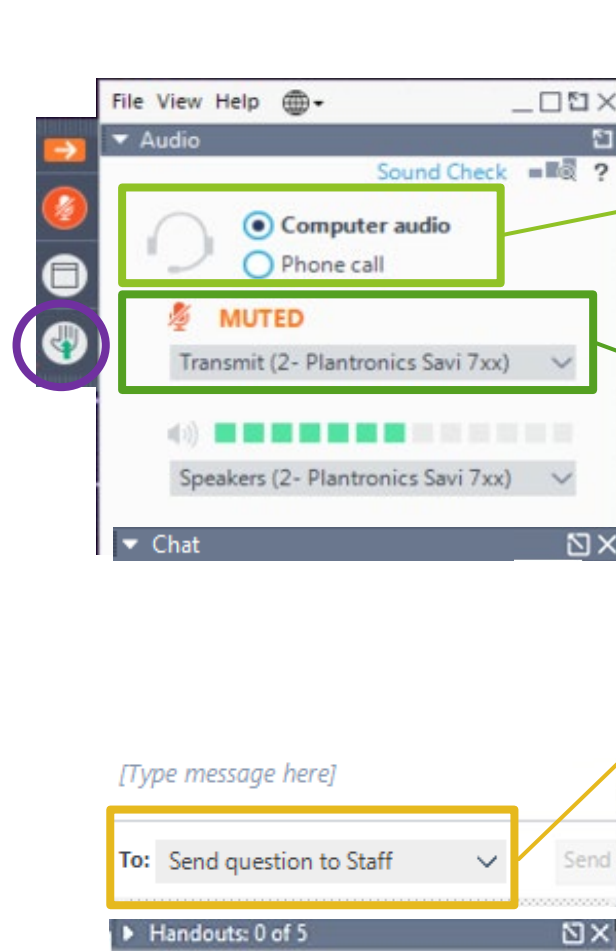# Using GoToWebinar

This is your dashboard.
Use the arrows in the grey bars to see your options.

Two Options for Audio:
1. Use your **computer's** speakers
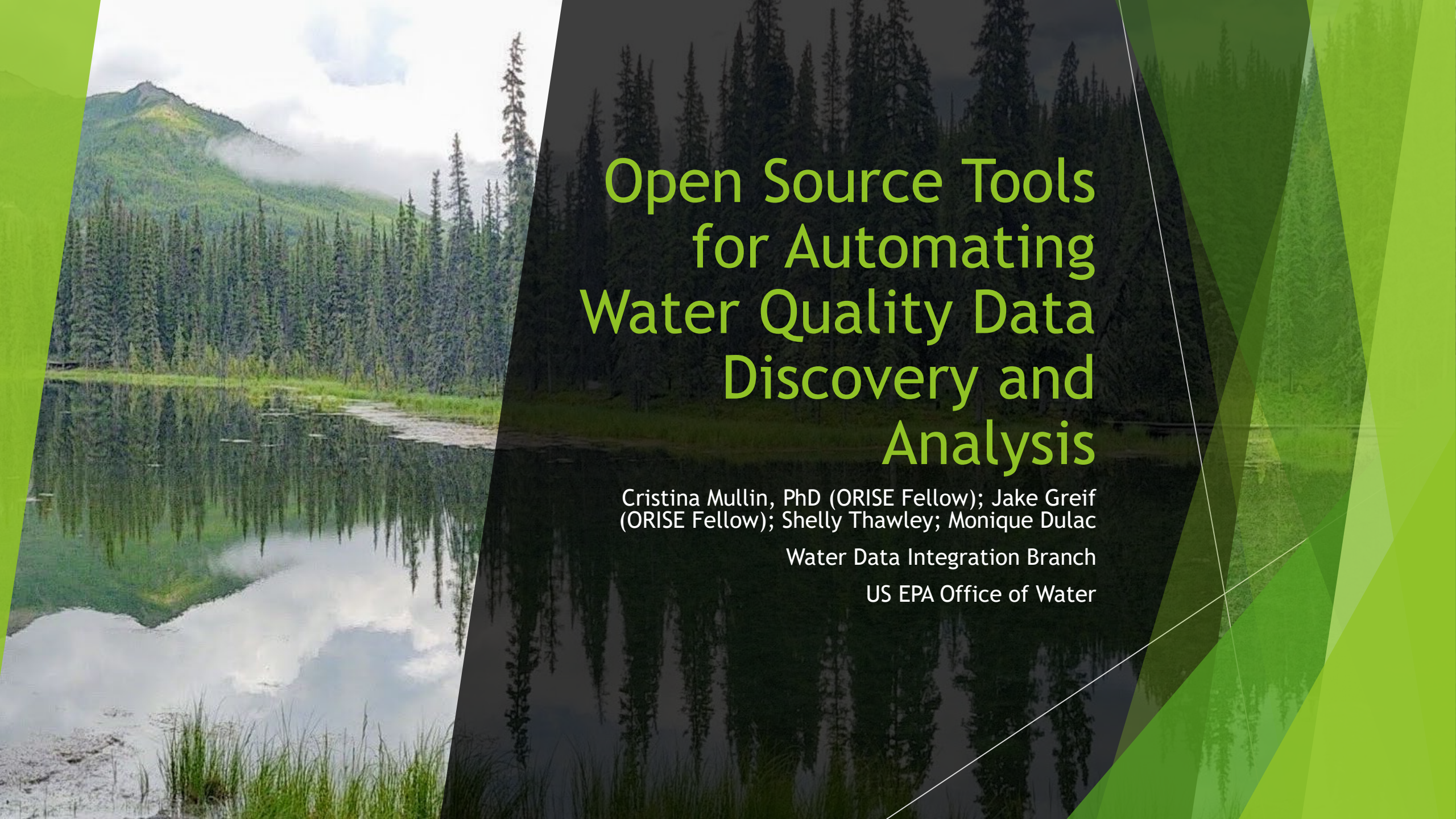2. Call in by **phone** using the call-in number and access code in the Audio section of the control box.

All participants are **muted** by default.

Questions: Use the "send question" option in the chat panel to ask questions throughout the presentation.

Ask for help! Use the questions box to let us know if you have technical difficulties. You can also **raise your hand** for help.

If you haven't downloaded the handout, it's available here.

Open Source Tools for Automating Water Quality Data Discovery and Analysis

Cristina Mullin, PhD (ORISE Fellow); Jake Greif (ORISE Fellow); Shelly Thawley; Monique Dulac

Water Data Integration Branch

US EPA Office of Water

Background

Open-source solutions are increasing in use

Numerous examples of automated work flows now exist

Review of existing approaches

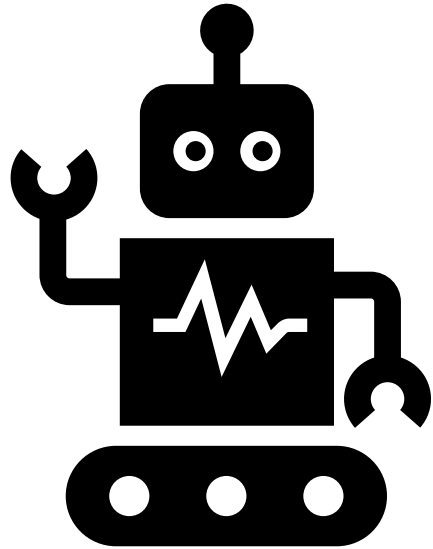Look for ways to better assist with data access and manipulation

What can we do to add value?

Hub for stakeholder community

Develop assessment tools that serve a broad range of stakeholder needs
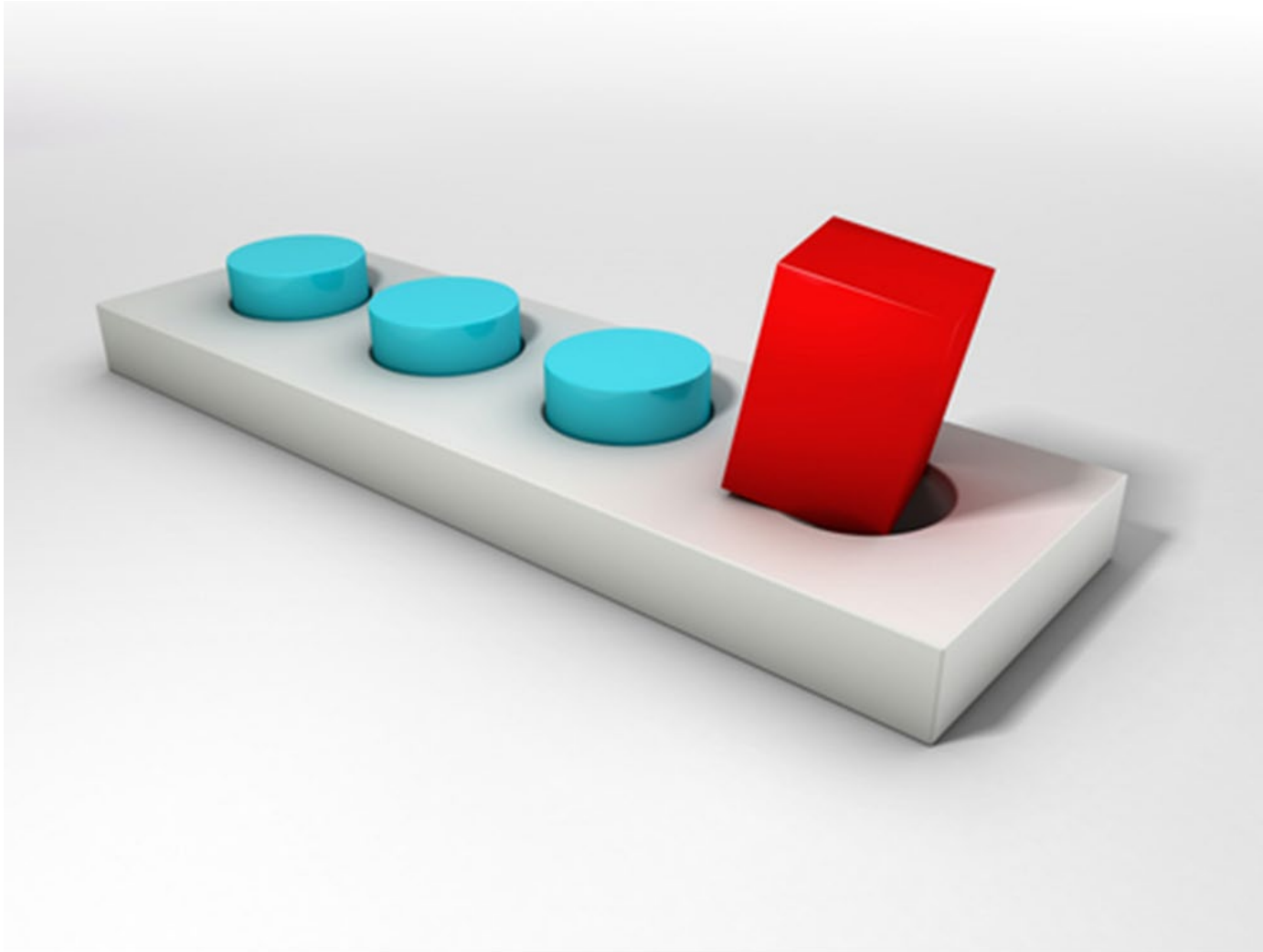
Automation of water quality analysis processes

Improves work efficiency
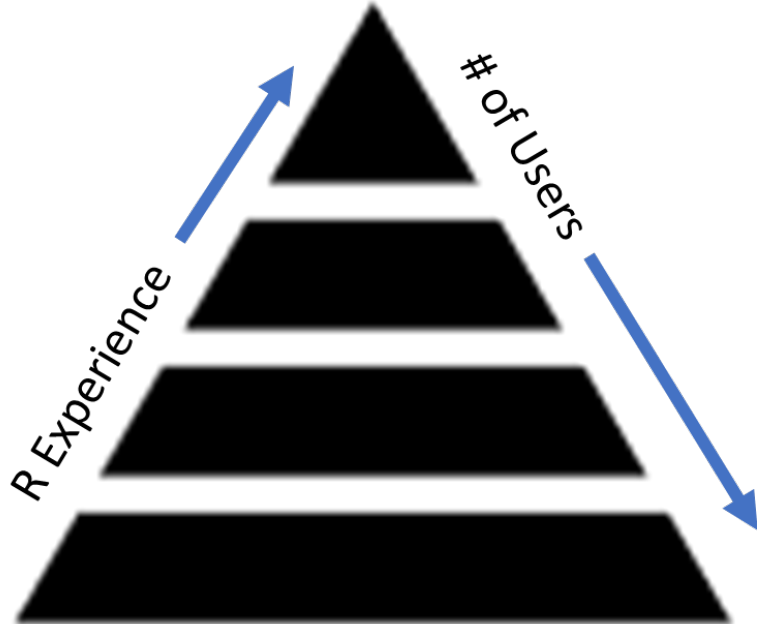
Improves accuracy

Reproduceability

Water Quality Assessment Process

One size does not fit all

Focus on common methods

Provide user flexibility

TADA Vision
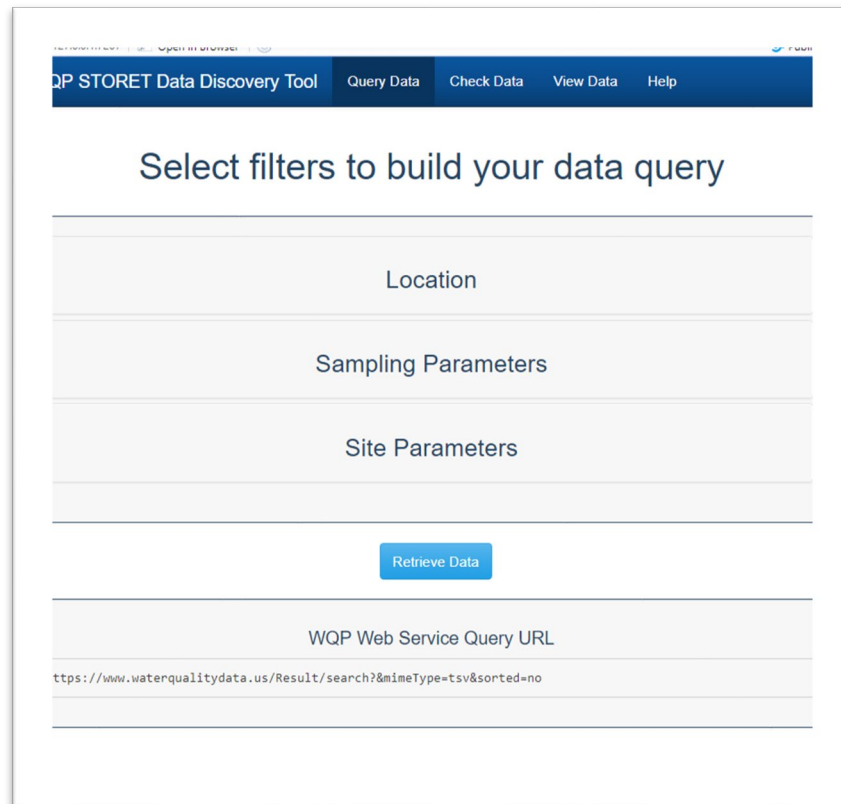
R Package
◦ Series of functions to assist common assessment methods
◦ Maximize user configuration

User Interface
◦ rShiny (with EPA Geoplatform for mapping)
◦ Assist with input generation
◦ Assist with workflow

Data Discovery and Analysis Tools

https://www.epa.gov/waterdata/water-quality-portal-data-discovery-tool

https://github.com/USEPA/Water-Quality-Data-Analysis-Tool.git

Up Next…

Common Methods (R Package) – Cristina Mullin

TADA GeoViewer (GUI) – Monique Dulac

R Development Plan – Jake Greif

Federal Clean Water Act Assessment Process

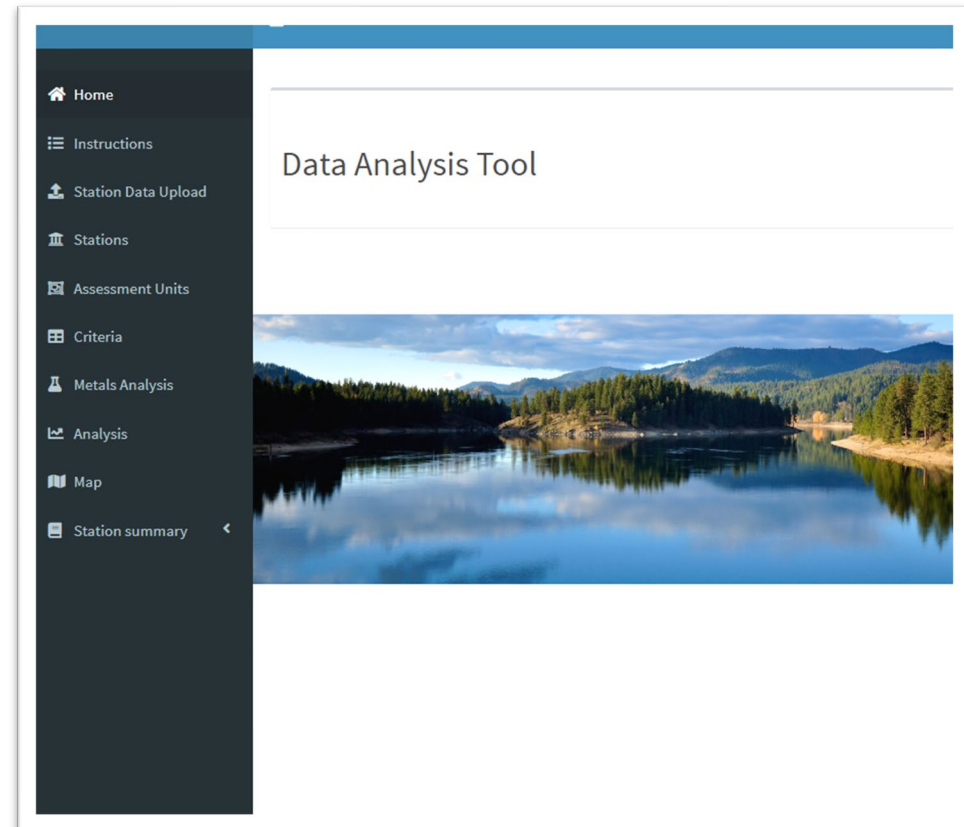- Section 305(b): Entities monitor, assess and report on the quality of their waters relative to designated uses established in accordance with Water Quality Standards
  - EPA WQP/WQX (Water Quality Portal/Water Quality eXchange)
  - EPA ATTAINS (The Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System)
  - EPA Criteria Search Tool

- Section 303(d): EPA is authorized to assist Entities with 303(d) programs in listing impaired waters
  - EPA ATTAINS

Goal: To help entities perform assessments more efficiently!

Data System Compatibility: Crosswalks

WQP, ATTAINS, and Criteria Search Tools

Benefits:
- Enhances automation capabilities of TADA
- Encourages greater use of EPA resources as a whole
- Reduces redundancies

Challenges:
- One to many relationships
- These tools are dynamic and require regular updates
- Use designations are inconsistent across the Criteria Search Tool and ATTAINS
- Limited number of entities can take advantage of these crosswalks

Requirements Gathering
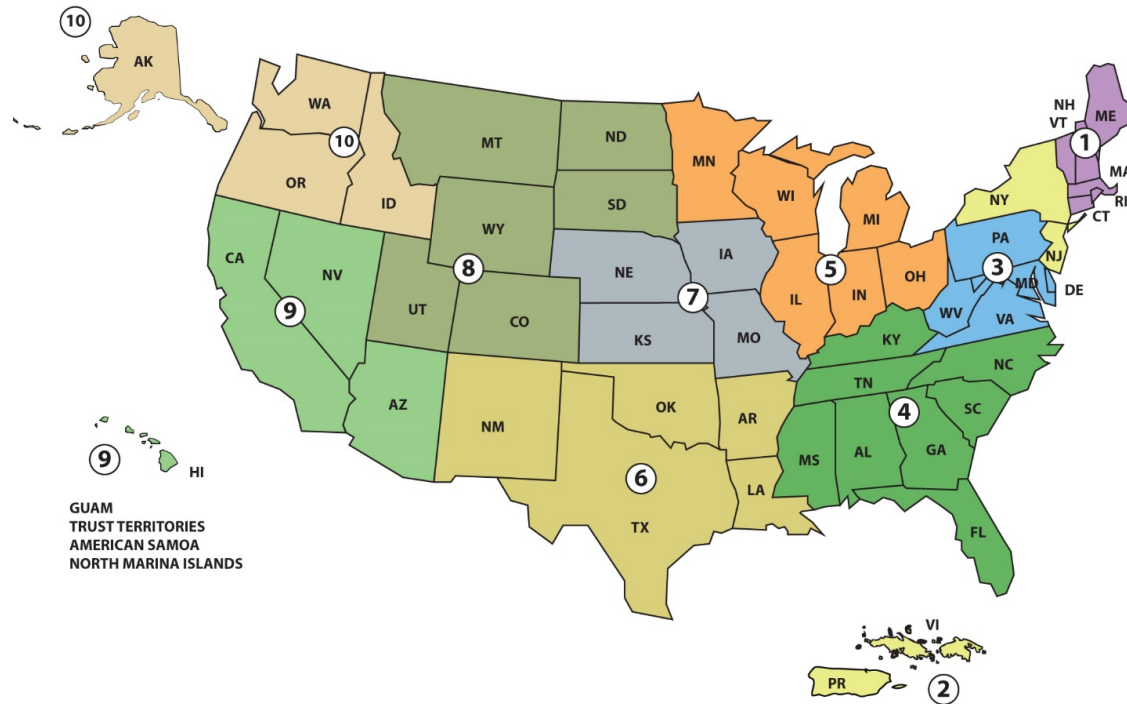
- Systematic review of assessment methodologies

- Community driven requirements - TADA working group (41 participants – not including WDIB folks)
  - 12 total EPA and USGS staff (HQ)
  - Staff from EPA regions 2, 3, 6, 7, and 8
  - State(s) from all regions except region 3
  - Tribe(s) from regions 1, 6 and 8

- Supplemental 1-on-1 calls with other stakeholders

Scope of TADA

- WQP data

- Sampling station or assessment unit spatial scale (or both with additional spatial aggregation requirements needed)

- Assessment units are matched with designated uses

- Criteria/standards and methodologies are assigned to a paired parameter and use

- Analysis results and assessment determinations (ATTAINS compatible)

INPUTS & OUTPUTS

R Shiny Input File Generators

INPUTS
WQP Monitoring Data
Data Cleaning & Preparation Logic
Assessment Units
Designated Uses
Criteria/Standards & Methodologies
Previous Decisions & Human Review
Impairment/Attainment Logic

ANALYSIS
R package

FEEDBACK
Information, new requirements, expertise, and user feedback

OUTPUT TABLES
A) Parameter and Designated Use (duration and magnitude criteria)
B) Parameter and Designated Use (frequency criteria, PARAM_ATTAINMENT_CODE)
C) Designated Use (USE_ATTAINMENT_CODE)
Overall Assessment Unit Determinations (supporting or not)

TADA Inputs

INPUTS
WQP Monitoring Data
Data Cleaning & Preparation Logic
Assessment Units
Designated Uses
Criteria/Standards & Methodologies
Previous Decisions & Human Review
Impairment/Attainment Logic

- WQP monitoring data
  - Single sample data (not high frequency continuous)
  - Sample media water (not biological, air, sediment, etc.)
  - Assessment period (e.g., typically 3-5 years)
  - Data cleaning and preprocessing
    - Quality control checks, unit conversations, synonym groupings, parameter groupings for multi-part assessments, etc.

- Sampling station or assessment unit spatial scale (or both with additional spatial aggregation requirements needed)
  - WQP monitoring stations are matched with assessment units
  - Options to assign site specific criteria if needed

- Assessment units and stations are matched with designated uses

TADA Inputs

INPUTS
WQP Monitoring Data
Data Cleaning & Preparation Logic
Assessment Units
Designated Uses
Criteria/Standards & Methodologies
Previous Decisions & Human Review
Impairment/Attainment Logic

Criteria/standards and methodologies are assigned to a paired parameter and use

Common methodologies are supported as opposed to specific parameters

Priority Methodologies

INPUTS

WQP Monitoring Data

Data Cleaning & Preparation Logic

Assessment Units

Designated Uses

Criteria/Standards & Methodologies

Previous Decisions & Human Review

Impairment/Attainment Logic

- Magnitude, duration (temporal aggregation)
  - 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)

- Address non-detects

- Minimum sample size

- Aggregate depth profile data

- Acute vs chronic

- Seasonality

TADA Outputs: Assessment determinations

- ATTAINS compatible tables

A. Parameter and use
  ◦ Meeting/not meeting magnitude & duration criteria

B. Parameter and use
  ◦ Meeting/not meeting frequency criteria

C. Overall designated use
  ◦ Supported or not supported
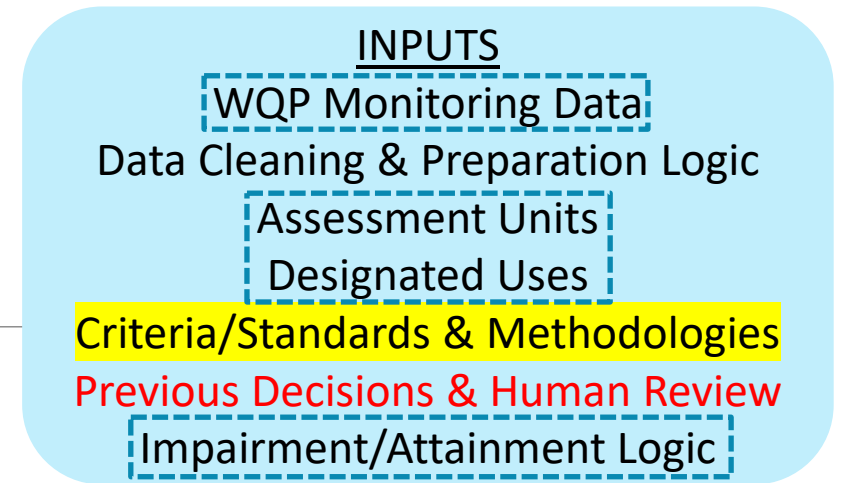
INPUTS
WQP Monitoring Data
Data Cleaning & Preparation Logic
Assessment Units
Designated Uses
Criteria/Standards & Methodologies
Previous Decisions & Human Review
Impairment/Attainment Logic

ANALYSIS
R package

OUTPUT TABLES
A) Parameter and Designated Use (duration and magnitude criteria)
B) Parameter and Designated Use (frequency criteria, PARAM_ATTAINMENT_CODE)
C) Designated Use (USE_ATTAINMENT_CODE)
Overall Assessment Unit Determinations (supporting or not)

Conceptual Example

DO and AQU

1 row of the DefineCriteriaAndMethods file

Joined data frame with WQP results, station/AU, and use data

Magnitude and Duration Analysis

Magnitude and duration analysis results will be available for all parameter and use combinations

A

Characteristic	Use	Duration-Result Value	Duration-Criteria	Magnitude-Criteria Value	Magnitude-Criteria Context	Y= meeting criteria; N=not meeting criteria
DO	AQU	5 mg/L	7-day average	3 mg/L	LowerLimit	Y
DO	AQU	7 mg/L	7-day average	3 mg/L	LowerLimit	Y

Frequency Analysis (PARAM_ATTAINMENT_CODE)

B

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

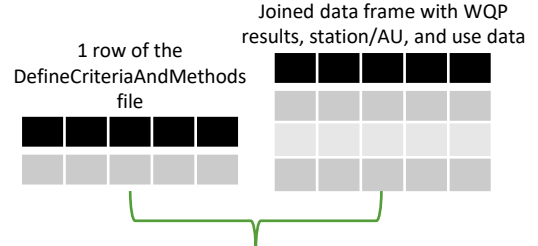Frequency analysis results will be available for all parameter and use combinations

pH and AQU

1 row of the DefineCriteriaAndMethods file

Joined data frame with WQP results, station/AU, and use data

Magnitude and Duration Analysis

Characteristic	Use	Duration-Result Value	Duration-Criteria	Magnitude-Criteria Value	Magnitude-Criteria Context	Y= meeting criteria; N=not meeting criteria
pH	AQU	3	30-day Min	6-8	Range	N
pH	AQU	5	30-day Min	6-8	Range	N

Frequency Analysis (PARAM_ATTAINMENT_CODE

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

C

Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

AU or station use analysis results will be provided for all designated uses

A

Characteristic	Use	Duration-Result Value	Duration-Criteria	Magnitude-Criteria Value	Magnitude-Criteria Context	Y=meeting criteria; N=not meeting criteria
DO	AQU	5 mg/L	7-day average	3 mg/L	LowerLimit	Y
DO	AQU	7 mg/L	7-day average	3 mg/L	LowerLimit	Y

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

Frequency analysis results will be available for all parameter and use combinations

B

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

C

Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

AU or station use analysis results will be provided for all designated uses

DO

Parameter 3-8

1 row of the DefineCriteriaAndMethods file

Magnitude... A...

A

Main table (in red box):

Characteristic	Use	Duration-Result Value	Duration-Criteria	Magnitude-Criteria Value	Magnitude-Criteria Context	Y=meeting criteria; N=not meeting criteria
pH	AQU	3	30-day Min	6-8	Range	N
pH	AQU	5	30-day Min	6-8	Range	N

Left table (partially obscured):

Characteristic	Use	Duration-Result Value	Dura... Cri...
DO	AQU	5 mg/L	7-... ave...
DO	AQU	7 mg/L	7-... ave...

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

Frequency analysis results will be available for all parameter and use combinations

B

Frequency Analysis (PARAM_ATTAINMENT_CODE

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

C

Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

AU or station use analysis results will be provided for all designated uses

DO

1 row of the
DefineCriteriaAndMe...
file

Magnitude...
A...

A

Characteristic	Use	Duration-Result Value	Dur... Cri...
DO	AQU	5 mg/L	7-... ave...
DO	AQU	7 mg/L	7-... ave...

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Frequency Analysis (PARAM_ATTAINMENT_CODE)

B

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAIN MENT_CODE	ParameterStatus
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

Frequency analysis results will
be available for all parameter
and use combinations

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAIN MENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

C

Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

AU or station use analysis
results will be provided for
all designated uses

DO ... **Parameter 3.8** ...

1 row of the DefineCriteriaAndMe... file

Magnitude ...
A...

(A)

Characteristic	Use	Duration-Result Value	Dur... Cri...
DO	AQU	5 mg/L	7-... ave...
DO	AQU	7 mg/L	7-... ave...

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Frequency Analysis (PARAM_ATTAINMENT_CODE)

(B)

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

Frequency analysis results will be available for all parameter and use combinations

Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

(C)

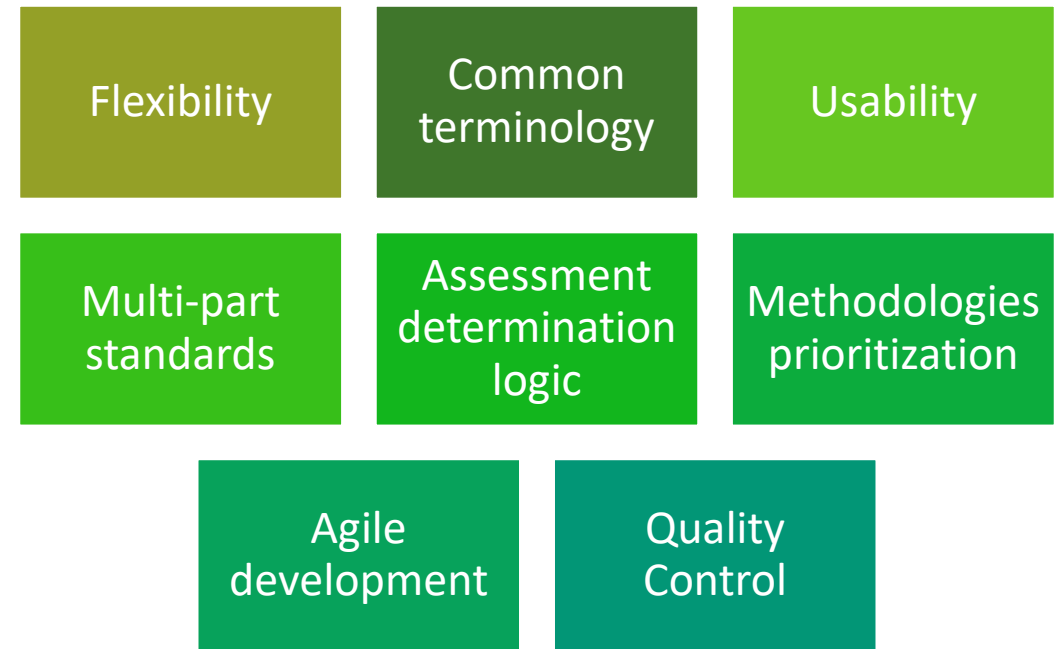| Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

AU or station use analysis results will be provided for all designated uses

1 row of the DefineCriteriaAndMe... file

DO...

Parameter 3.8...

Magnitu...
A...

A

Characteristic	Use	Duration-Result Value	Dur... Cri...
DO	AQU	5 mg/L	7-... ave...
DO	AQU	7 mg/L	7-... ave...

Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Frequency Analysis (PARAM_ATTAINMENT_CODE)

Characteristic	Use	Freq-result (# or % exceptions over assessment period)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

Frequency analysis results will be available for all parameter and use combinations

B

Characteristic	Use	Freq-result (# or % exceptions over assessment period)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
D...	AQU	15%	10%, percentile	Not meeting criteria	Cause

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

C

Use	USE_ATTAINMENT_CODE	AU or Station
AQU	Not Supporting	Station

AU or station use analysis results will be provided for all designated uses

Summary

PROGRESS

Requirements gathering	Data source compatibility	Draft input & output files
Cataloging methodologies	Documentation of functionality	Data retrieval
	Geoplatform station/AU associations	

NEXT STEPS & CHALLENGES

Flexibility	Common terminology	Usability
Multi-part standards	Assessment determination logic	Methodologies prioritization
Agile development	Quality Control	
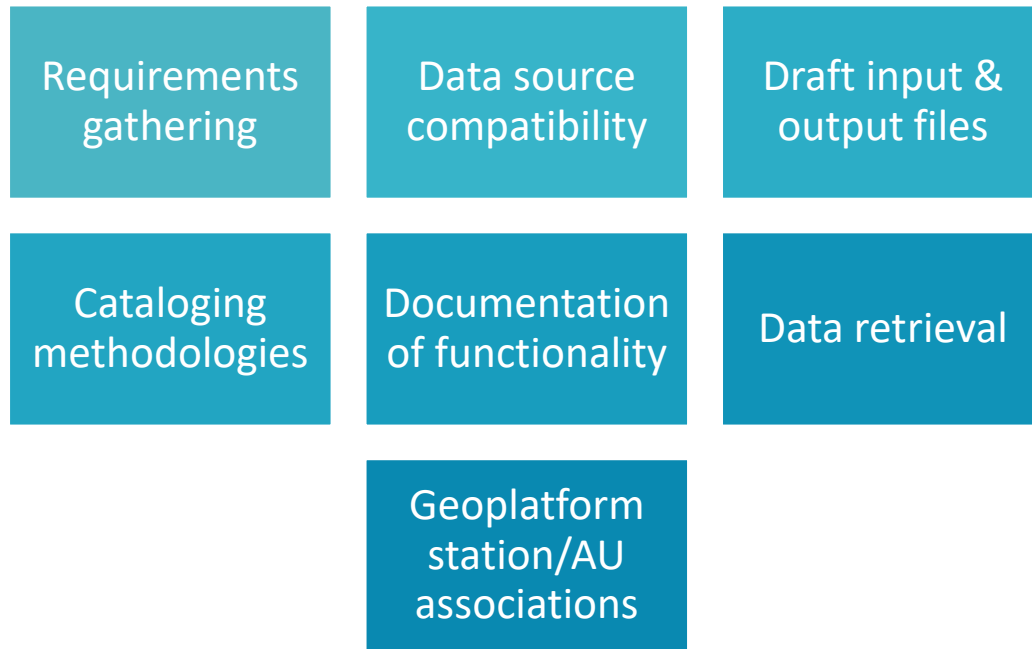
Questions?

TADA GeoViewer (Match Stations with AUs Input File Generator)

Requirements/Community Requests

- Manually match WQP stations with Assessment Units (AUs)

- Automatically match WQP stations with ATTAINS AUs

- Automatically match WQP stations with user uploaded AU layers

- Add/remove stations to/from an AU

- Exclude stations

- Group stations (e.g., with limited data)

- Review station data availability (e.g., results for each parameter group)

- **The excel file output from the TADA GeoViewer can be directly used as an input file for the TADA R package**

TADA GeoViewer Demo

Questions?

Why R?

- Open-source

- Powerful programming language for statistics

- Wealth of online resources

- Flexible and manageable packages

- R Shiny- graphical user interface

- One draw back
  ◦ Steep learning curve for nonprogrammers
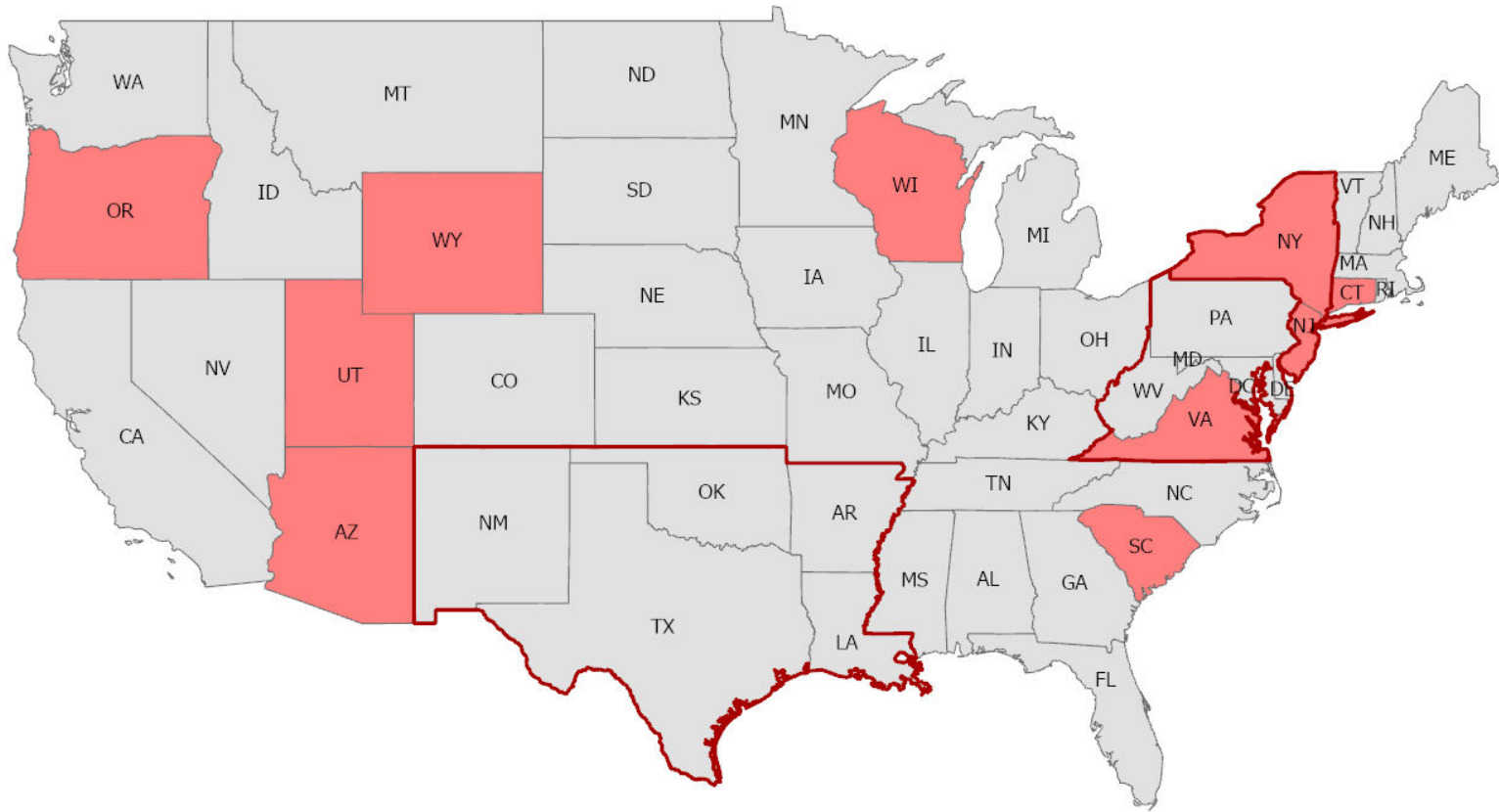
Example Assessment Tools Around the Country

- EPA Regions 2, 3, and 6

- Tribal Nations
  - Penobscot Nation

- Tools vary in...
  - Scope
  - Application
  - Stage of development

TADA Architecture

Workflow

Data Retrieval & Cleaning → User/Auxiliary Inputs → Analysis → ATTAINS Compatible Output
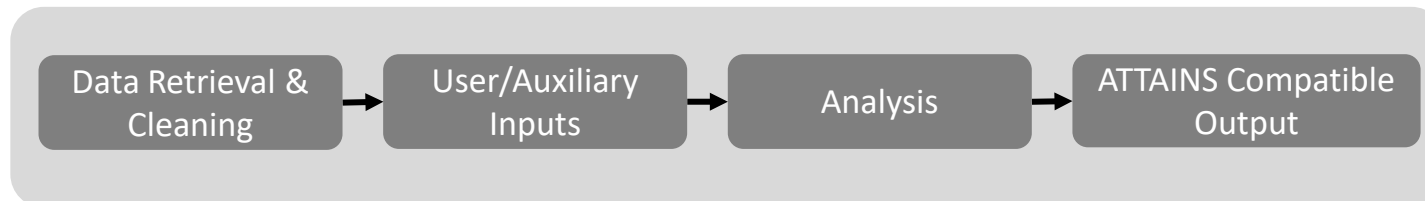
R Shiny Widgets

Each one is standalone
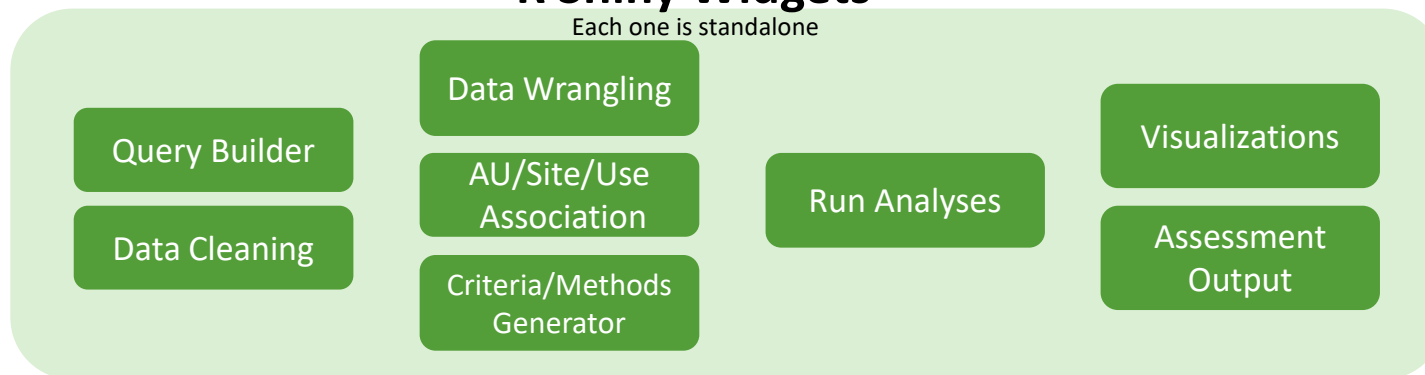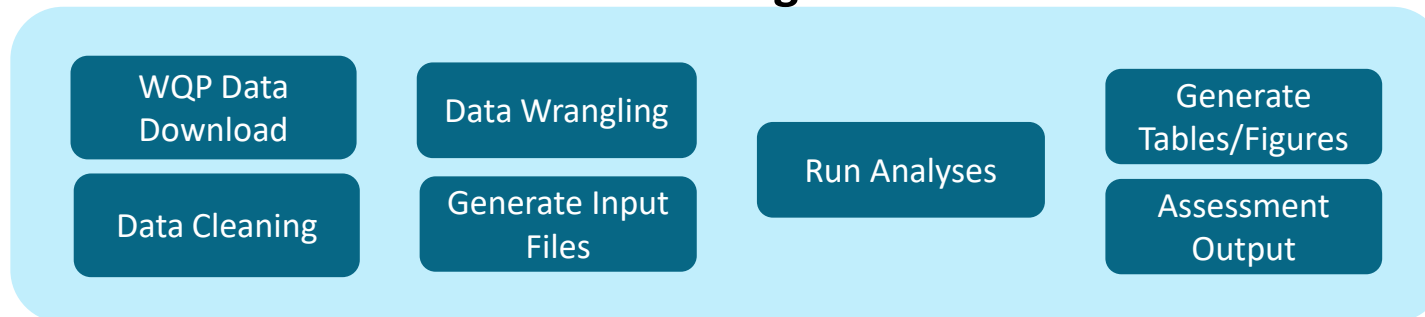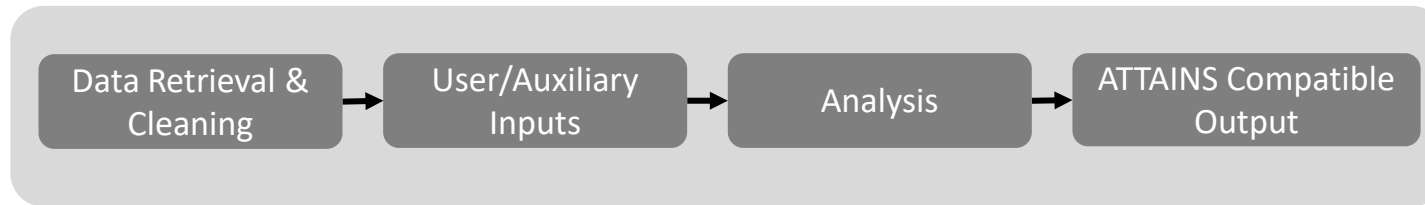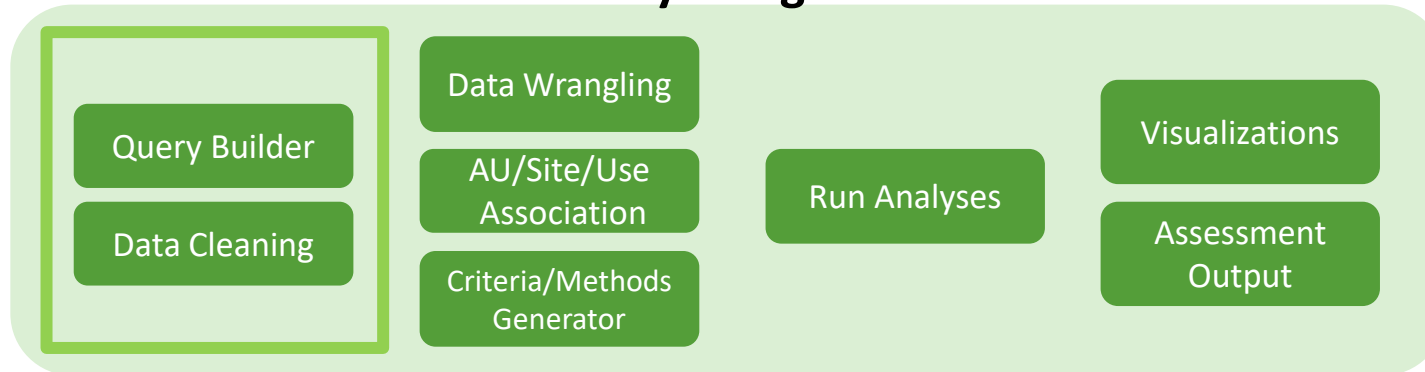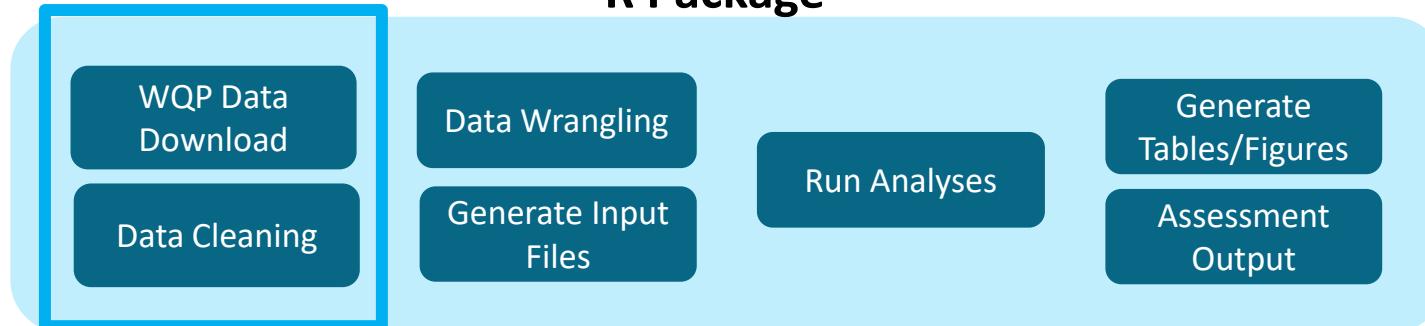
Query Builder

Data Cleaning

Data Wrangling

AU/Site/Use Association

Criteria/Methods Generator

Run Analyses

Visualizations

Assessment Output

R Package

WQP Data Download

Data Cleaning

Data Wrangling

Generate Input Files

Run Analyses

Generate Tables/Figures

Assessment Output

TADA Architecture

TADA Architecture

Workflow

| Data Retrieval & Cleaning | → | User/Auxiliary Inputs | → | Analysis | → | ATTAINS Compatible Output |

R Shiny Widgets

- Query Builder
- Data Cleaning
- Data Wrangling
- AU/Site/Use Association
- Criteria/Methods Generator
- Run Analyses
- Visualizations
- Assessment Output

R Package

- WQP Data Download
- Data Cleaning
- Data Wrangling
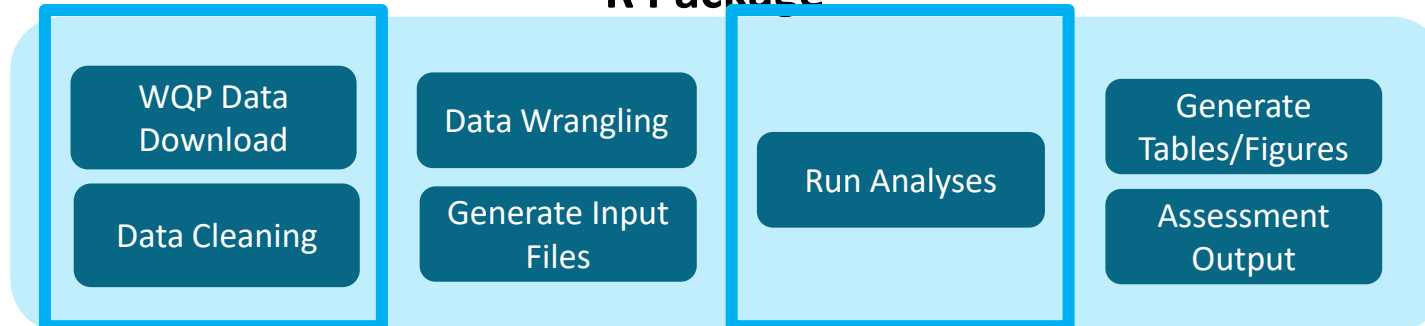- Generate Input Files
- Run Analyses
- Generate Tables/Figures
- Assessment Output
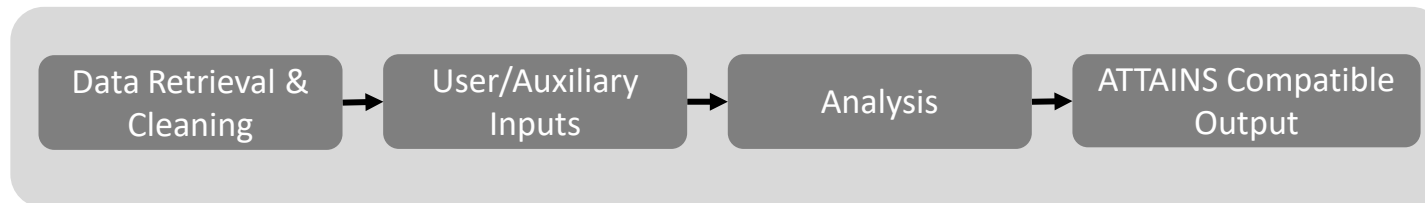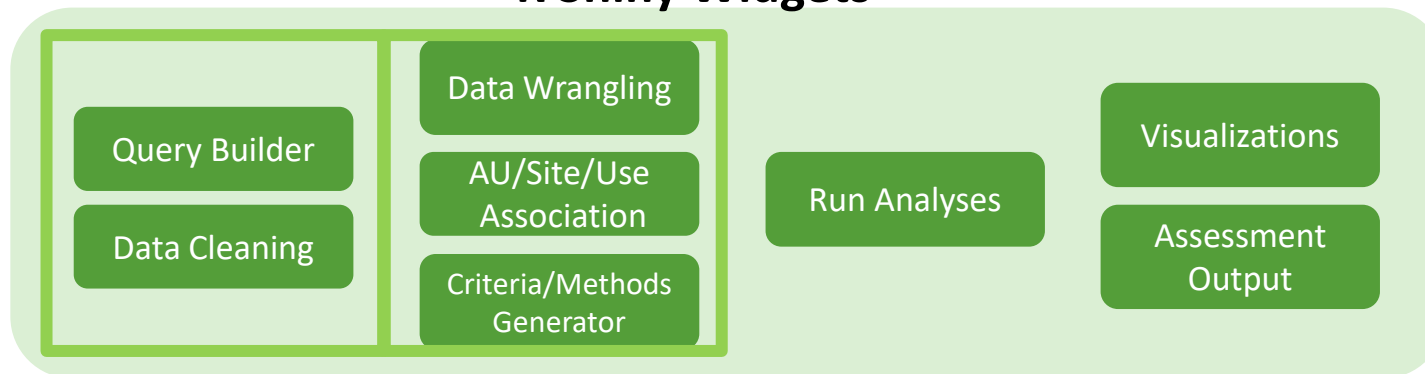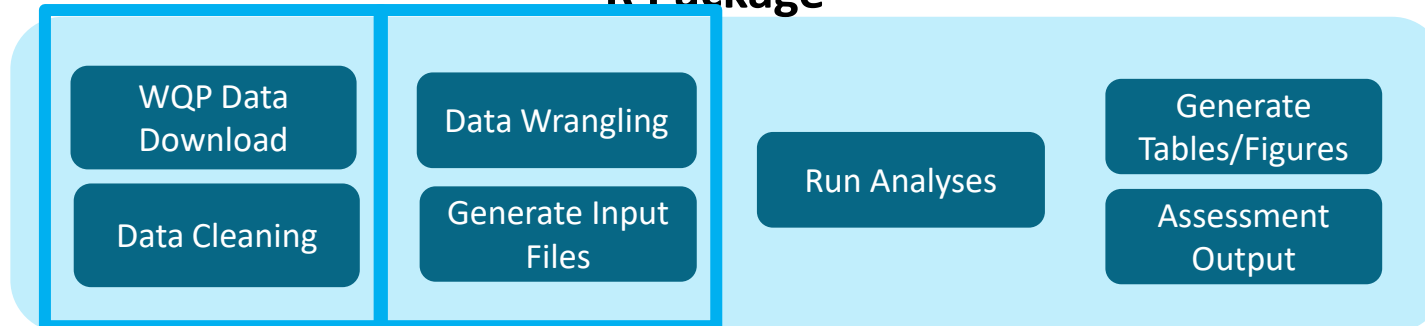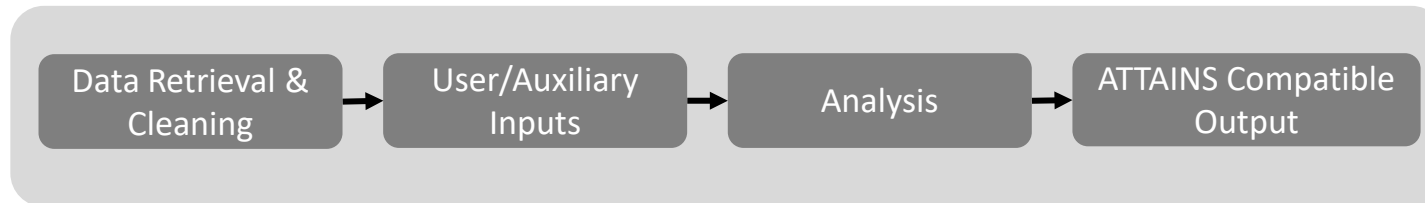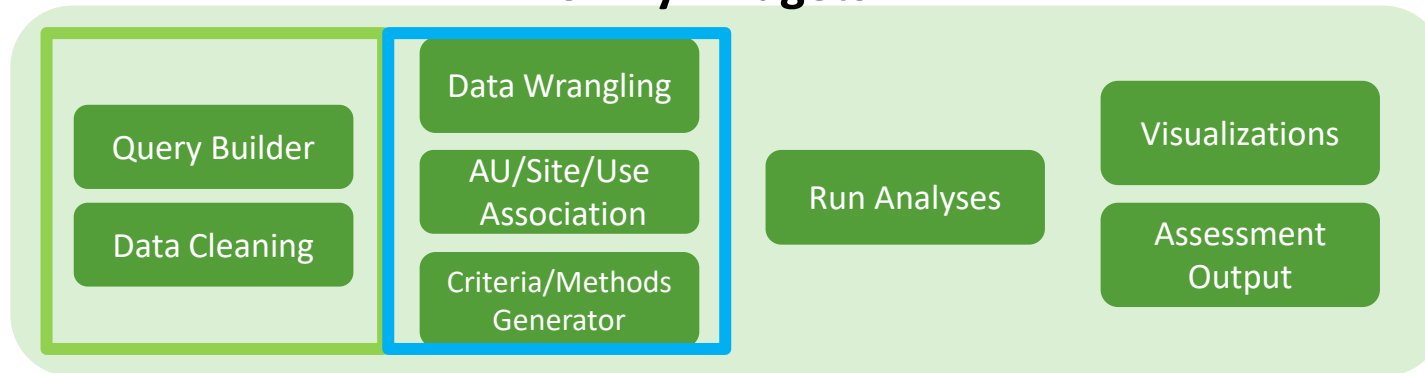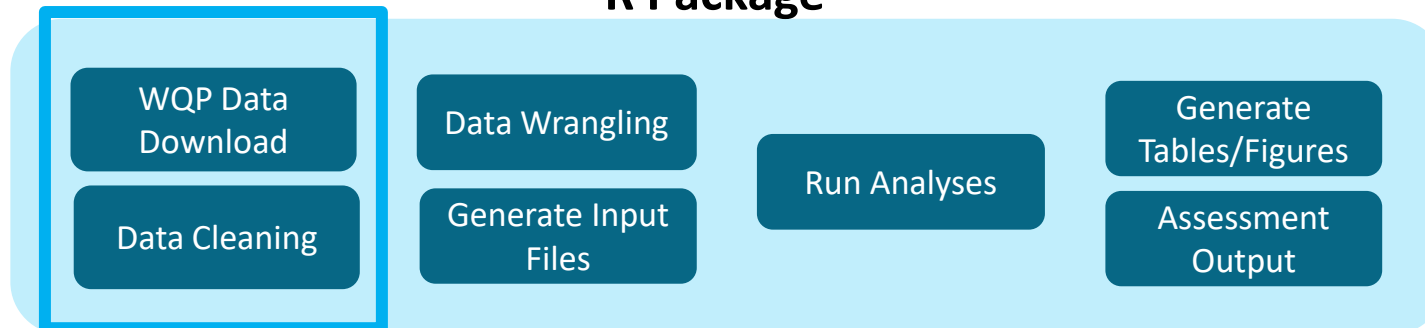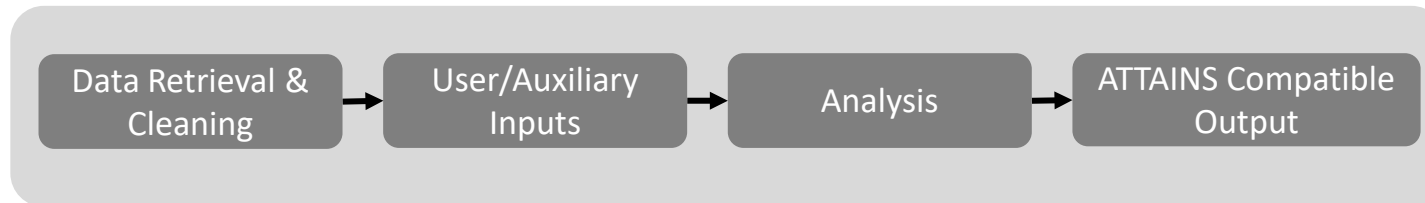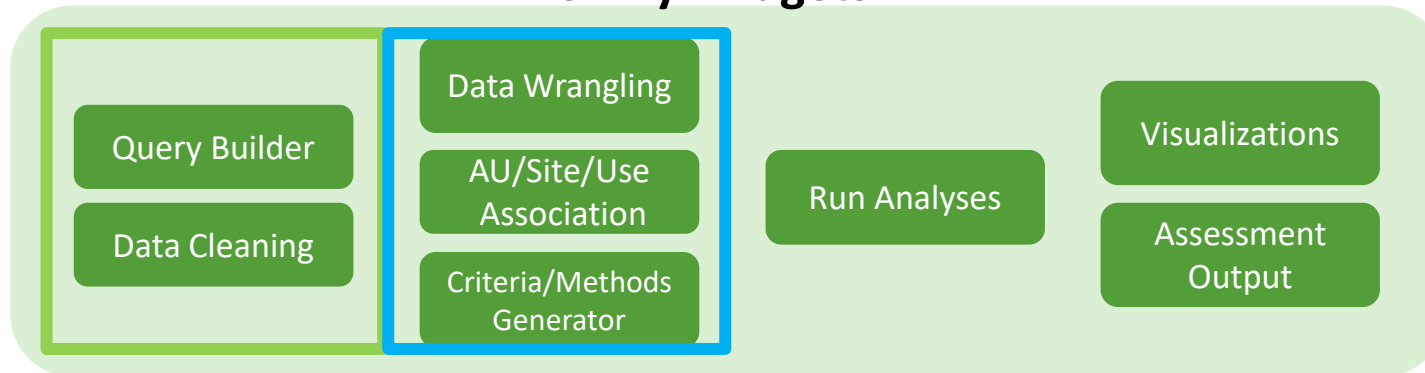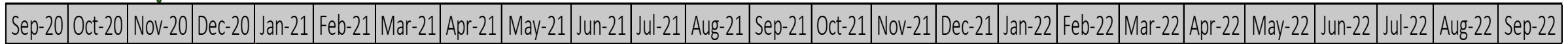
TADA Project 2-Year Development Plan

Waterfall TADA development: Gather requirements & design tool architecture (perform research & stakeholder engagement)

Agile TADA development: Share modules/widgets after each step of development, perform usability testing, seek continuous feedback, make improvements, train and assist users, and maintain TADA over time

Requirements | Design | Development | Testing

Cumulative outcomes
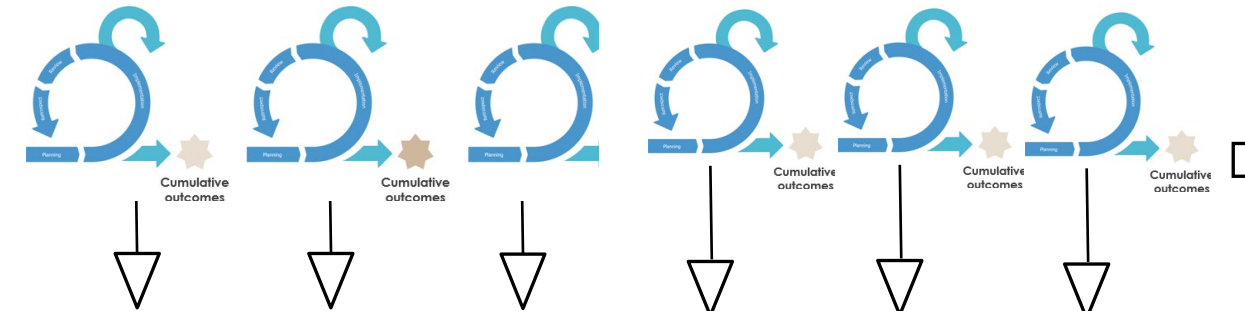
Host regular meetings with open-source water quality data analysis community to deploy tool

| Sep-20 | Oct-20 | Nov-20 | Dec-20 | Jan-21 | Feb-21 | Mar-21 | Apr-21 | May-21 | Jun-21 | Jul-21 | Aug-21 | Sep-21 | Oct-21 | Nov-21 | Dec-21 | Jan-22 | Feb-22 | Mar-22 | Apr-22 | May-22 | Jun-22 | Jul-22 | Aug-22 | Sep-22 |

Assemble project team, begin TADA conceptualization and management

TADA working group end date

TADA working group start date

- Started with waterfall approach, moving to agile

- EPA-led with future contract support

- Engaging active community in code development and testing

Value of this work

- Increase efficiencies, save time, and reduce costs of WQ assessments

- Improve data use and data system interoperability
  - OST's Criteria Search Tool
  - ATTAINS
  - WQX

- Systematic review of and increased communication about assessment processes & methodologies better enable science-based water resource decisions

- Open-source tools provide a path for collaboration across EPA regions, states, tribal nations, and other stakeholders

Connect with us

Mullin.Cristina@epa.gov

Greif.Jacob@epa.gov