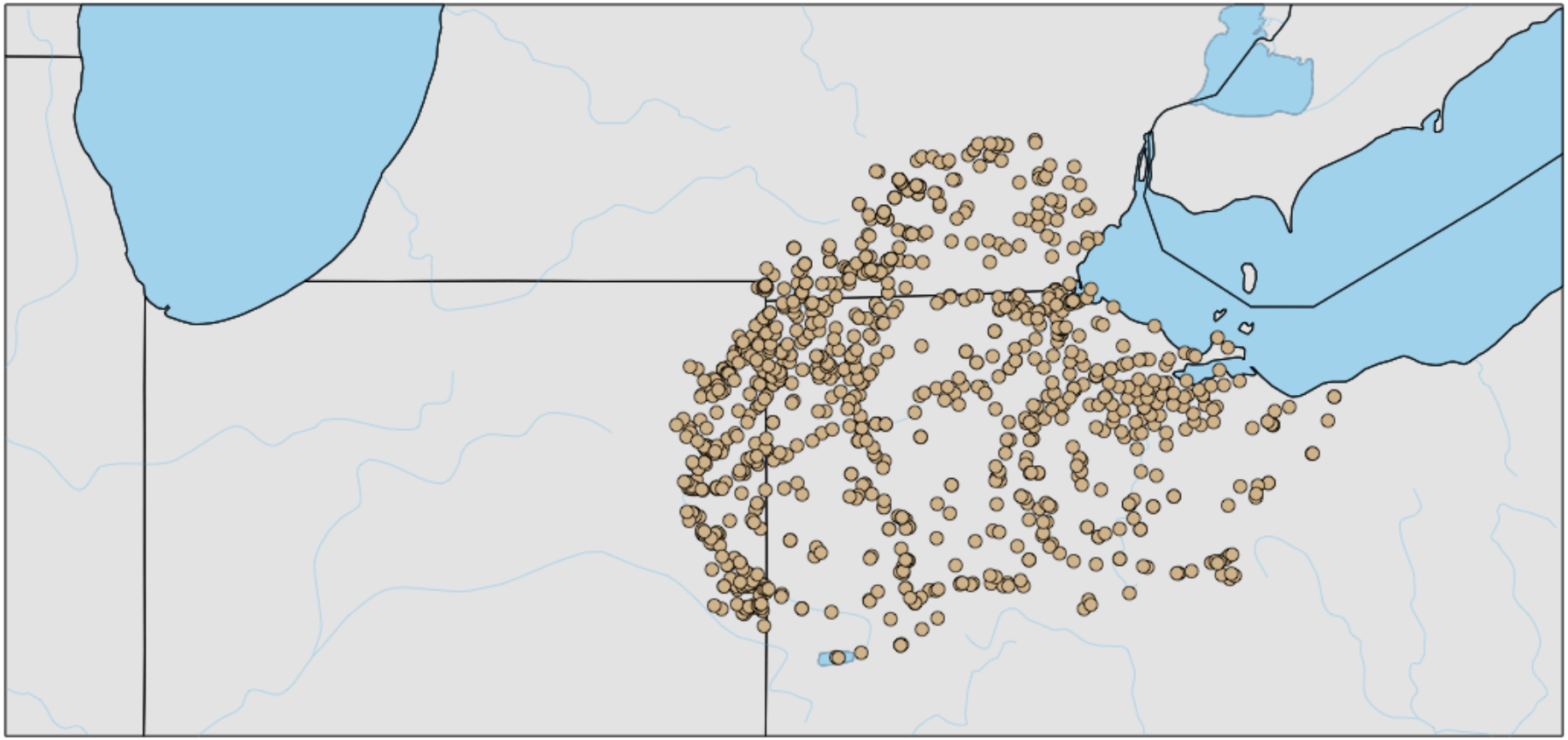


What agencies have sampled for nutrients in the western Lake Erie basin?

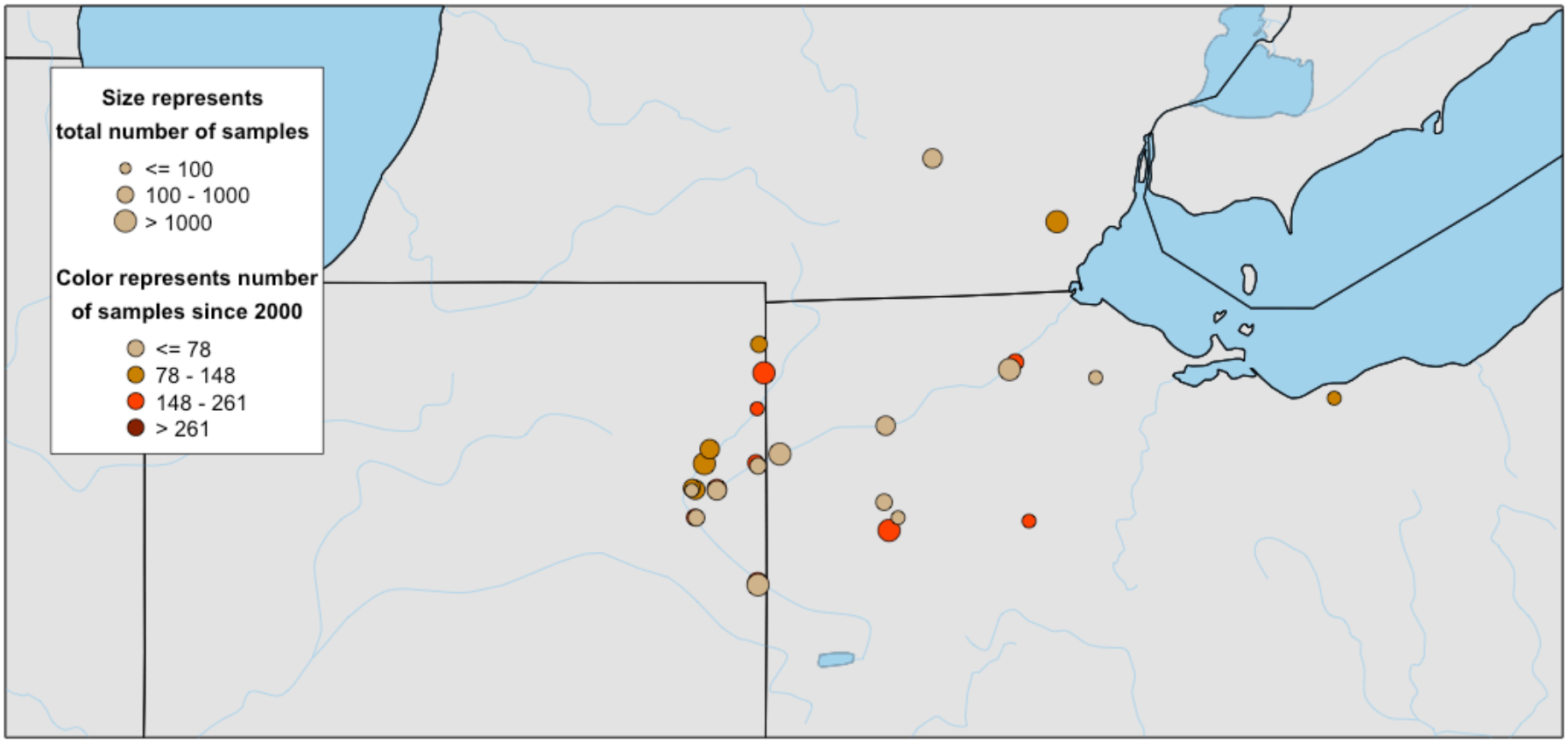
	OrganizationFormalName	count
1	USGS Ohio Water Science Center	427
2	USGS Michigan Water Science Center	351
3	Division of Surface water (Ohio)	221
4	Indiana STORET	201
5	USGS Indiana Water Science Center	107
6	Michigan Department of Environmental Quality	74
7	Division of Drinking and Ground Water (Ohio)	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

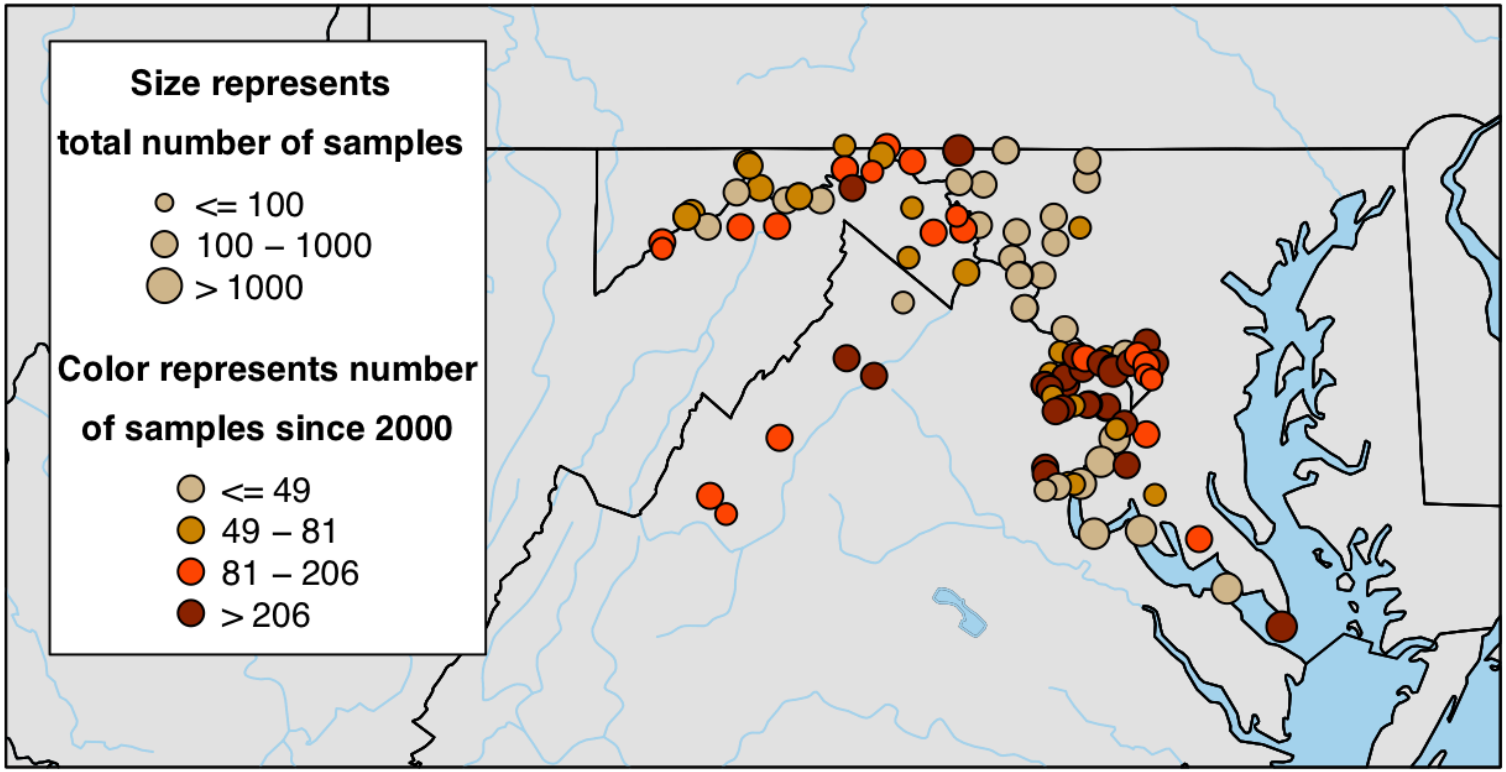
```
data <- readWQPdata(huc="0207*",
characteristicName="Phosphorus”)
```

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

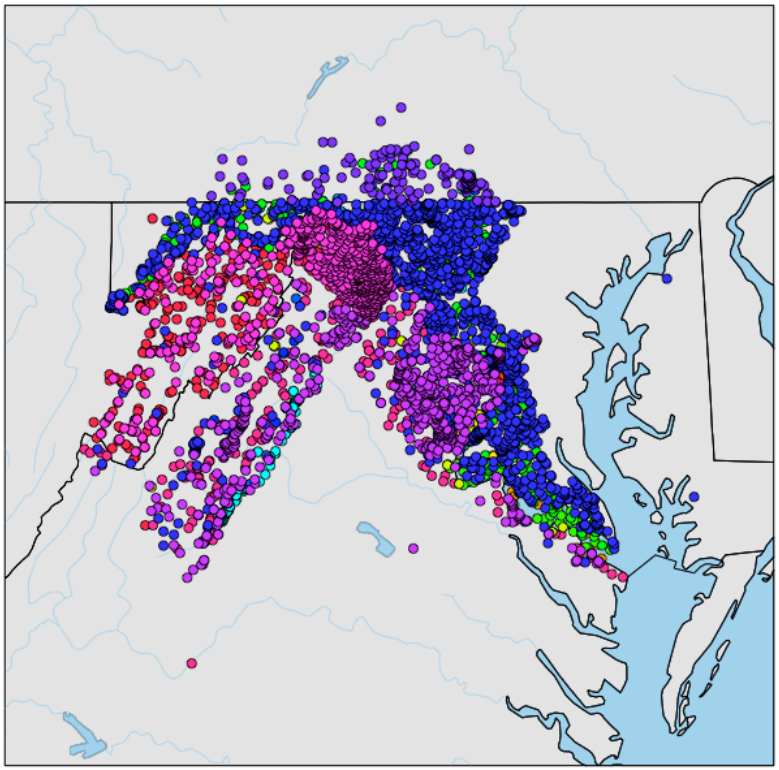
is changed to

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

And we have changed to the Potomac

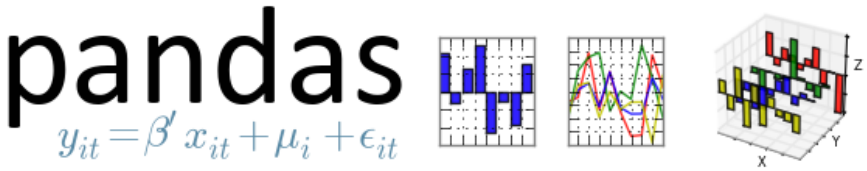


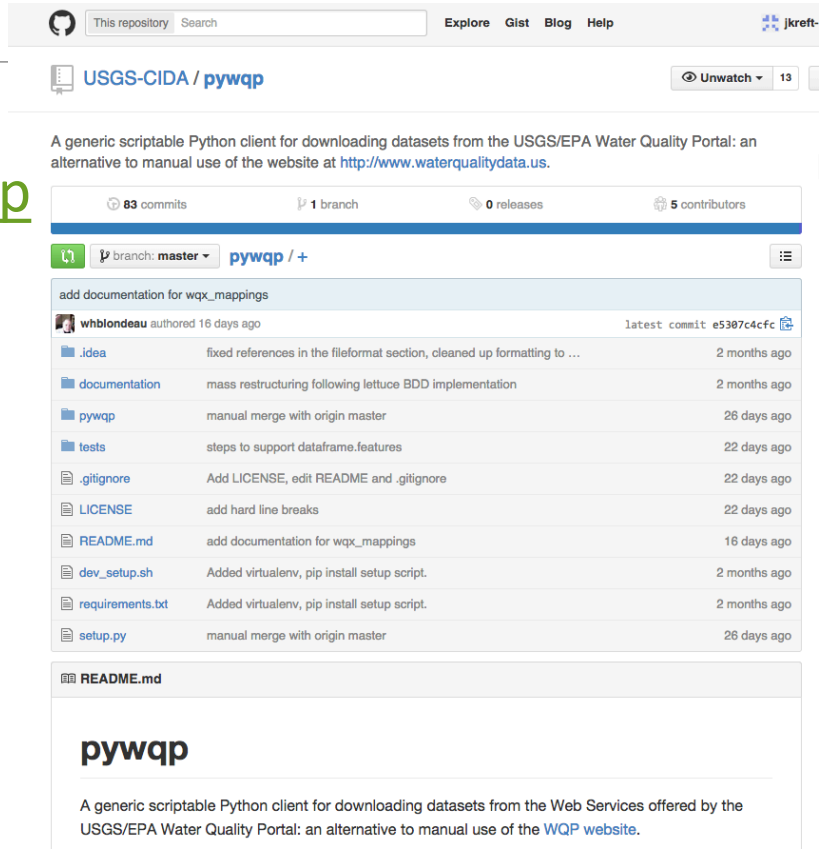
Potomac Organizations and sites that collect nutrient data



Water Quality Portal Python Client

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





Portal FY16 Activities

Development

- Identify additional data partners and sources
- Examine opportunities to improve data quality
- Discuss continuous data solution

Outreach and planning

- 5 year Strategic Planning exercise
- Highlight new community tools
- Promote data sharing
- Attend meetings
- Conduct webinars

Mission and Vision

Vision

Be the premiere source for water quality data for everyone, everywhere.

- Vision: what is our desired end state? What are the major issues or problems? What would success look like?

Mission

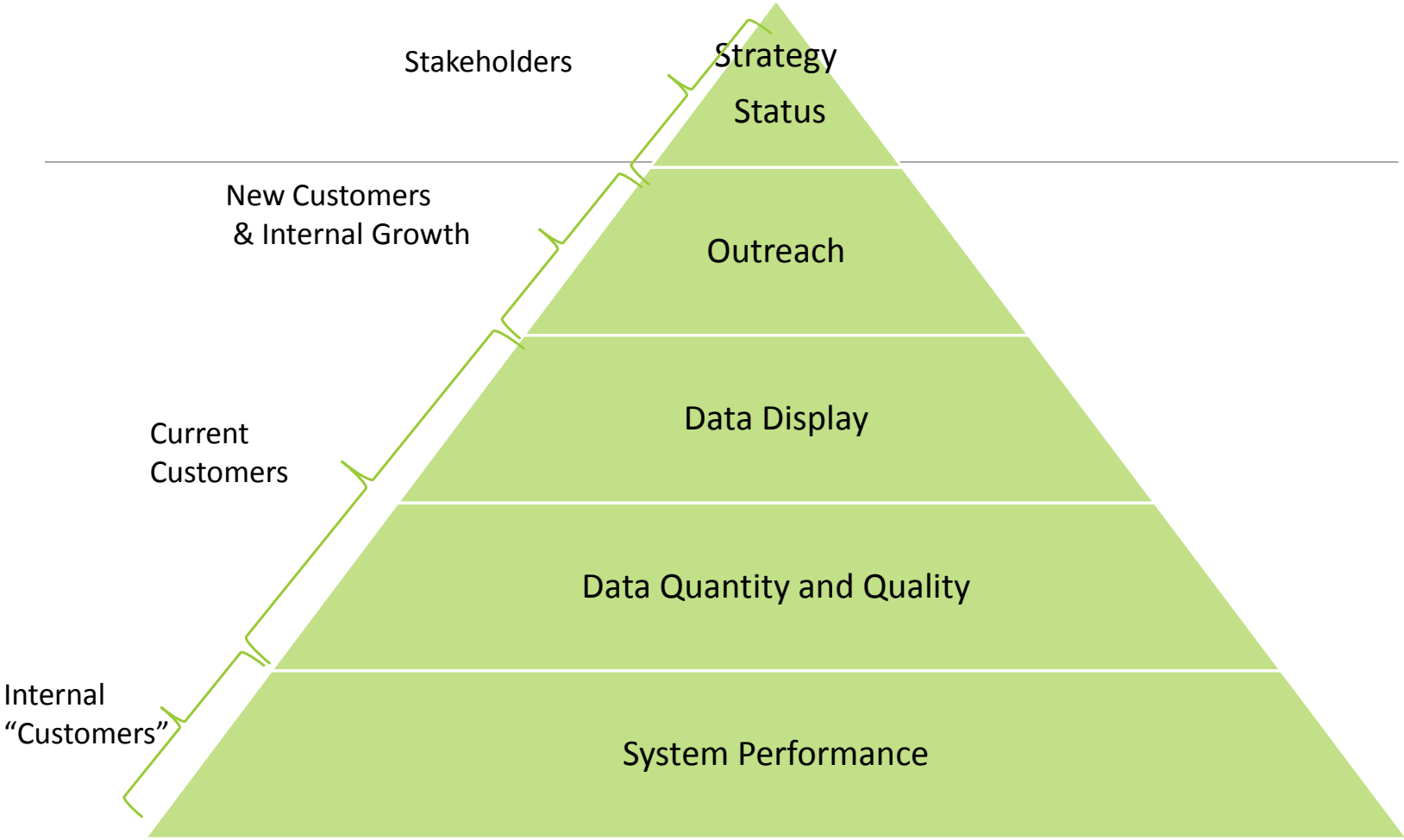
Provide easy access to all water quality data, facilitate improvements in data quality, and enhance data discovery and data summaries to inform sound water-quality decision making at local, state, regional, and national scales.

- Mission: Why we exist, Who are we, what do we do, why do we do this work and for whom?

Scope

Water quality data collected from discrete samples of ambient surface and groundwater in the United States.

- Scope: What types of data, and where, will we focus on?



Relationship of 5 Portal Strategy Sections to Customer Groups