TADA Reference Articles

Welcome to TADA: Tools for Automated Data Analysis!

lifecycle experimental

R-CMD-check passing

Tools for Automated Data Analysis, or TADA, is being developed to help States, Tribes (i.e., Tribal Nations, Pueblos, Bands, Rancherias, Communities, Colonies, Towns, Indians, Villages), federal partners, and any other Water Quality Portal (WQP) users (e.g. researchers) efficiently compile and evaluate WQP data collected from water quality monitoring sites. TADA is both a stand-alone R package, and a building block to support development of the TADA R Shiny application. We encourage you to read this package's LICENSE and README files (you are here).

- How to use TADA:
  - Function Reference
  - Example Workflow 1: Water Quality Portal Data Discovery and Cleaning (Beginner)
  - Example Workflow 2: 2023 Shepherdstown Training (Advanced)
- How to Contribute
  - We encourage stakeholders to test the functionality and provide feedback. Moreover, open source software provides an avenue for water quality data originators and users to develop and share code, and we welcome your contributions! We hope to build a collaborative community dedicated to this effort where TADA users and contributors can discover, share and build the functionality over time.
- More information on how TADA leverages the WQX QAQC Service
- More about the TADA Project

Installation

You must first have R and R Studio installed to use the TADA R Package (see instructions below if needed). You can install and load the most recent version of the TADA R Package on GitHub by running:

```
if(!"remotes"%in%installed.packages()){
install.packages("remotes")
}

remotes::install_github("USEPA/TADA", ref = "develop", dependencies = TRUE)
```

Water Quality Portal (WQP) & Tools for Automated Data Analysis (TADA)

Cristina Mullin PhD & Jesse Boorman-Padgett

Water Data Integration Branch

EPA Office of Water

Wednesday, March 6th, 2024, 3:00pm – 4:30pm

Agenda

- Water Quality Portal (WQP) Overview (15 mins)
- WQP Demo (15 mins)
- Tools for Automated Data Analysis (TADA) Overview (15 mins)
- TADA Demo (45 mins)

Reusable

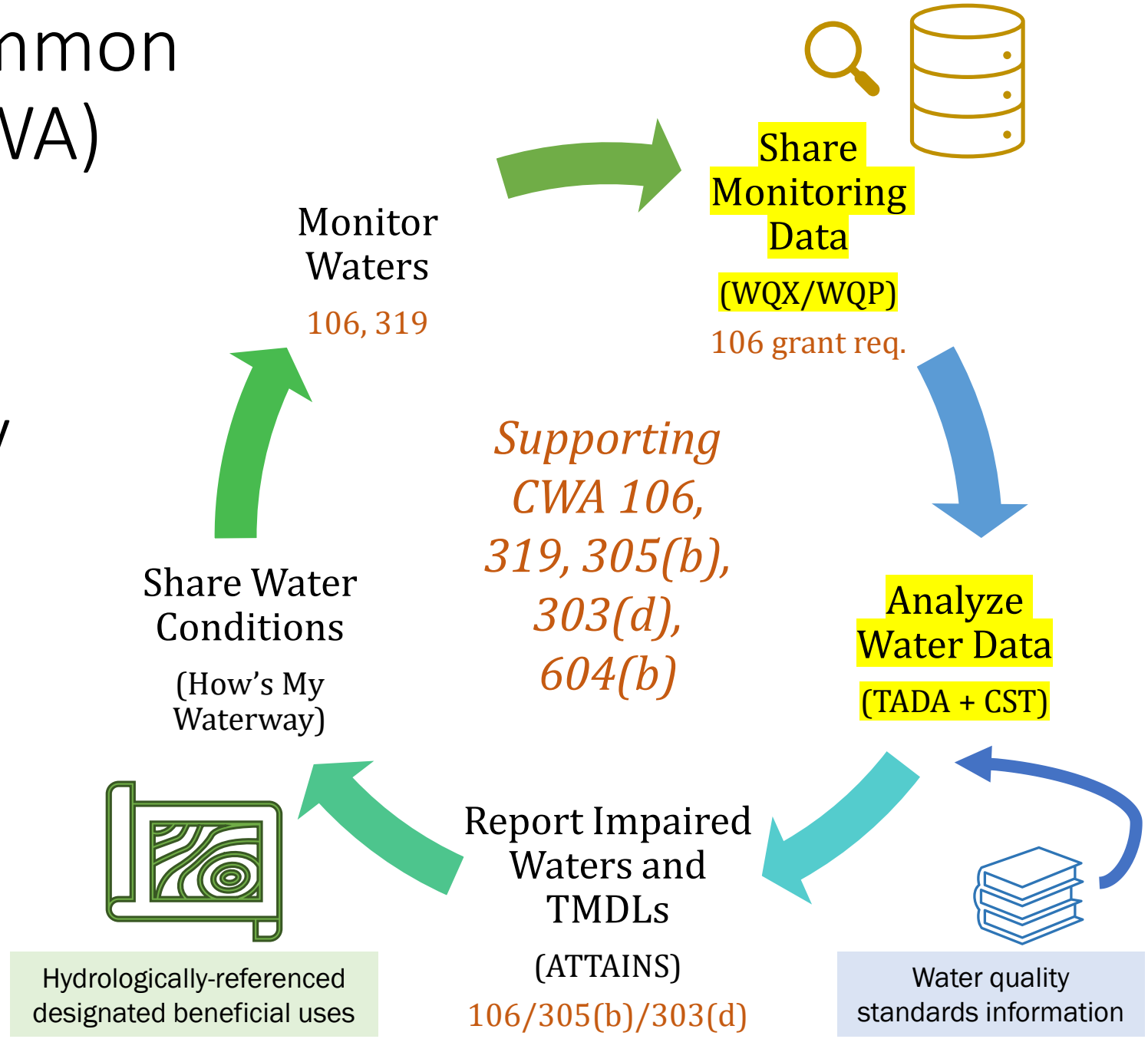
Interoperable

Accessible

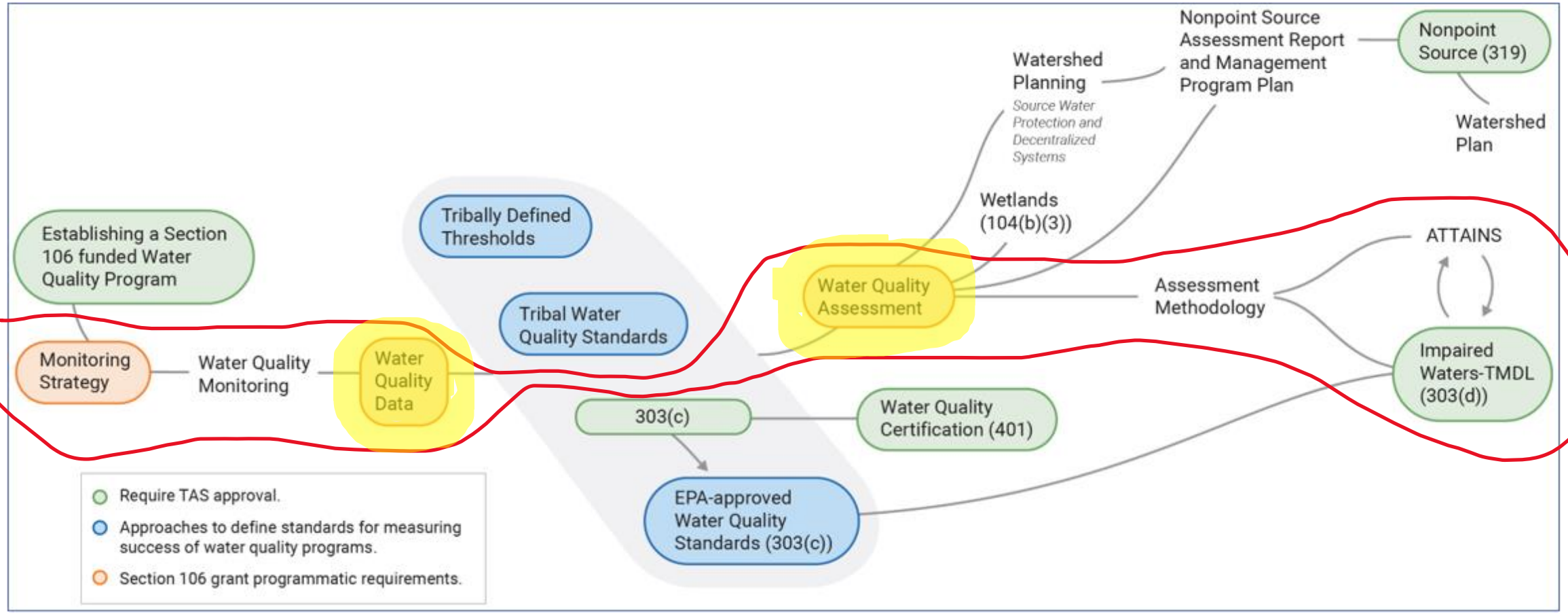
Findable

Helping Answer Common Clean Water Act (CWA) Questions

- Is my water safe?
- Does it meet water quality standards?
  - Aquatic Life
  - Drinking
  - Fishing
  - Recreation

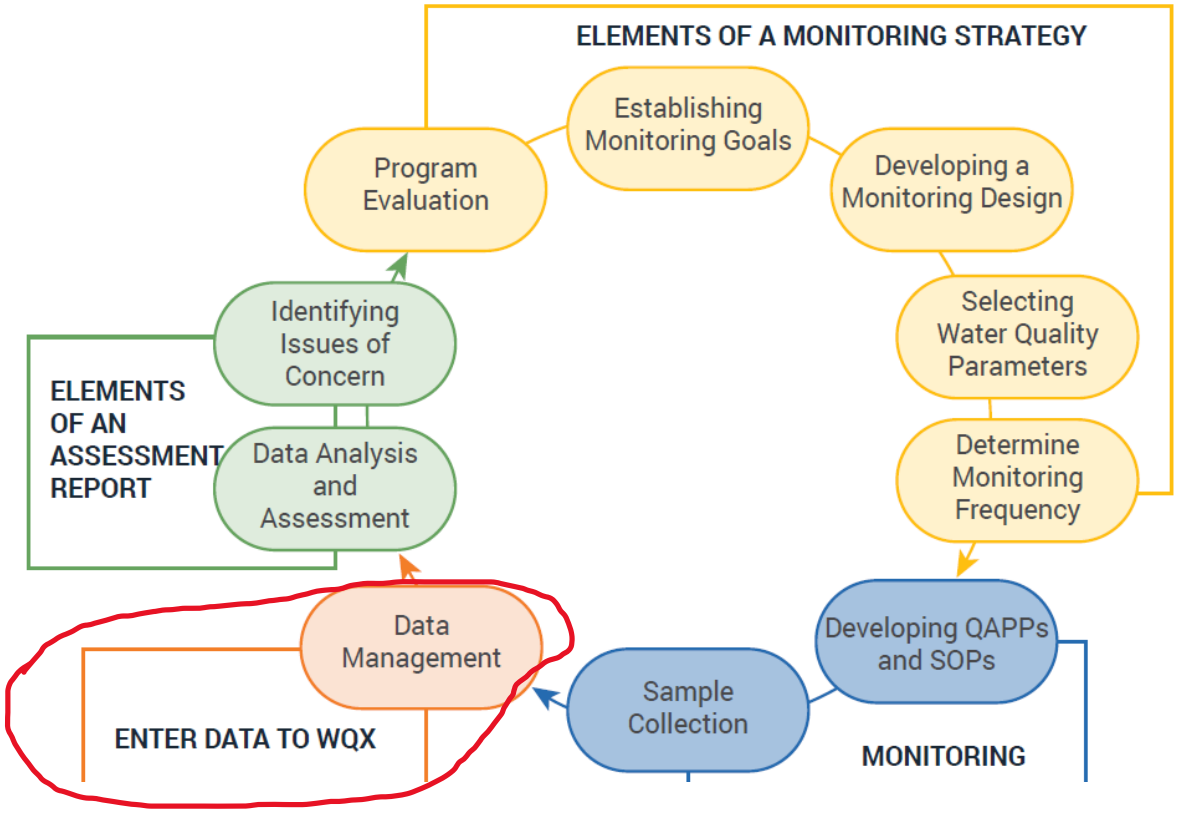


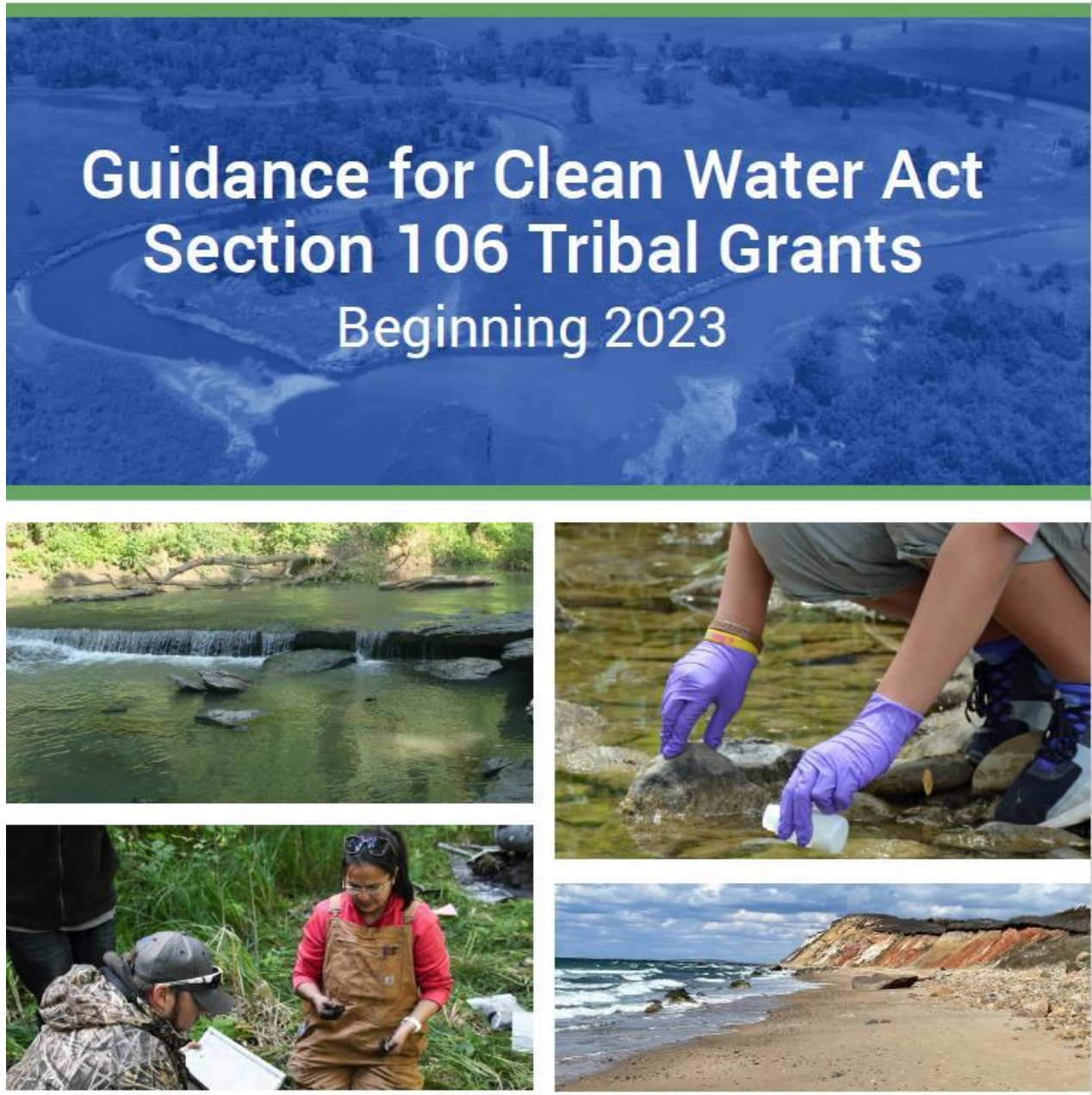
Clean Water Act - Tribal Program Road Map



Study Design → Data Collection / Monitoring → Data Mngmnt / Sharing → Data Analysis / Assessment → Action / Communication

Clean Water Act Section 106 – Data Management Req.





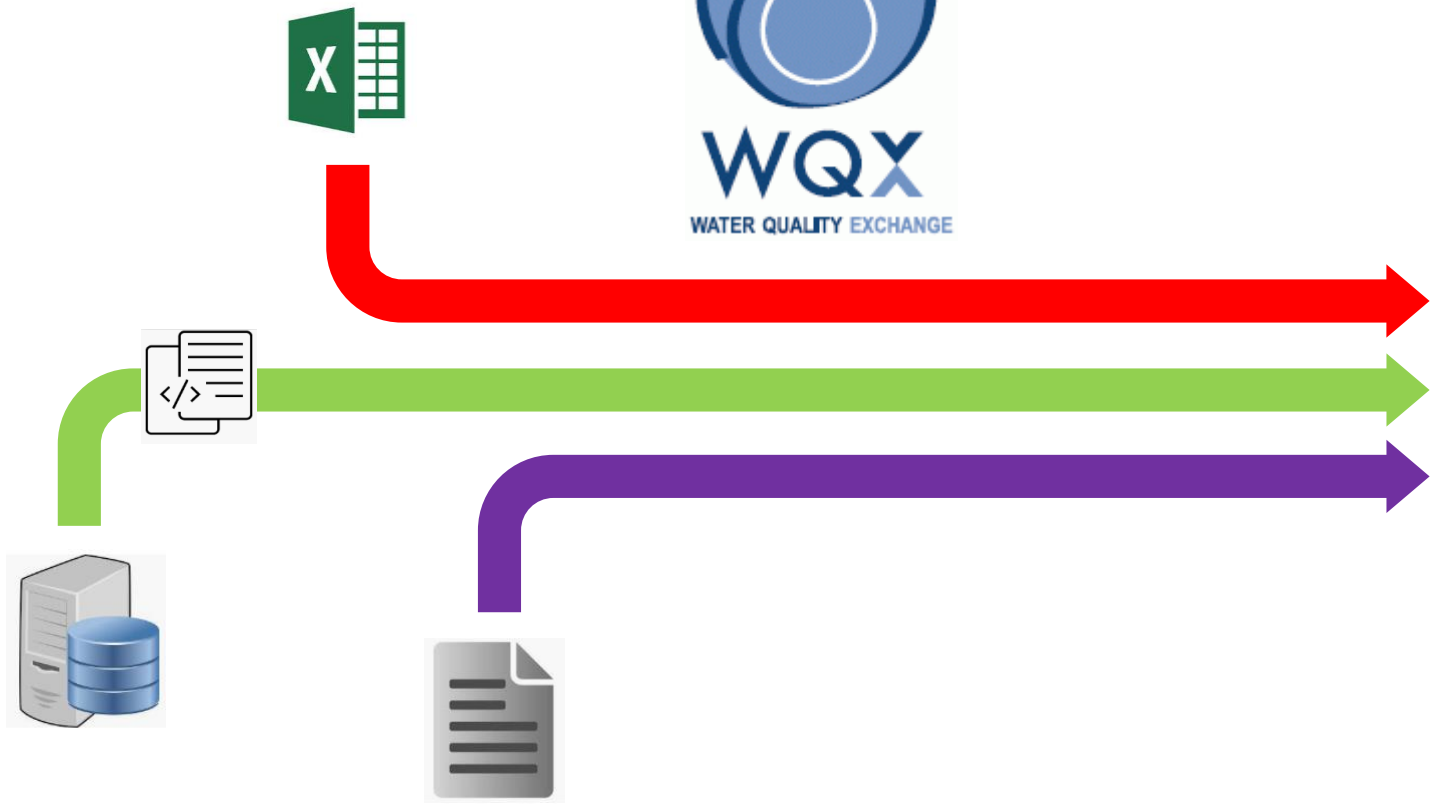
Water Quality eXchange

NATIONAL WATER QUALITY MONITORING COUNCIL

USGS science for a changing world

EPA United States Environmental Protection Agency

WQX WATER QUALITY EXCHANGE



Water Quality Portal

- 424,691,590 sample results
- 1,085,011 sites
- 1,590 organizations

Water Quality Portal (WQP)

EPA United States Environmental Protection Agency

USGS science for a changing world

WQX WATER QUALITY EXCHANGE

dataRetrieval USGS R

Water quality monitoring data is foundational to being able to answer important questions

- Is my water safe?
- Is there enough?

Format is the same for everyone who wants to share data

- Water quality monitoring and data management is complicated
- Standardized, electronic data is more valuable than data in file cabinets (reusable, sharable, discoverable, interoperable, and includes important metadata)

Usable data translates to knowledge, public awareness, and action

- Reuse adds value!
- Supports CWA assessments and other water quality research
- Serves as the backbone for water data tools like HMW

Standardized data formats

Stacked Style Dataset

- AKA “Tall” “Narrow”

All measures in one column

Additional information about those measures

- How data is stored/served by the WQP
- Good for data management
- Allows for metadata

Result Identifier	Characteristic Name	Sample Fraction	Measure Value	Unit
NWIS-114877794	Stream width measure		3	ft
NWIS-114877795	Temperature, water		16.8	deg C
NWIS-114877797	Stream flow, instantaneous		0.19	ft3/s
NWIS-114877798	Specific conductance	Total	696	uS/cm @25C
NWIS-114877799	Acidity, (H+)	Total	0.00001	mg/l
NWIS-114877800	Oxygen	Dissolved	11.4	mg/l
NWIS-114877801	pH	Total	8.3	std units

WQX/WQP data are relational

Data of different types are each managed in their own table

We establish relationships between certain pieces of information in the tables

The related pieces of information are often (but not always) ID or “key” fields

This allows for more detailed information to be stored in separate tables, allowing for useful queries of the database

Ex. This is what allows users to query across place, time, program, and result type in the WQP

All Organizational Data

Org ID	Org Name
001	Minnesota Depart
002	Kentucky Departm
003	University of Was

All Project Data

Org ID	Prjct ID	Project Name
001	CYAN	Cyano Monitoring
001	PRB	Probabalistic mon
001	TRND	Trends Program

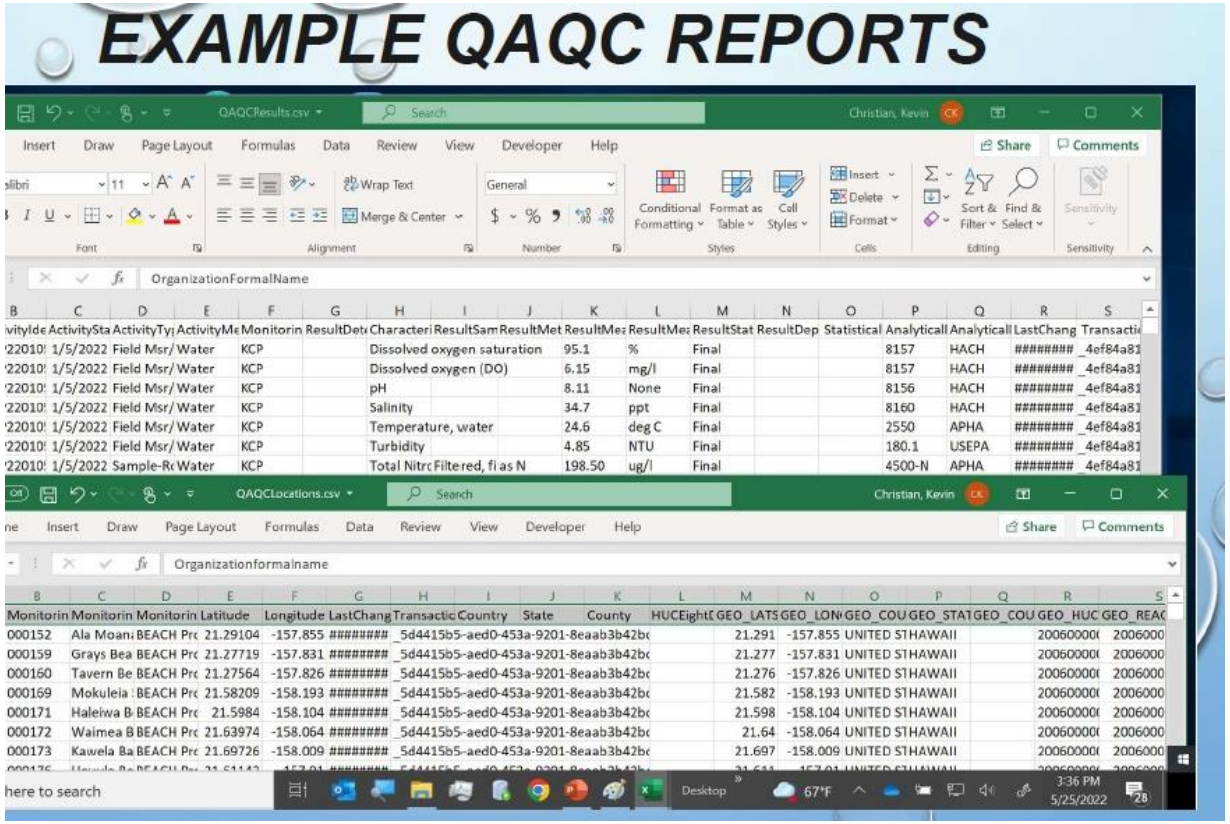
All Location Data

Org ID	Loc ID	Location Name
001	DBL	Deep Blue Lake
001	MR	Muddy Run
001	FC	Fishing Creek

Activities & Results Data & Metadata

Org ID	Proj ID	Loc ID	Result ID	Sample Date	Characteristic	Value	Unit
001	PRB	MR	001	5/3/2012	pH	7.6	SU
001	PRB	MR	002	4/17/2016	Temperature	19.1	Deg C
001	PRB	MR	003	4/23/2020	Conductivity	236	µS/cm

WQX QAQC Service



TADA (0.0.1) Reference Articles

WQX QAQC Service User Guide

TADA Team

2023-09-15

Source: `vignettes/WQXValidationService.Rmd`

On this page

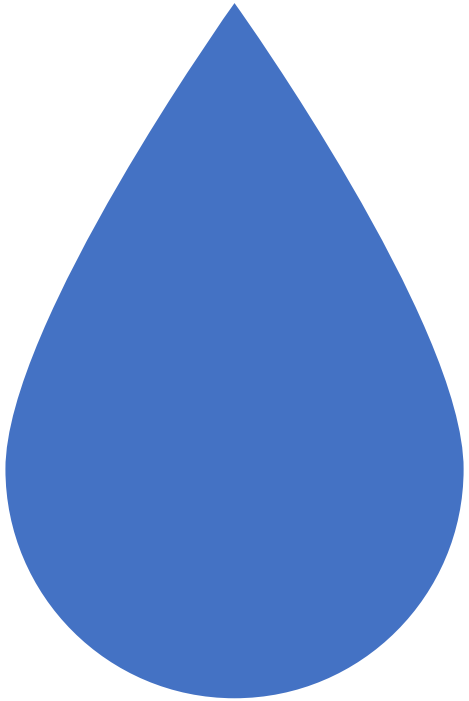
- TADA Leverage eXchange (WC
- Background
- Available Tests
- Providing Feed Tables

TADA Leverages the Water Quality eXchange (WQX) QAQC Service

This is a overview of the the WQX Quality Assurance and Quality Control (QAQC) data submission service, and how TADA leverages that service to flag potentially invalid data in the Water Quality Portal (WQP). It will cover: 1) an overview of all available WQX QAQC tests for data submissions, 2) which of these QAQC tests are also available in TADA for flagging potentially invalid WQP data, and 3) how to interpret and provide feedback on the validation reference tables referenced by WQX and TADA for this QAQC service.

Background

The WQX expectation for submissions is that users submit only QAQC'd data and utilize WQX elements to ensure the data is of "documented quality". The WQX team has historically hosted data quality working groups aimed at creating best practices and required data elements for WQX 3.0 for specific parameter groups such as nutrients, metals and biological data. These resources have supported users to submit data of documented quality. This approach has been



Water Quality Portal Demo

Retrieving data from the WQP

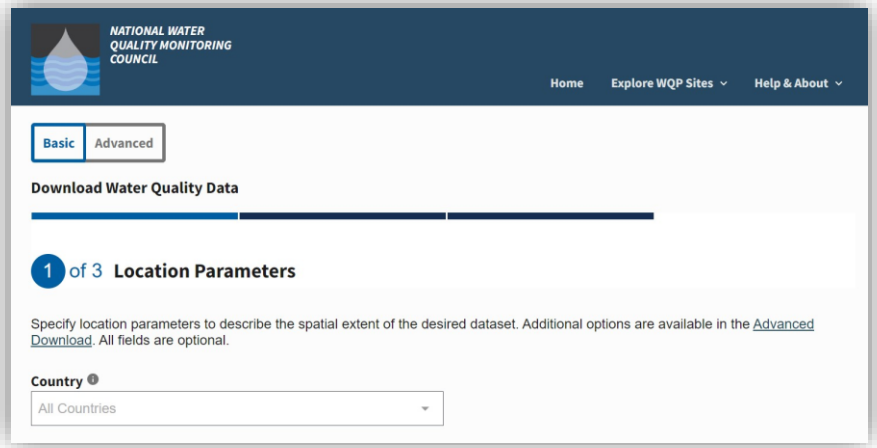
Several Options:

1. WQP Web Interface
   - WQP Demo on How to Download Data (2015)
   - WQP Demo on How to Download Data (2019)

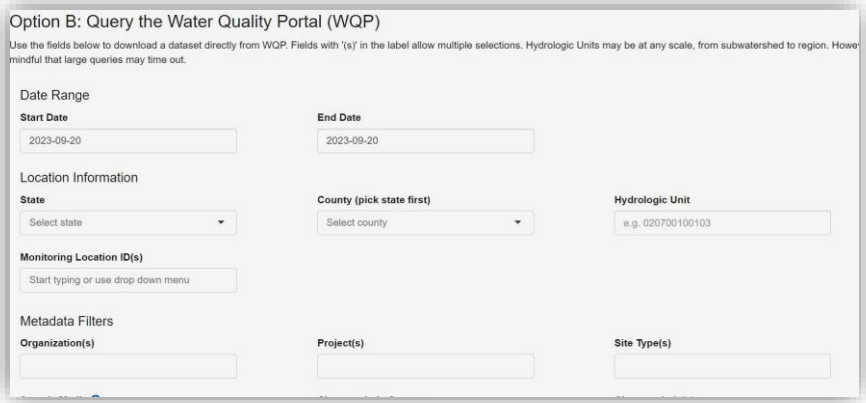
2. How's My Waterway

3. TADA

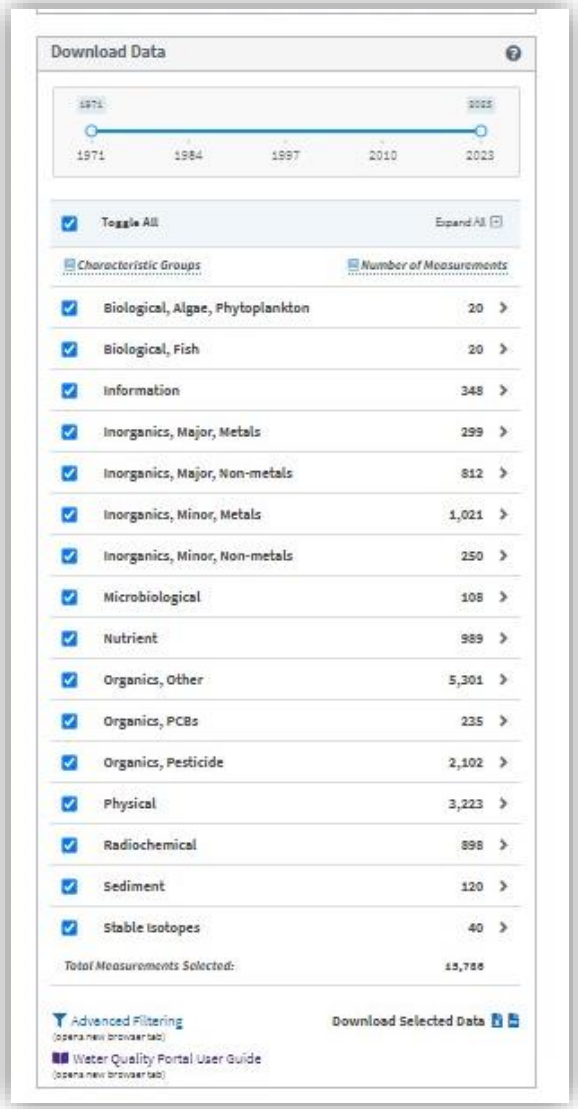
The Water Quality Portal Website



TADA



How's My Waterway



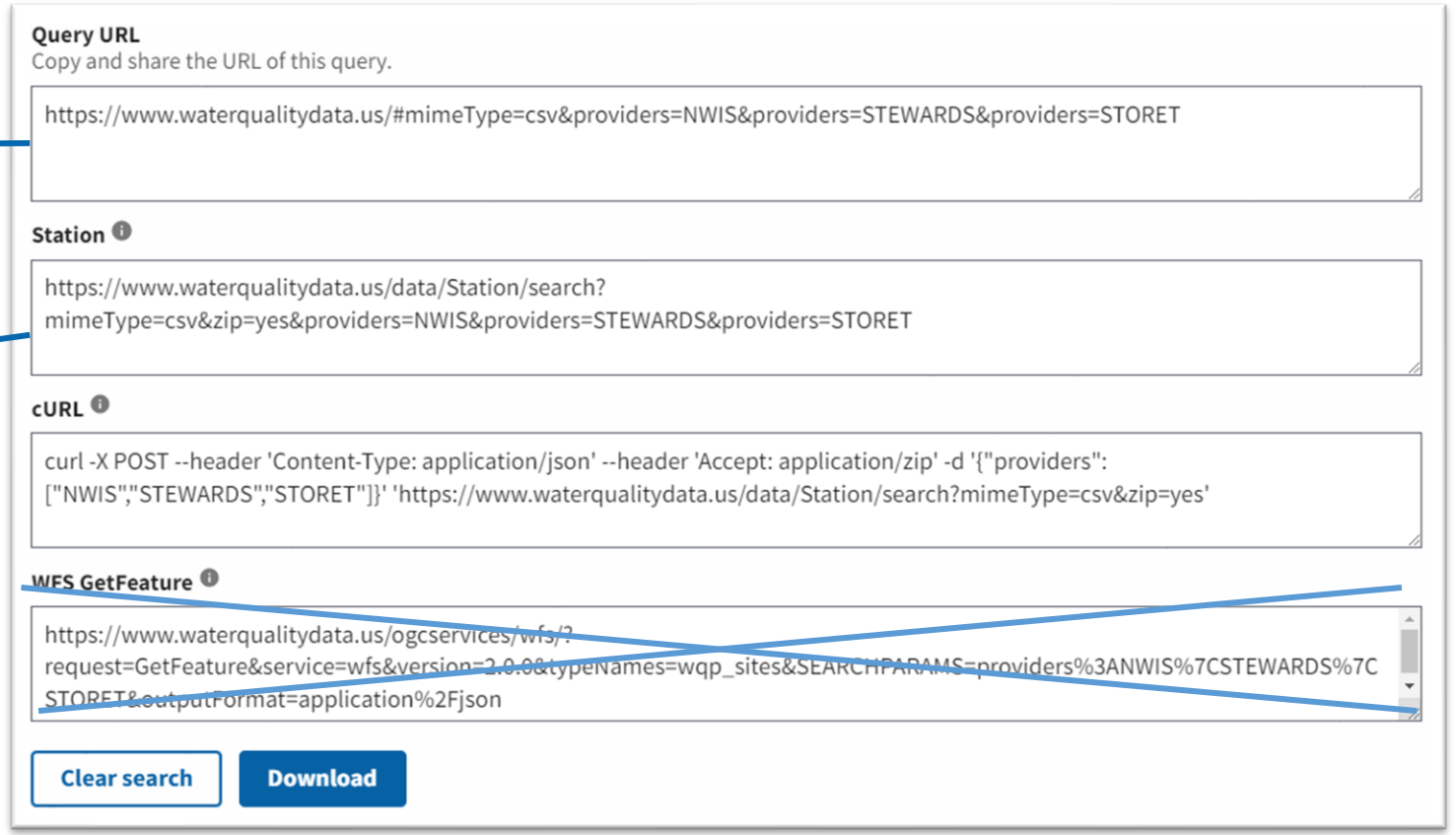
Using Data from WQP Web Services

- Web services are URLs that provide the instructions from your query

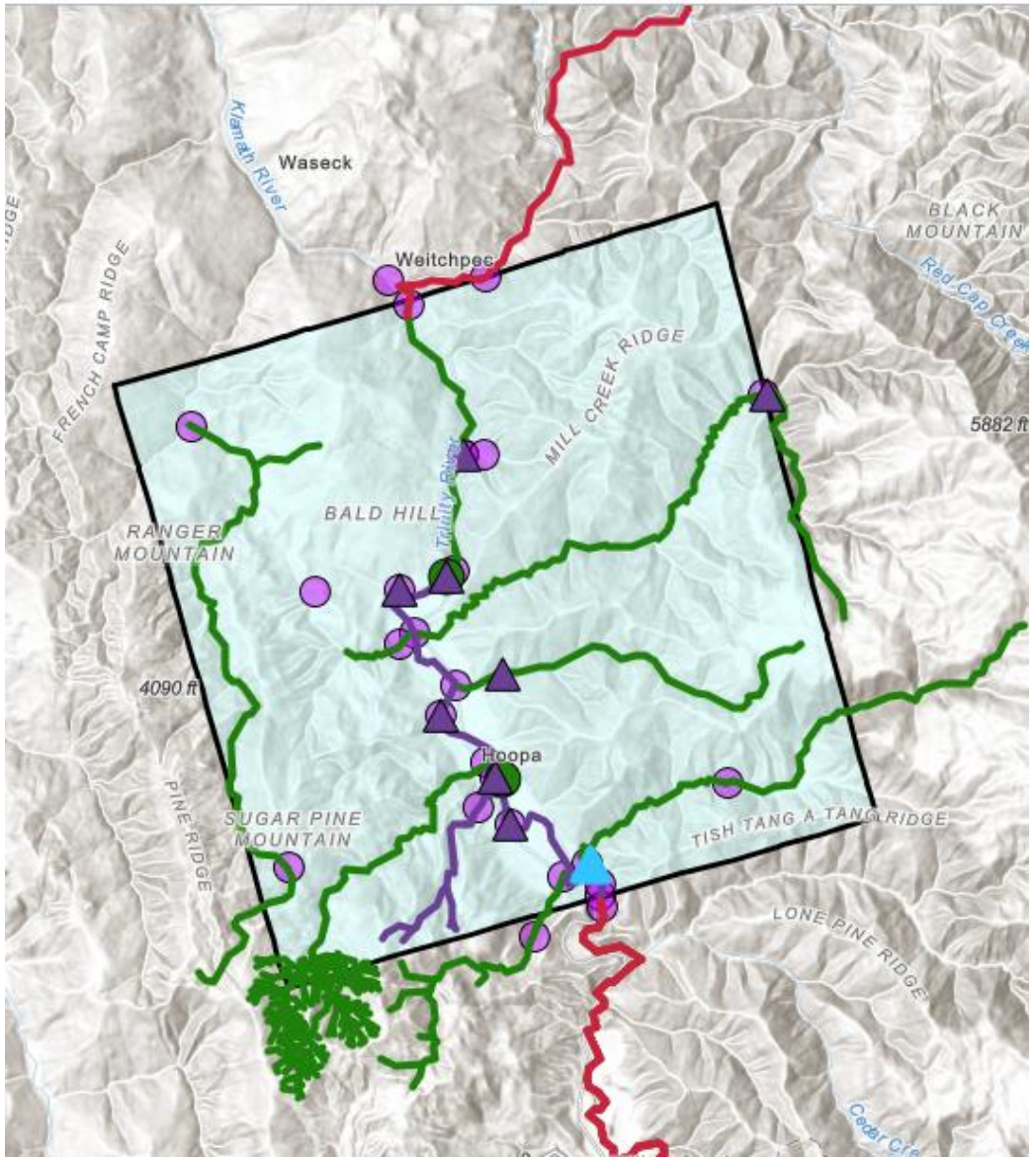
Save the URL to your Query

URL of your data download

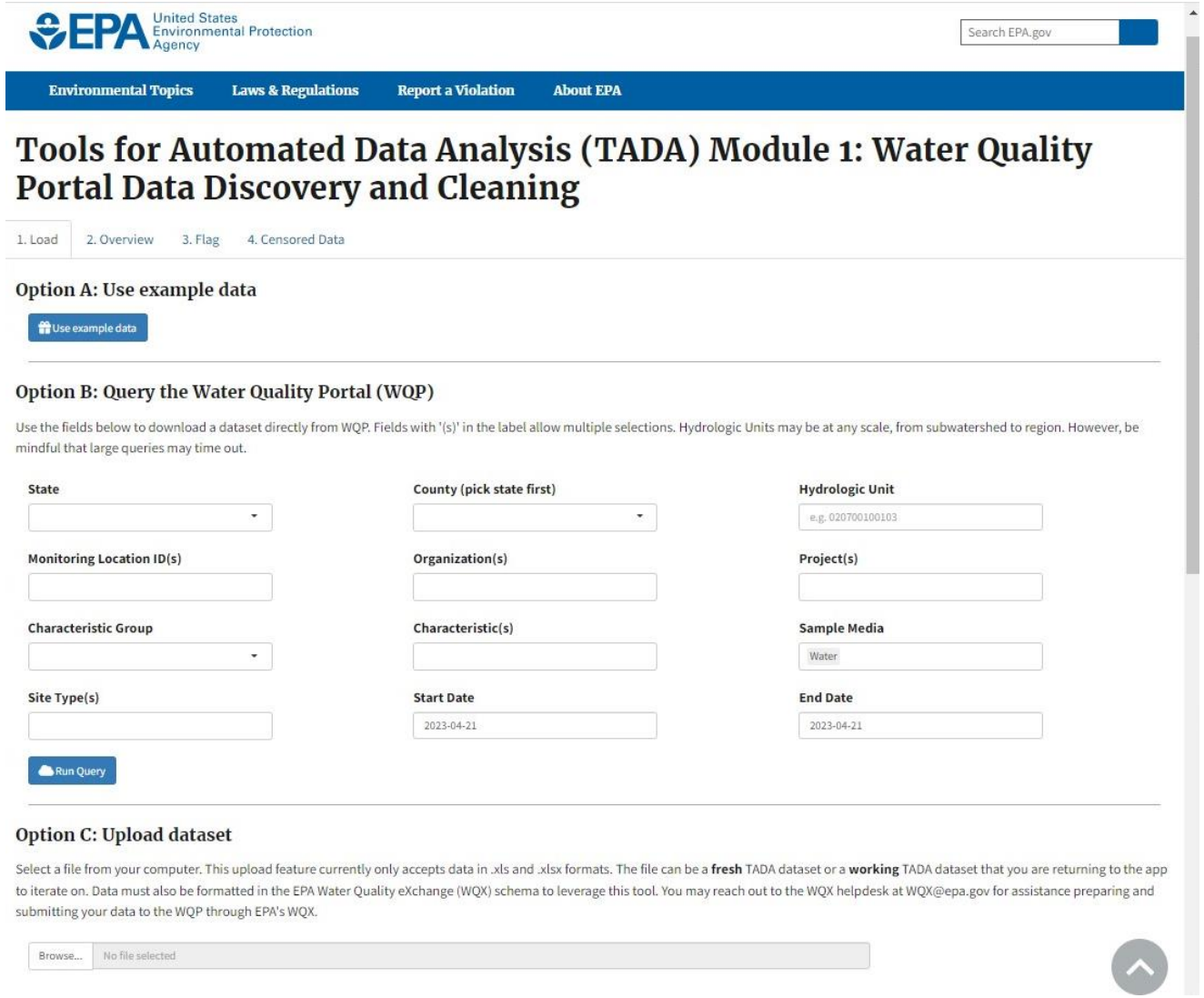
Use this web service URL in any data application that can read data, like MS Excel, R, Access, Arc Online, etc.



How's My Waterway Data Retrieval



TADA: WQP Data Retrieval in R



NATIONAL WATER QUALITY MONITORING COUNCIL

EPA United States Environmental Protection Agency

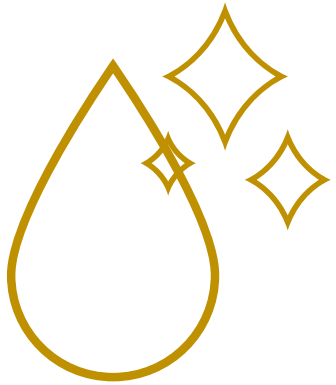
USGS science for a changing world

WQX WATER QUALITY EXCHANGE

dataRetrieval USGS R

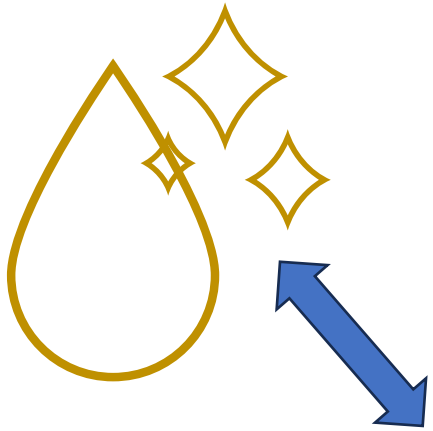
TADA Vision

Discover, wrangle, and QC data

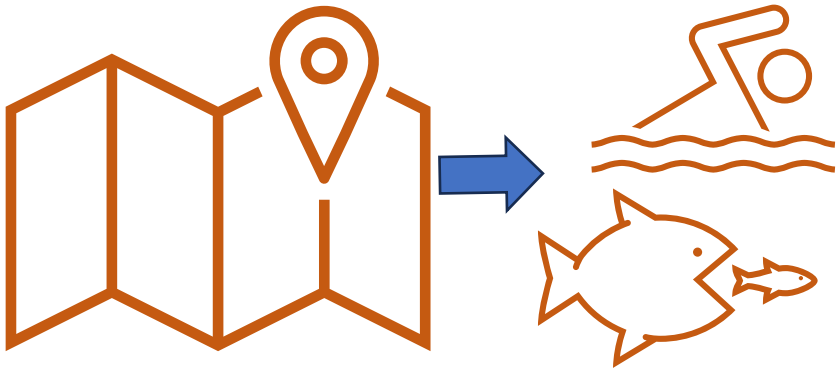


TADA Vision

Discover, wrangle, and QC data



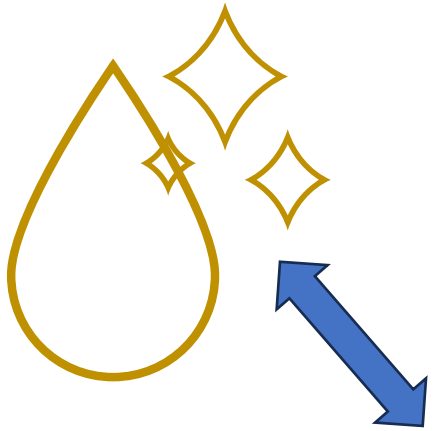
Assign beneficial uses



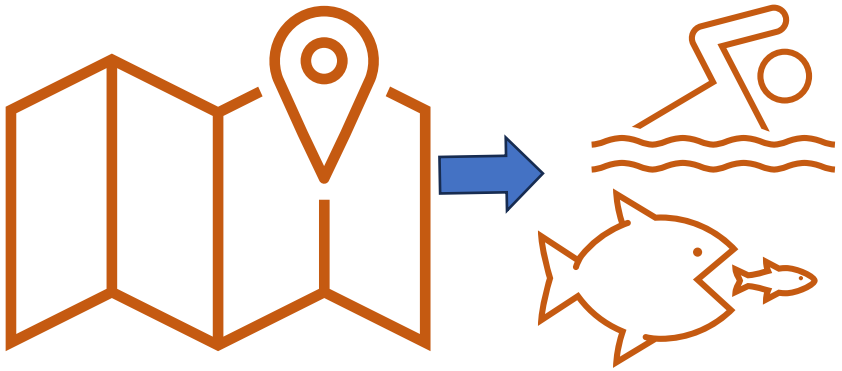
Assessment unit overlay with monitoring locations

TADA Vision

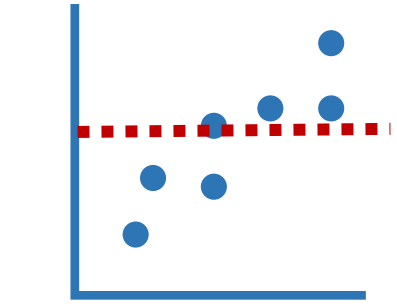
Discover, wrangle, and QC data



Assign beneficial uses



Assessment unit overlay with monitoring locations



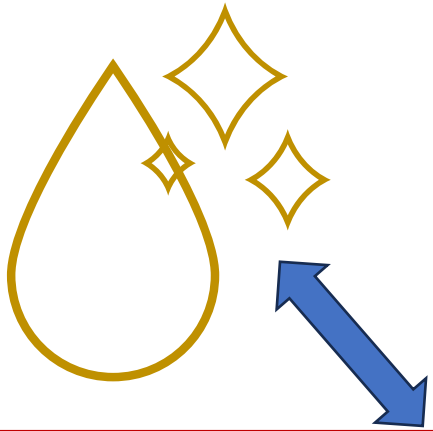
Beneficial uses determine numeric criteria used

Assessment methods guide impairment decisions based on:

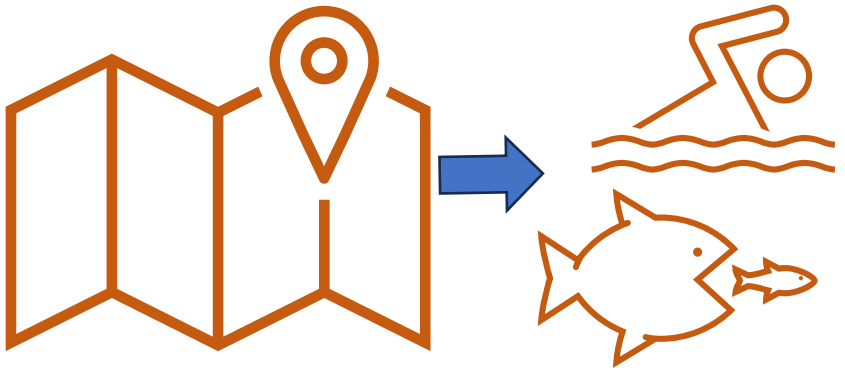
- Period of record
- Aggregated measurements
- Frequency
- Duration
- Magnitude
- Season
- Correction factors
- Covariates
- Site-specific criteria

TADA Vision

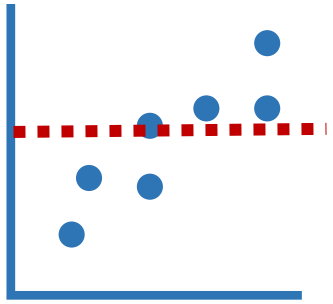
Discover, wrangle, and QC data



Assign beneficial uses



Assessment unit overlay with monitoring locations



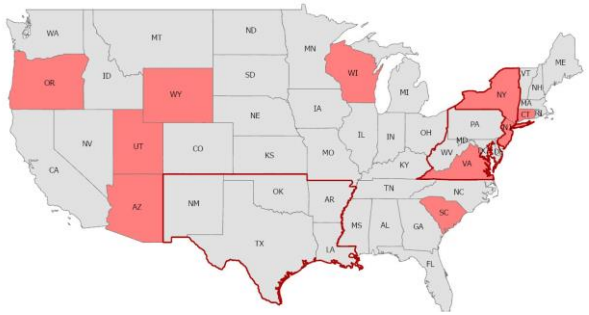
Beneficial uses determine numeric criteria used

Assessment methods guide impairment decisions based on:

- Period of record
- Aggregated measurements
- Frequency
- Duration
- Magnitude
- Season
- Correction factors
- Covariates
- Site-specific criteria

WQP TADA Working Group

- Established in 2020
  - Built a community
  - Sharing examples from around the country
  - Extensive assessment process and methodologies discussions
  - Join us!



- Growing over time
- Currently meeting once every other month

- EPA
  - HQ: WDIB, MAB, WB, OST, ORD, OECA, OLEM, ARD
  - Regions: 1, 2, 3, 6, 7, 8
- At least 1 State from each EPA Region
  - VA, WI, MA, SC, MT, UT, AZ, IL, WA, OR, KS, CT, CA, MN, OK, MO, NJ, IN, NY, AR, AK, LA, MD, GA, DC
- Tribal Nations from Regions 1, 6 and 8
  - Penobscot Nation, Ute Mountain Ute, Absentee Shawnee Tribe, Pueblo of Tesuque, Owens Valley Indian Water Commission, others early on (staff turnover)
- Federal Agencies/Universities/Other
  - USGS, TetraTech, Colorado State, Long Island Sounds Study, Oak Ridge National Laboratory

Inventory of Open-Source R Tools for Water Analyses

- Over 50 resources to learn from, and build on
- Working Group helps share knowledge, examples, and set priorities
- Faster progress through collaboration and iteration (learning from each other)

TADA Working Group: Identified Requirements/Priorities

Scope

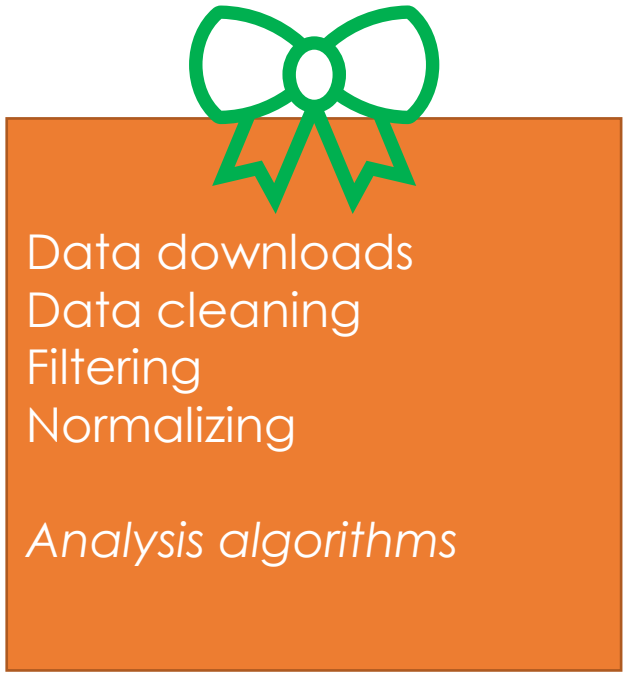
- Focusing on quantitative (numeric) water data in the WQP to start
- Focusing on frequently assessed parameters
- Common assessment processes and methodologies

Arsenic	Nickel
Boron	Nitrate
Chlorophyll a	Total Nitrogen, mixed forms
Chromium	pH
Chromium(VI)	Total Phosphorus, mixed forms
Cadmium	Depth, Secchi disk depth
Copper	Selenium
Dissolved oxygen (DO)	Silver
Dissolved oxygen saturation	Temperature, water
Escherichia coli	Total suspended solids
Lead	Chromium(III)
Mercury	Zinc

Common Methodologies

- Spatial aggregation – assessment unit and station level assessments
- Characteristic specific assessment start and end dates
- Magnitude, duration (temporal aggregation)
- Criteria context – upper or lower limit, range
- n-day mean, n-day mean maximum or mean minimum, n-hour mean, geometric mean, arithmetic mean, n-day rolling average
- Frequency criteria (e.g., 10% rule, 1-in-3 years rule applied using binomial test or percentile)
- Custom input equations needed to calculate criteria (e.g., for ammonia and certain metals)
- Incorporating depth
- Acute vs chronic
- Seasonality

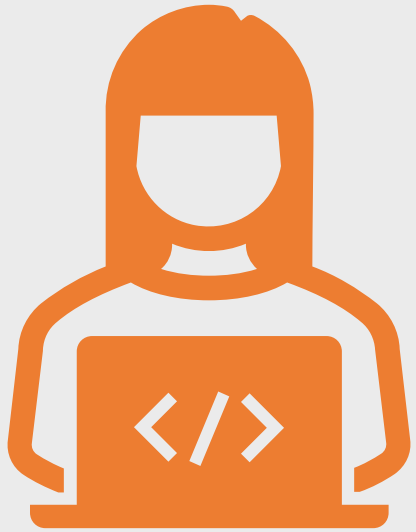
R Package



- Focus on algorithms specific to WQP data
- Series of functions to assist common analysis processes
- Data is flagged but not automatically removed or modified
- Can be easily incorporated into existing tools or data processing methods
- Highly customizable
- Companion to **dataRetrieval** package

R Shiny User Interface

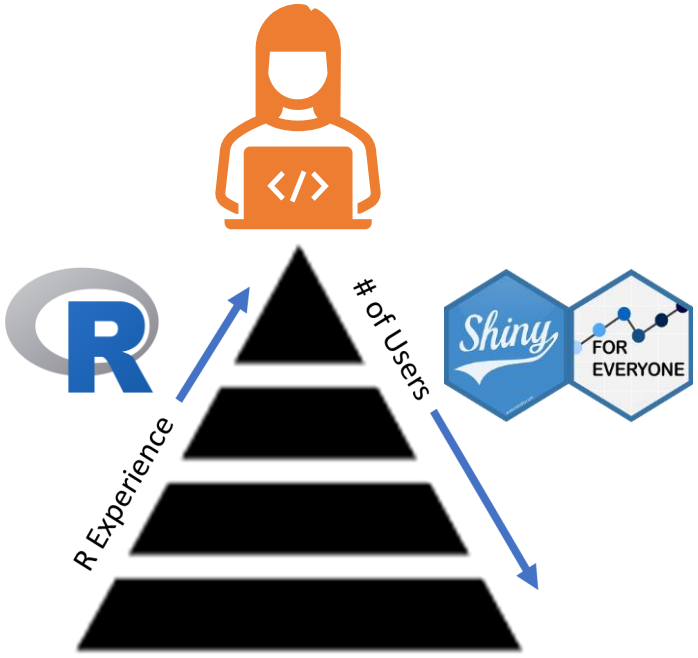
- R Shiny
  - Makes use of R package
- Developed independently
- Guides user through process
- Interactive
  - data exploring
  - cleaning
  - graphing
  - etc.
- Web based



TADA Products

- Different tools for different users
  - R Package (coders)
  - R Shiny Application (non-coders)
- User Guide on GitHub Pages
- EPA TADA Website
- R and R Shiny Learning Resources for Water Community, Collaborative Effort Between TADA Working Group & North American Lake Management Society (NALMS)
- Working Group SharePoint & inventory of open-source R code and WQP tools – please reach out to learn more!

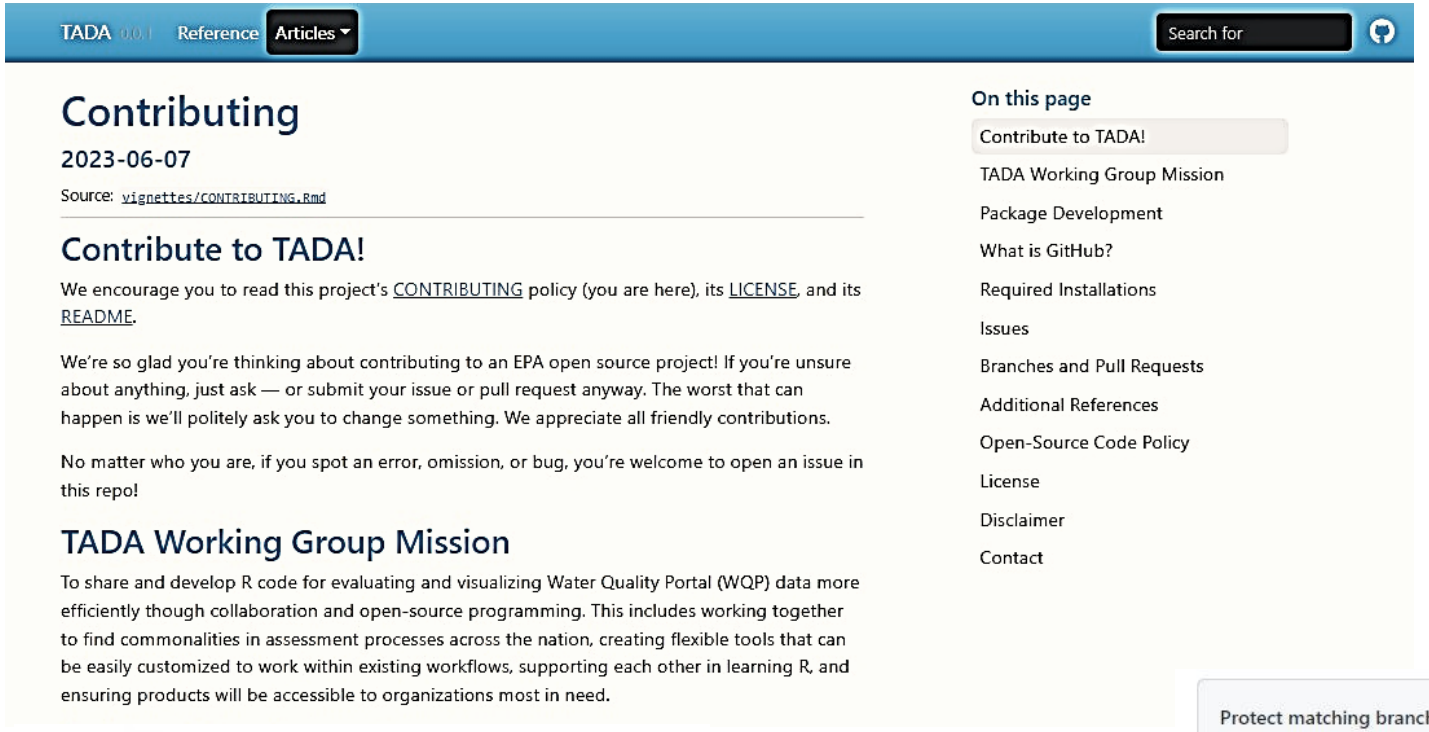
"Serve as a hub for an open-source water quality community"

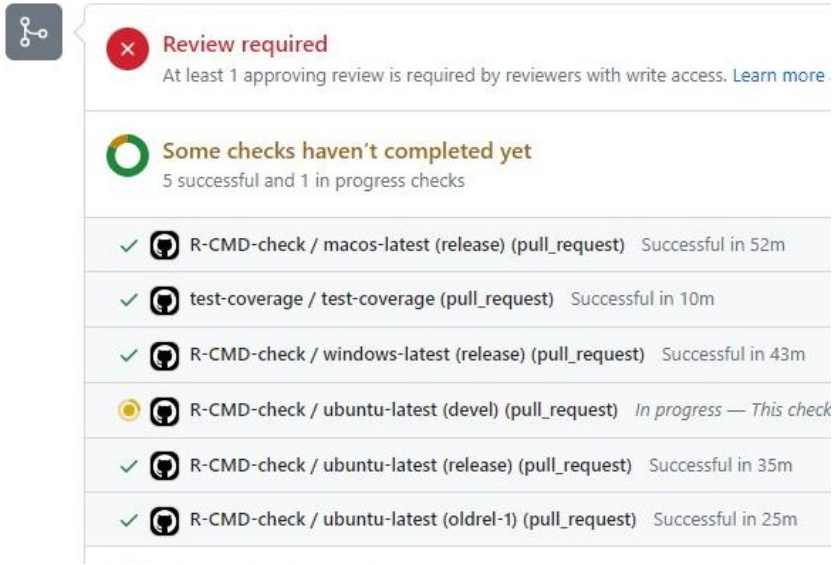


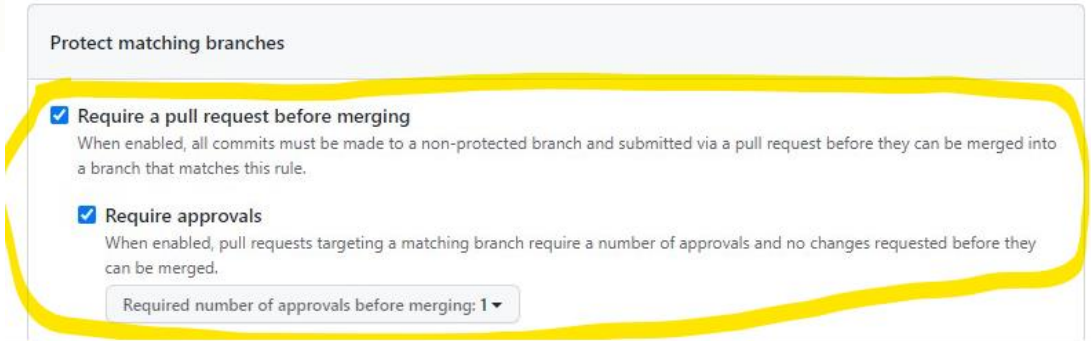
Working Group Mission

*To share and develop **R code** for evaluating and visualizing **WQP** data more efficiently though collaboration and open-source programming. This includes working together to find commonalities in assessment processes across the nation, creating flexible tools that can be easily customized to work within existing workflows, supporting each other in learning R, and ensuring products will be accessible to organizations most in need.*

Involving users in the development process

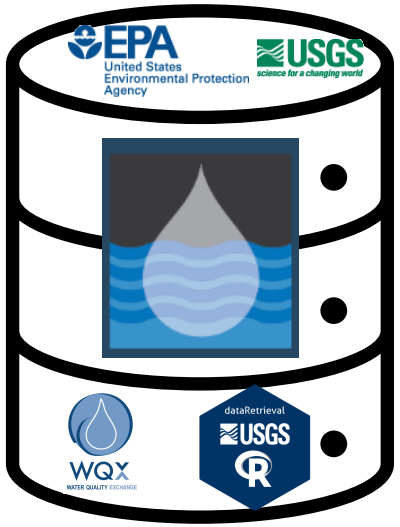




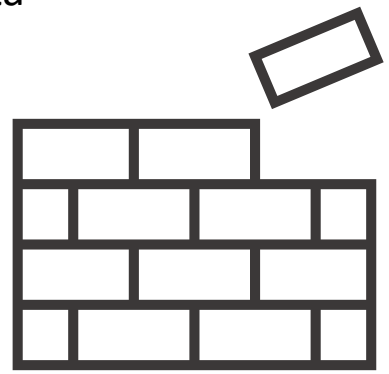


https://usepa.github.io/TADA/articles/CONTRIBUTING.html

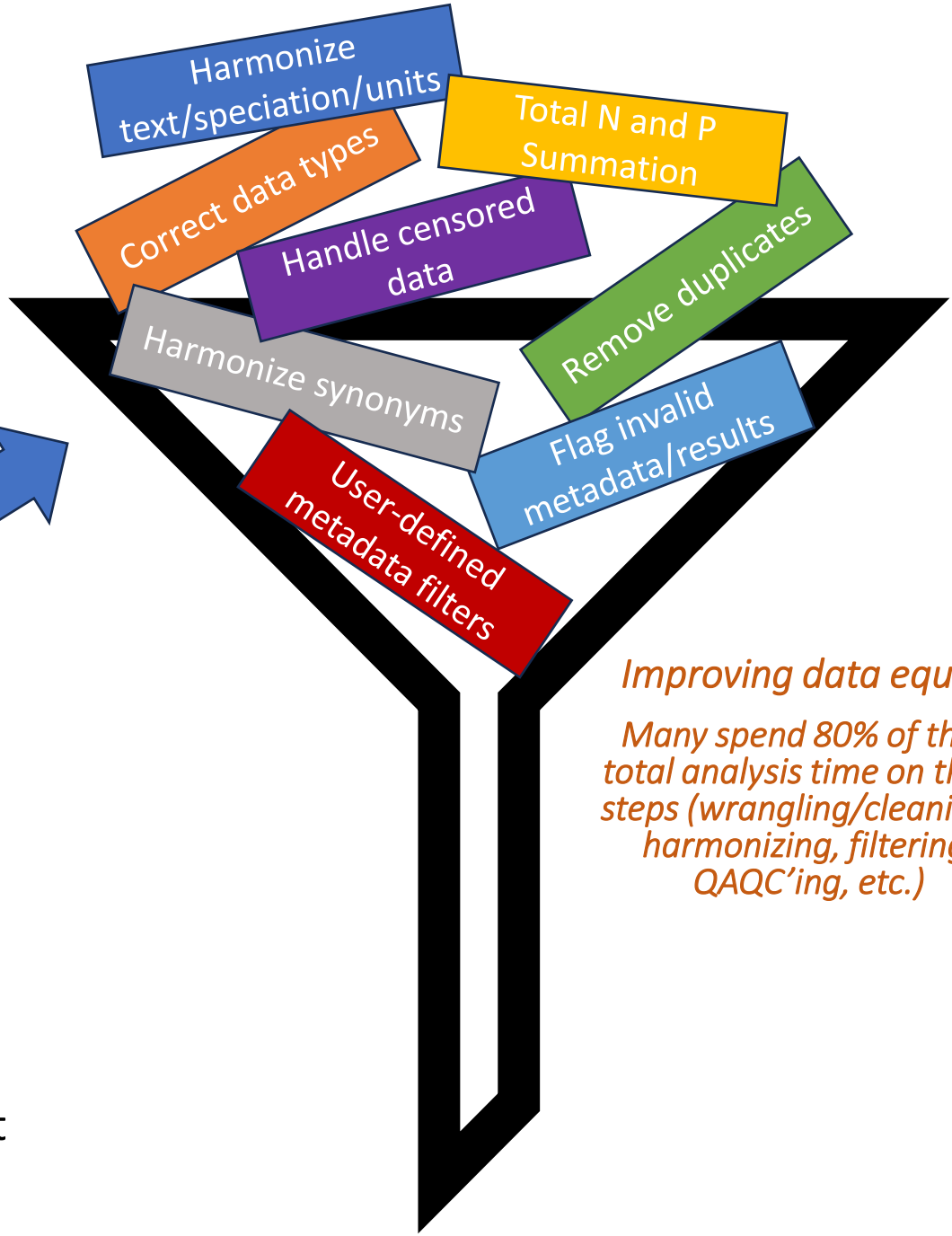
Module 1



Retrieve data from WQP



Construct a unified dataset containing key metadata



Improving data equity

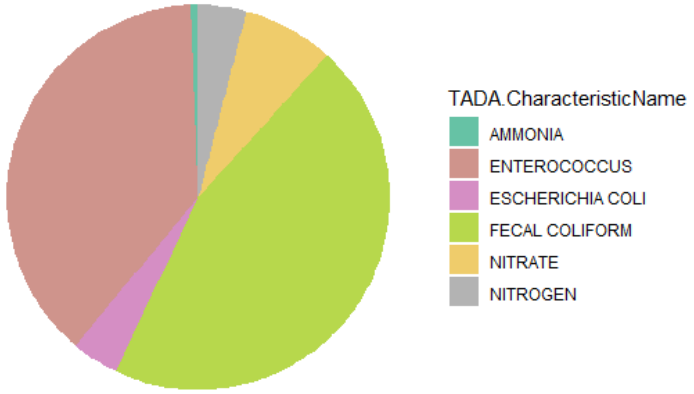
Many spend 80% of their total analysis time on these steps (wrangling/cleaning – harmonizing, filtering, QAQC'ing, etc.)

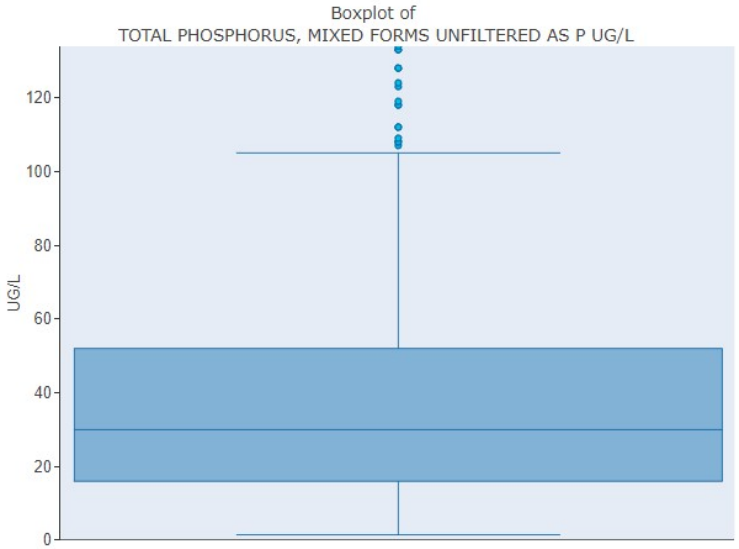
- Is the data of sufficient quality for my analysis?
- Is it relevant?
- Does it include key metadata?
- Can the data be harmonized and grouped in a way that makes sense for my analysis?

Modules 2-4

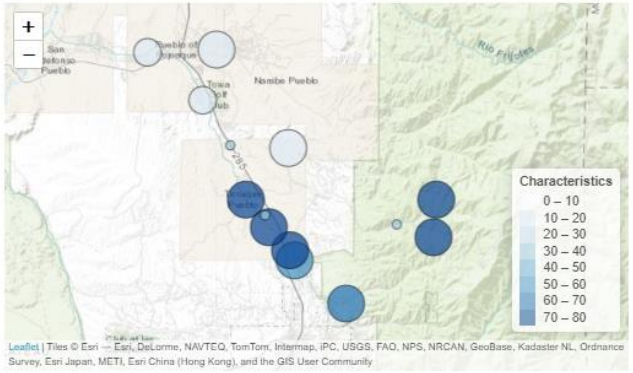
- Gathered requirements, at refining stage
- Started proof of concept, starting development soon
- Challenges
  - Formatting water quality standard information
    - Criteria Search Tool does not include narrative standards, duration and frequency, or methodologies
  - System (ATTAINS/WQP/CST) crosswalk development and maintenance (parameters, designated uses)

TADA Visualizations





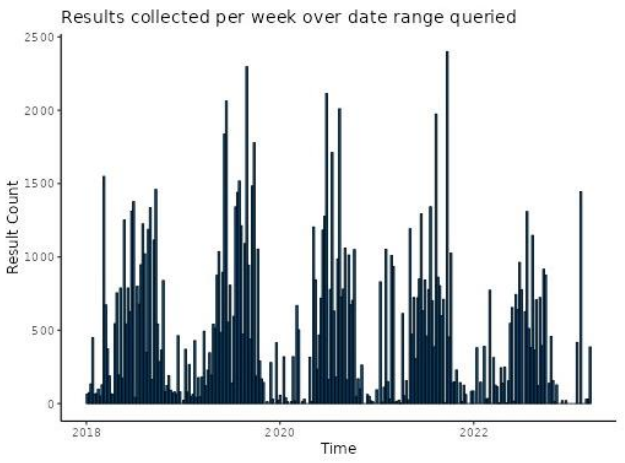
Your dataset contains **131,106** unique results from **221** monitoring location(s) and **6** unique organization(s).

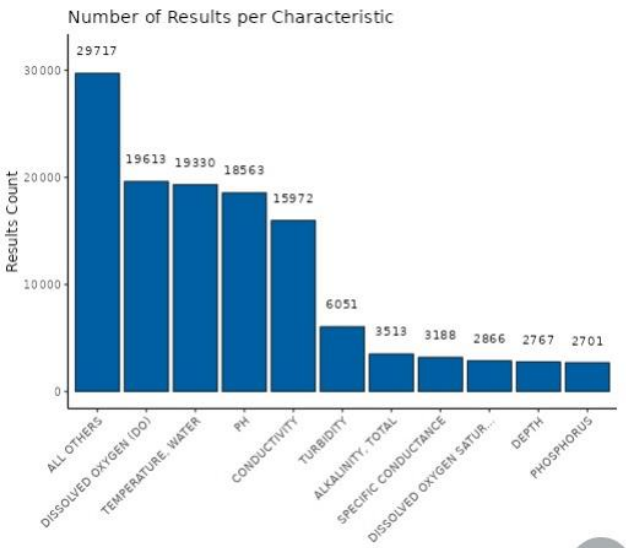


Show 10 entries

OrganizationFormalName	Result_Count
Chickasaw Nation Environmental Service	4580
Fond du Lac Band of Chippewa (MN)	20176
Pueblo Of Tesuque	6795
Pueblo of Pojoaque	1181
Red Lake DNR	81734
Sac and Fox Nation (Tribal)	9815

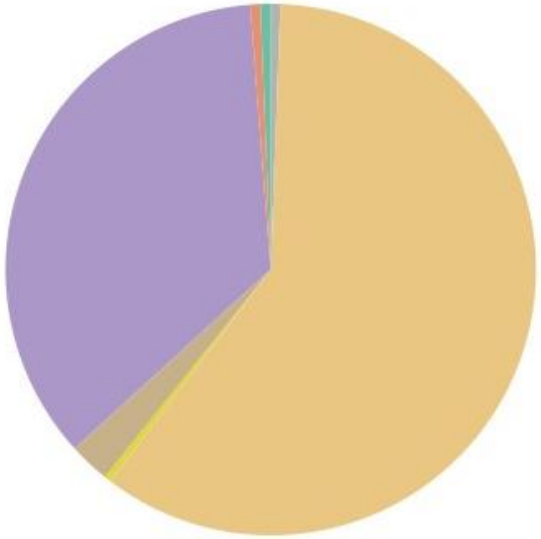
Showing 1 to 6 of 6 entries Previous 1 Next

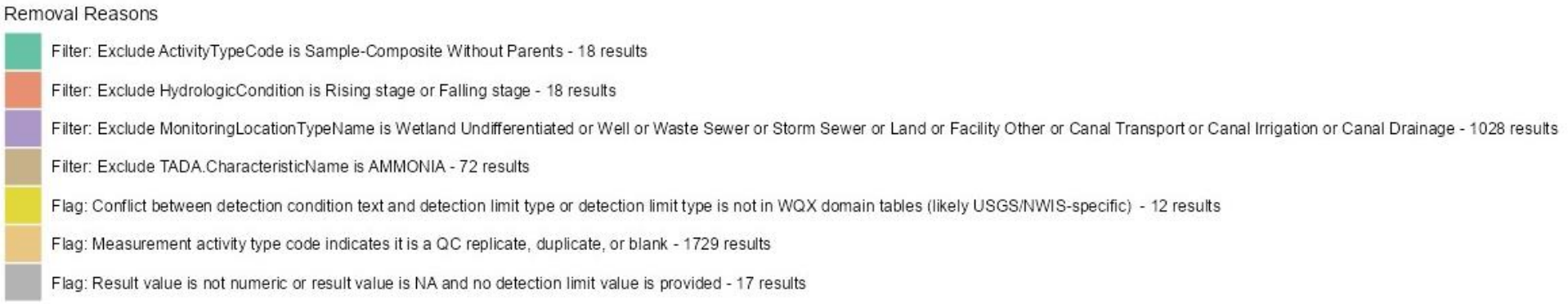




Summary of User Decisions

TADA.Remove	TADA.RemovalReason
TRUE	Flag: Measurement activity type code indicates it is a QC replicate, duplicate, or blank
TRUE	Filter: Exclude ActivityTypeCode is Sample-Composite Without Parents
TRUE	Filter: Exclude HydrologicCondition is Rising stage or Falling stage
FALSE	
TRUE	Filter: Exclude MonitoringLocationTypeName is Wetland Undifferentiated or Well or Waste Sewer or Storm Sewer or Land or Facility Other or Canal Transport or Canal Irrigation or Canal Drainage
TRUE	Flag: Result value is not numeric or result value is NA and no detection limit value is provided, Flag: Measurement activity type code indicates it is a QC replicate, duplicate, or blank
FALSE	
TRUE	Flag: Result value is not numeric or result value is NA and no detection limit value is provided
TRUE	Flag: Measurement activity type code indicates it is a QC replicate, duplicate, or blank, Filter: Exclude MonitoringLocationTypeName is Wetland Undifferentiated or Well or Waste Sewer or Storm Sewer or Land or Facility Other or Canal Transport or Canal Irrigation or Canal Drainage





TADA Limitations and Challenges

- Does not cover narrative/text data
- Water data only for now (could expand in future)
- Does not include WQP biological data profile (taxon)
- Will require some user upkeep / maintenance
- User decides which components to use (or not)
- Discrete data only for now (could expand in future)
- Human review required

Broader Impacts

Use of TADA has potential to greatly reduce total government costs across:

- State and tribal agencies
- EPA regions
- EPA HQ
- USGS, other federal agencies

Efficient, transparent, and reproducible assessments

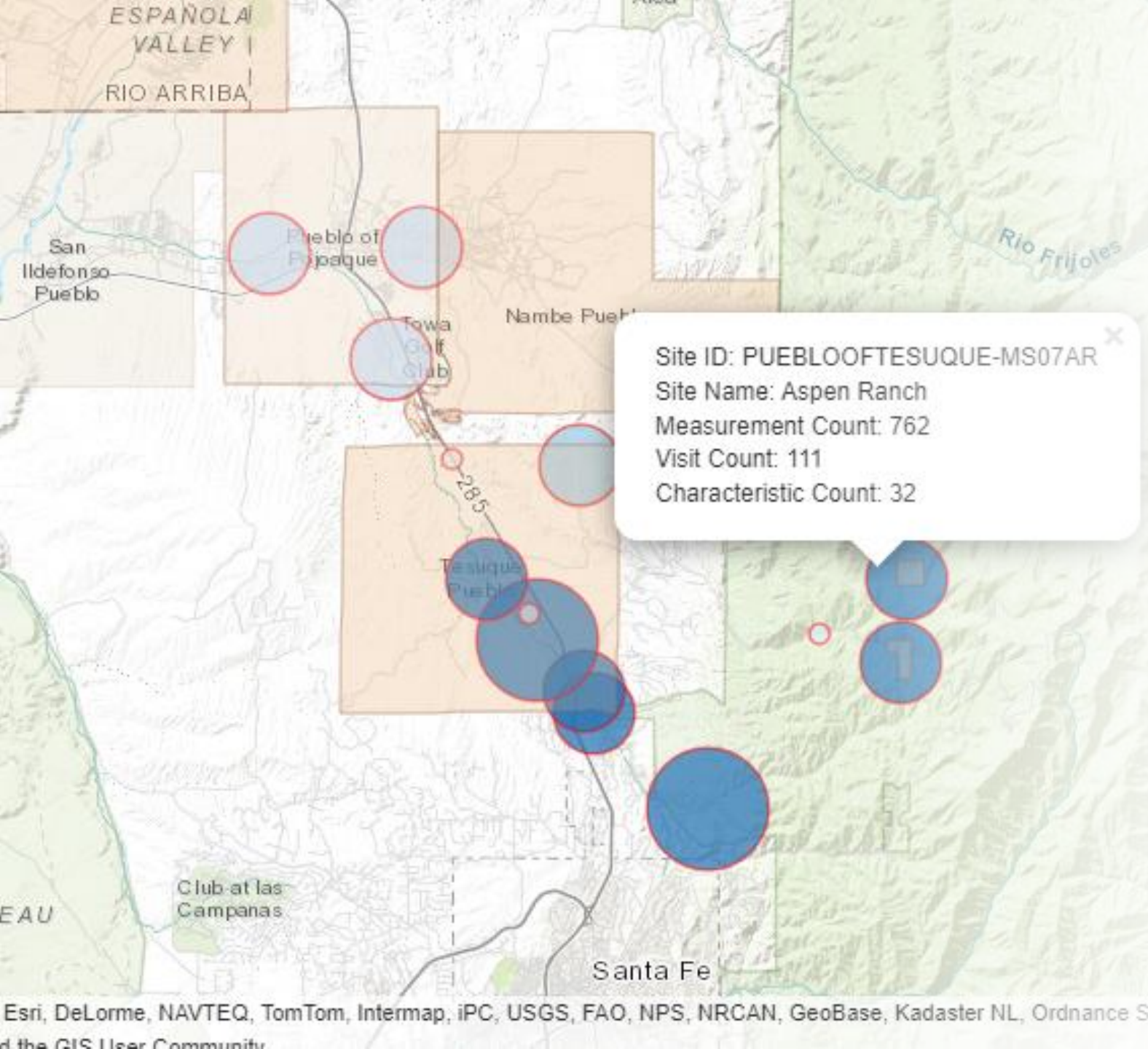
- Frees up time for other important tasks
- May facilitate assessing more waters
- Assists tribal onboarding to ATTAINS
- Helps discover and share commonalities in assessment processes nationally
- Improves interoperability across WQX/WQP, ATTAINS, and the Criteria Search Tool (CST)

Building data equity

- Facilitates use of other organizations data in State or Tribal assessments
- Makes the WQX QAQC service available on the WQP side
- Helps find and address data quality issues in WQX/WQP

TADA

WQ Assessment



TADA R Shiny App Demo

- TADA Shiny App: https://github.com/USEPA/TADAShiny
- Dev web application: https://owshiny-dev.app.cloud.gov/tada-dev/

Thank you for listening!

And a BIG thank you to all our TADA contributors!

TADA Working Group (all participants)
Jacob Greif, former EPA WDIB ORISE (now at EPA OW)
Elise Hinman, former EPA WDIB ORISE (now at USGS)
Hillary Marler, EPA OW
Michelle Thawley, EPA OW
Kathleen Healy, EPA ARD (WDIB Skills Marketplace)
Trip Hook, EPA OLEM (WDIB Skills Marketplace)
Justin Bousquin, EPA ORD
Elinor Keith, EPA OW
Jesse Boorman-Padgett, EPA OW
Laura Shumway, EPA OW
Laura DiCicco, USGS
Hui Zhou, ERG (EPA contractor)
Renee Myers (EPA contractor)
Brad Cooper, ERG (EPA contractor)
Florian Rupprecht (Hacktoberfest 2022 contributor)
Zachary Smith, NY DEP