TADA (Tools for Automated Data Analysis)

Cristina Mullin, PhD

Water Data Integration Branch

EPA Office of Water

Agenda

- Installation (~5 min)
- Presentation (~25 min)
- Hands-on/Q&A (~60 min)

TADA Installation (Demo)

- TADA R Package: https://github.com/USEPA/TADA
  - User Guide: https://usepa.github.io/TADA/
- TADA R Shiny App: https://github.com/USEPA/TADAShiny
  - Dev web application: https://owshiny-dev.app.cloud.gov/tada-dev/
    - NOTE: this space is still under development and currently may not handle multiple instances of the app at the same time. Recommend downloading locally for now, but feel free to try it. 😊
- Working Group SharePoint
- Inventory of open-source R code and WQP tools – please add any relevant packages/tools you are aware of: Inventory

Reusable

Interoperable

Accessible

Findable

Monitor Waters

106, 319

WQX/WQP

(Share Water Data)

106 grant req.

Water quality monitoring data

Supporting CWA 106, 319, 305(b), 303(d)

TADA

(Analyze Water Data)

Criteria Search Tool

Water quality standards information

ATTAINS

(Share Water Conditions)

106/305(b)/303(d)

How's My Waterway

(Discover Water Conditions)

Hydrologically-referenced designated beneficial uses

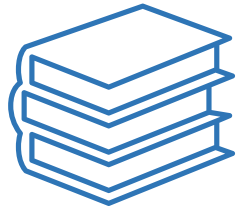
Vision: Efficient and Reproducible Water Quality Assessments

- Interconnected data resources and tools
- Streamlined water quality assessment and reporting

Water quality monitoring data

Hydrologically-referenced designated beneficial uses

Water quality standards information



Findable

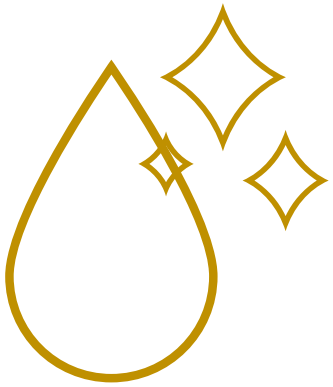
Accessible

Interoperable

Reusable

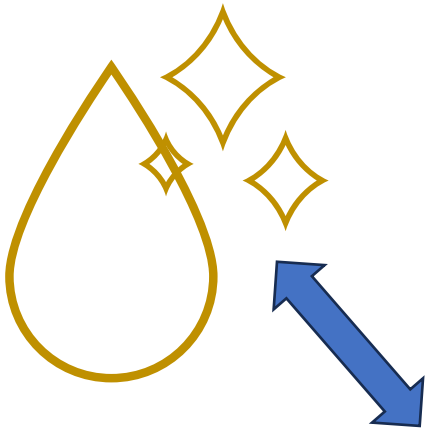
TADA Vision

Discover, wrangle, and QC data

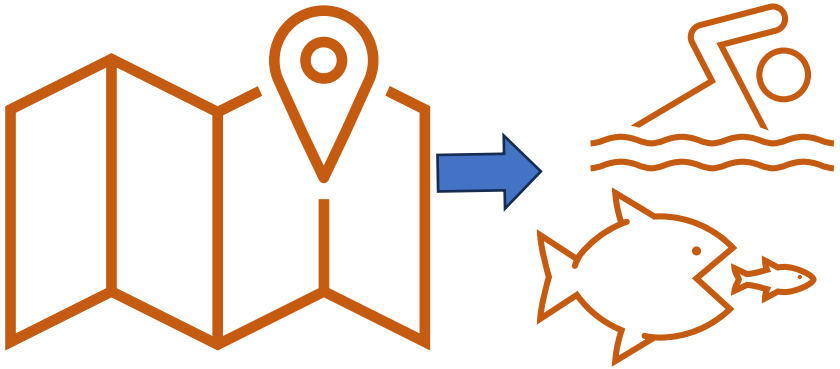


TADA Vision

Discover, wrangle, and QC data



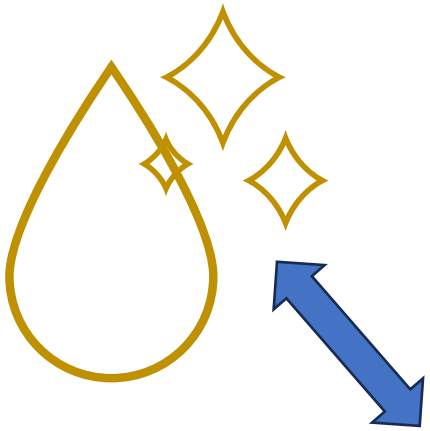
Assign beneficial uses



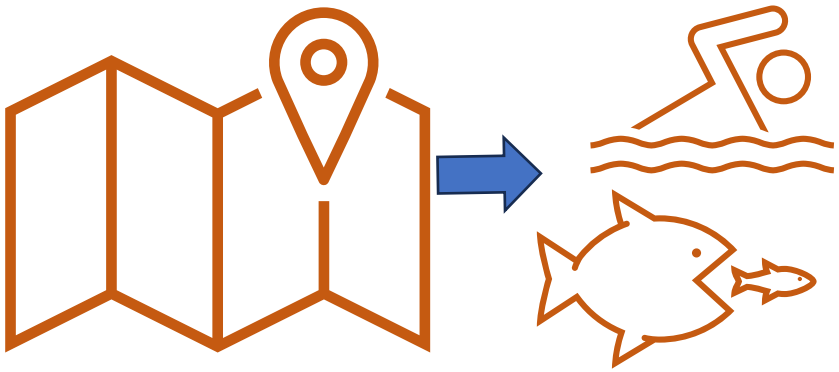
Assessment unit overlay with monitoring locations

TADA Vision

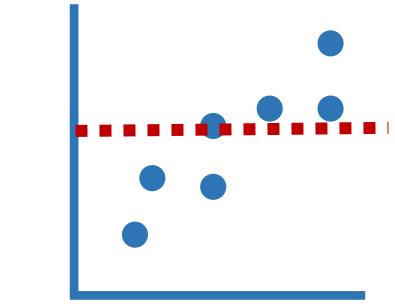
Discover, wrangle, and QC data



Assign beneficial uses



Assessment unit overlay with monitoring locations



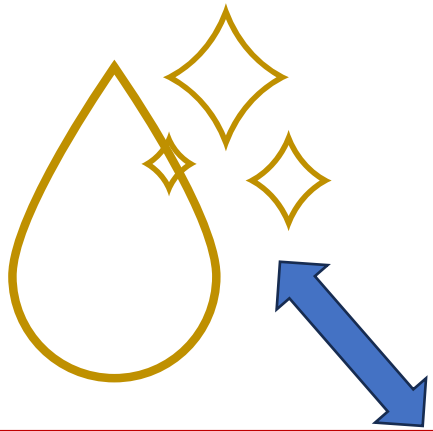
Beneficial uses determine numeric criteria used

Assessment methods guide impairment decisions based on:

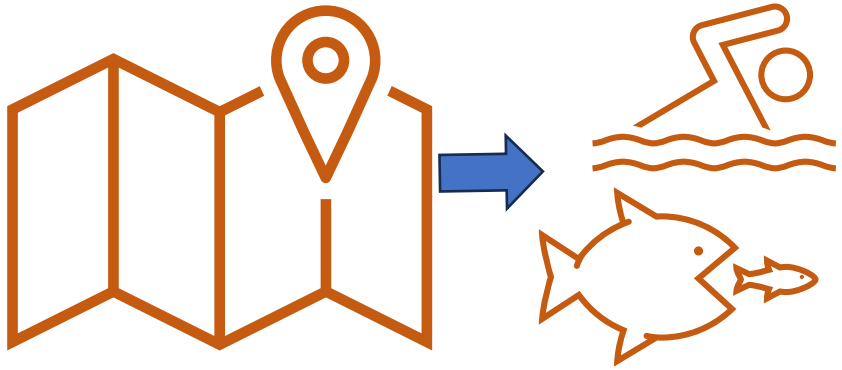
- Period of record
- Aggregated measurements
- Frequency
- Duration
- Magnitude
- Season
- Correction factors
- Covariates
- Site-specific criteria

TADA Vision

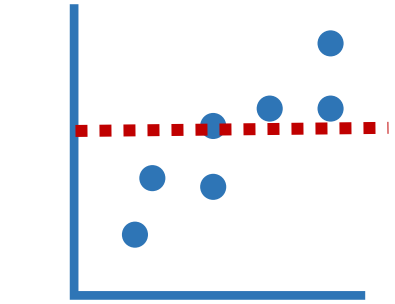
Discover, wrangle, and QC data



Assign beneficial uses



Assessment unit overlay with monitoring locations



Beneficial uses determine numeric criteria used

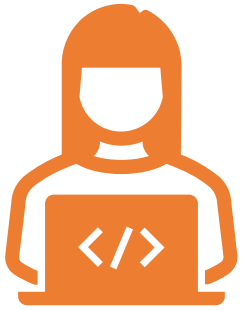
Assessment methods guide impairment decisions based on:

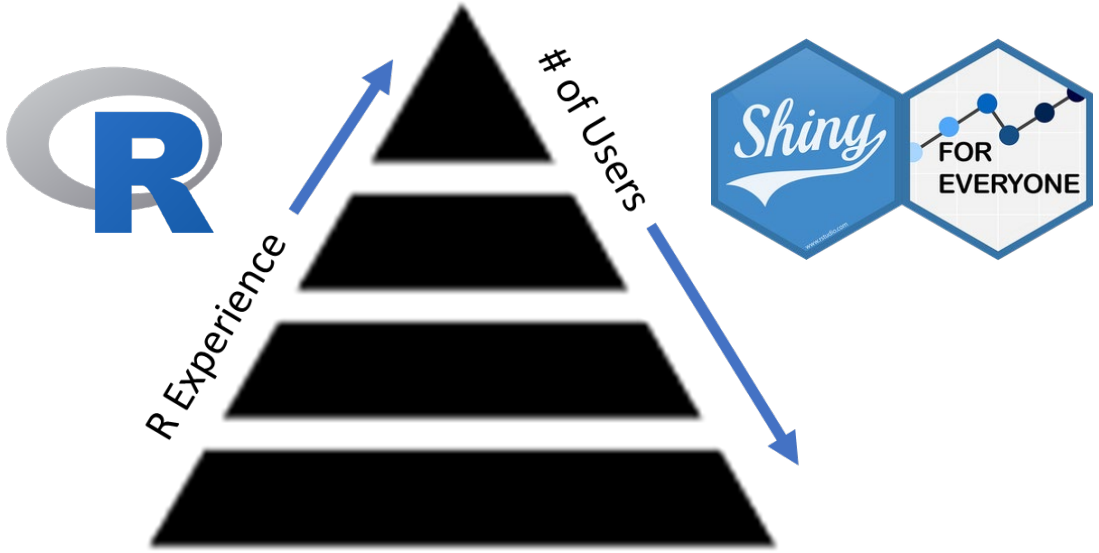
- Period of record
- Aggregated measurements
- Frequency
- Duration
- Magnitude
- Season
- Correction factors
- Covariates
- Site-specific criteria

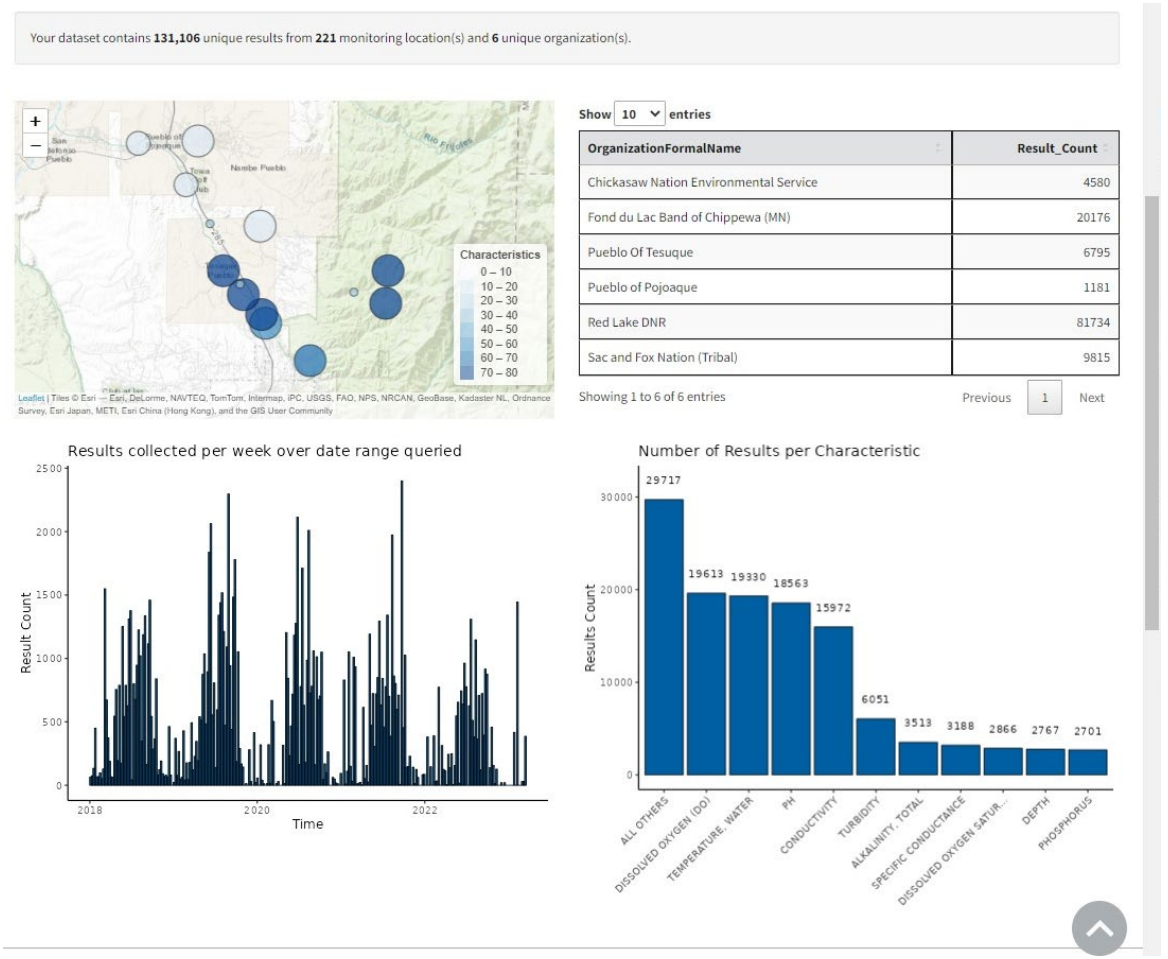
TADA (Tools for Automated Data Analysis)

Tools vary in…

- Scope
- Users/audience
- Stage of development







TADA Vision

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**

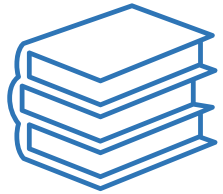
TADA Working Group: Mission

Working Group Mission Statement: 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.

Accessible, QC'd data

Hydrologically-referenced designated beneficial uses

Assessment Criteria & Methods Compendium



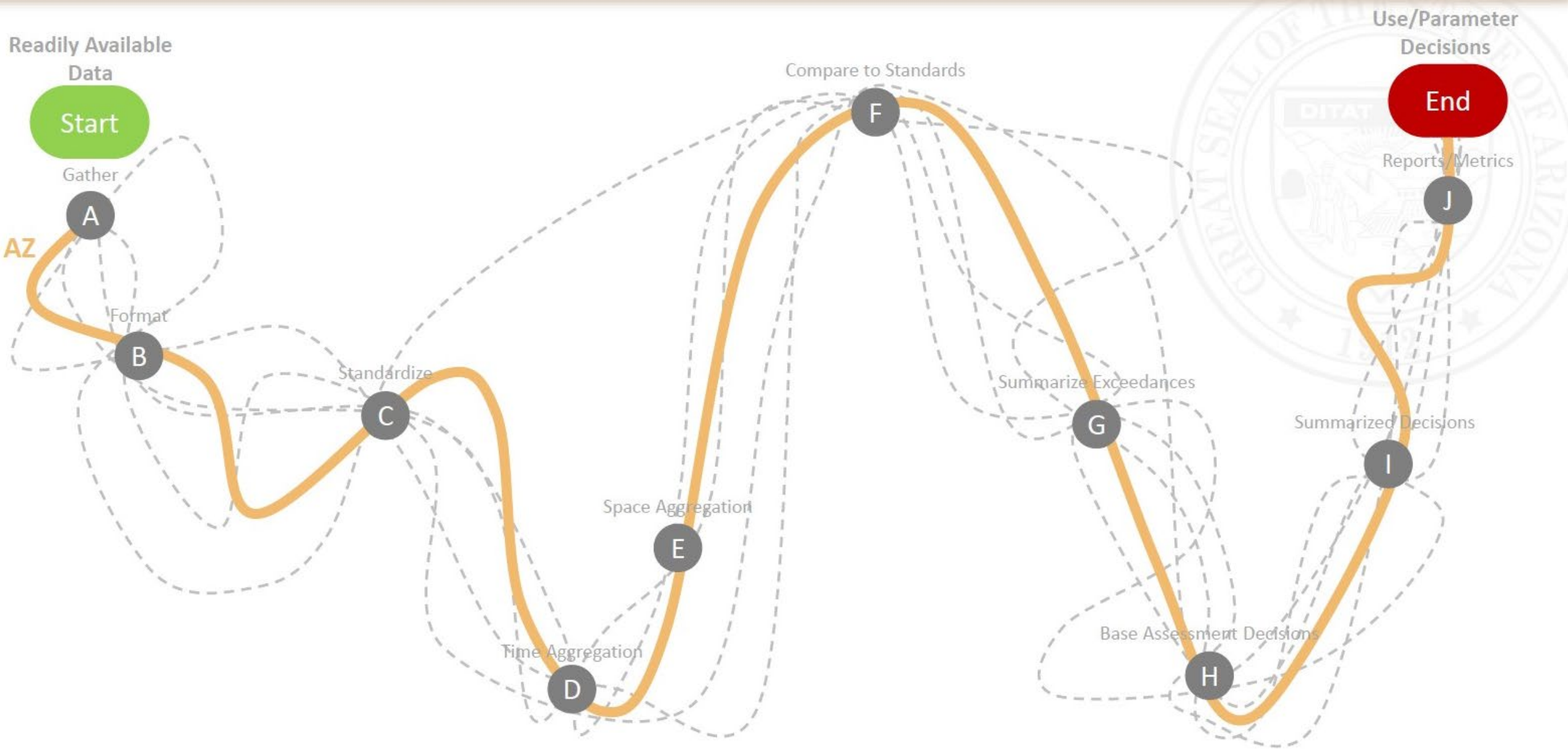
"Automating 80% of the workload is helpful"

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

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)

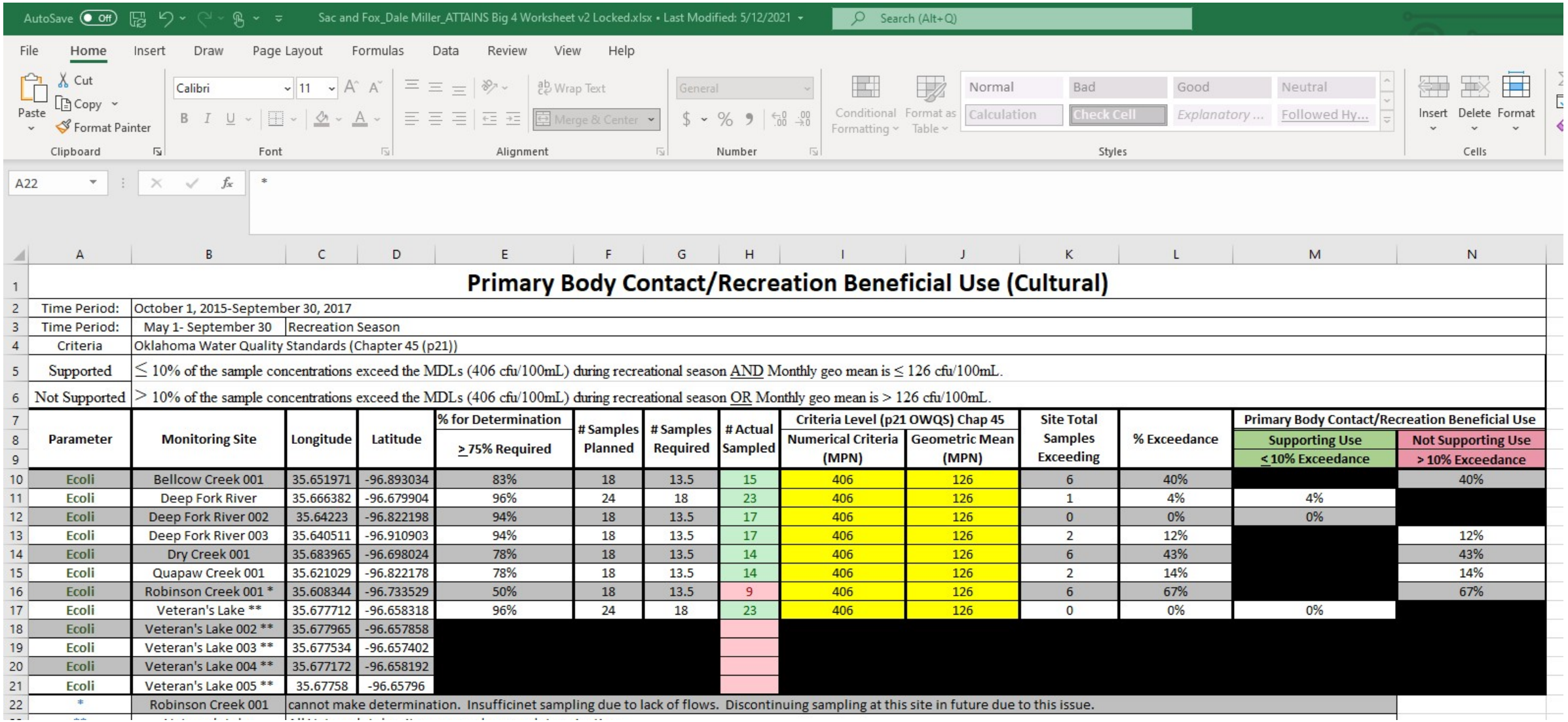
Getting There from Here



Getting There from Here

ADEQ Arizona Department of Environmental Quality

Sac and Fox – ATTAINS Big 4 Excel Worksheet



AutoSave Off | Sac and Fox_Dale Miller_ATTAINS Big 4 Worksheet v2 Locked.xlsx • Last Modified: 5/12/2021 | Search (Alt+Q)

File Home Insert Draw Page Layout Formulas Data Review View Help

Cut Copy Format Painter Paste | Calibri 11 | Wrap Text | Merge & Center | General | Conditional Formatting | Format as Table | Normal Bad Good Neutral Calculation Check Cell Explanatory ... Followed Hy... | Insert Delete Format

Clipboard Font Alignment Number Styles Cells

A22 *

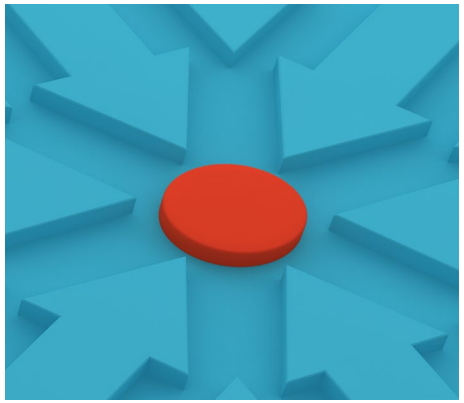
Primary Body Contact/Recreation Beneficial Use (Cultural)

Time Period:	October 1, 2015-September 30, 2017
Time Period:	May 1- September 30 Recreation Season
Criteria	Oklahoma Water Quality Standards (Chapter 45 (p21))
Supported	≤ 10% of the sample concentrations exceed the MDLs (406 cfu/100mL) during recreational season AND Monthly geo mean is ≤ 126 cfu/100mL.
Not Supported	> 10% of the sample concentrations exceed the MDLs (406 cfu/100mL) during recreational season OR Monthly geo mean is > 126 cfu/100mL.

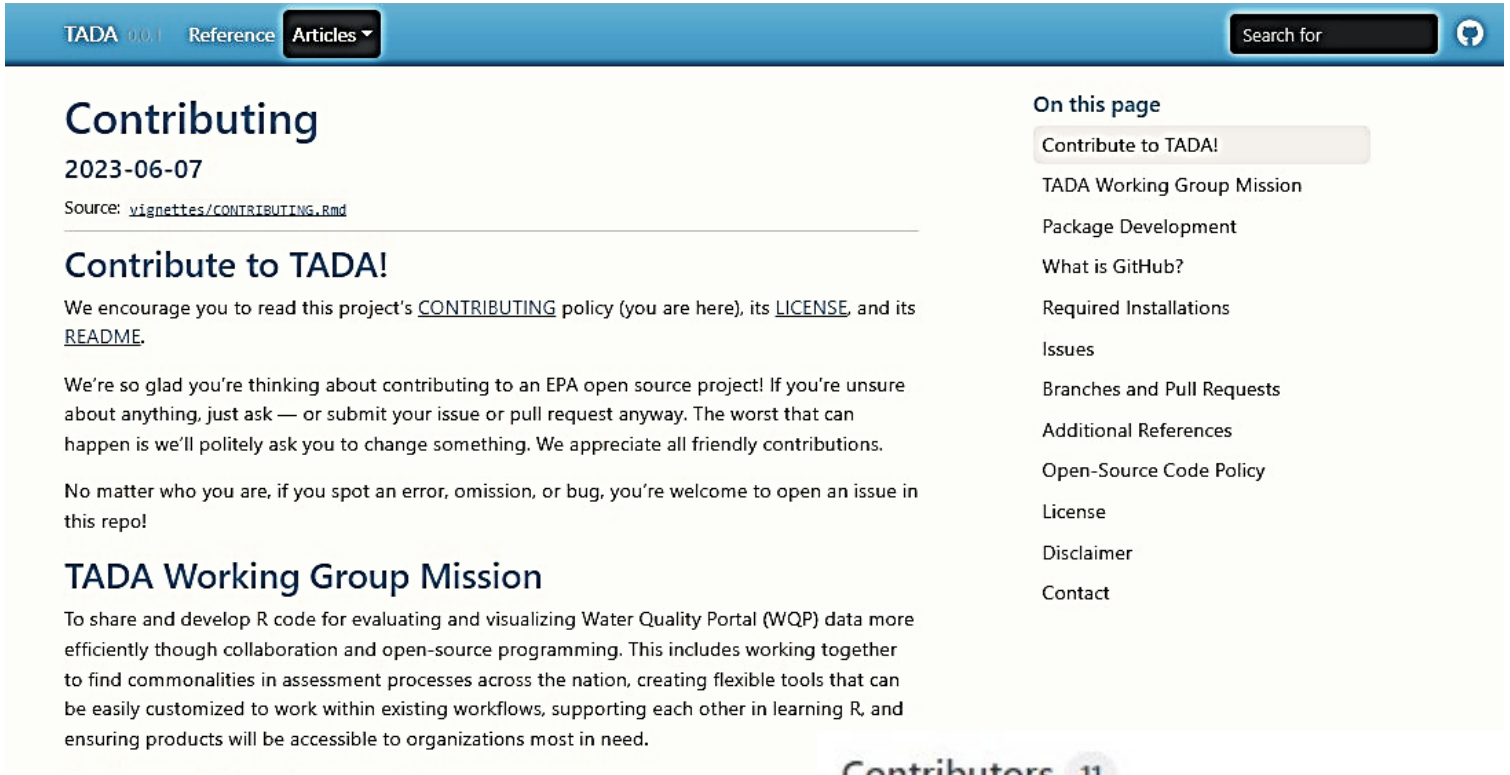
Parameter	Monitoring Site	Longitude	Latitude	% for Determination ≥ 75% Required	# Samples Planned	# Samples Required	# Actual Sampled	Criteria Level (p21 OWQS) Chap 45 Numerical Criteria (MPN)	Criteria Level (p21 OWQS) Chap 45 Geometric Mean (MPN)	Site Total Samples Exceeding	% Exceedance	Primary Body Contact/Recreation Beneficial Use: Supporting Use ≤ 10% Exceedance	Primary Body Contact/Recreation Beneficial Use: Not Supporting Use > 10% Exceedance
Ecoli	Bellcow Creek 001	35.651971	-96.893034	83%	18	13.5	15	406	126	6	40%		40%
Ecoli	Deep Fork River	35.666382	-96.679904	96%	24	18	23	406	126	1	4%	4%	
Ecoli	Deep Fork River 002	35.64223	-96.822198	94%	18	13.5	17	406	126	0	0%	0%	
Ecoli	Deep Fork River 003	35.640511	-96.910903	94%	18	13.5	17	406	126	2	12%		12%
Ecoli	Dry Creek 001	35.683965	-96.698024	78%	18	13.5	14	406	126	6	43%		43%
Ecoli	Quapaw Creek 001	35.621029	-96.822178	78%	18	13.5	14	406	126	2	14%		14%
Ecoli	Robinson Creek 001 *	35.608344	-96.733529	50%	18	13.5	9	406	126	6	67%		67%
Ecoli	Veteran's Lake **	35.677712	-96.658318	96%	24	18	23	406	126	0	0%	0%	
Ecoli	Veteran's Lake 002 **	35.677965	-96.657858										
Ecoli	Veteran's Lake 003 **	35.677534	-96.657402										
Ecoli	Veteran's Lake 004 **	35.677172	-96.658192										
Ecoli	Veteran's Lake 005 **	35.67758	-96.65796										
*	Robinson Creek 001	cannot make determination. Insufficinet sampling due to lack of flows. Discontinuing sampling at this site in future due to this issue.											

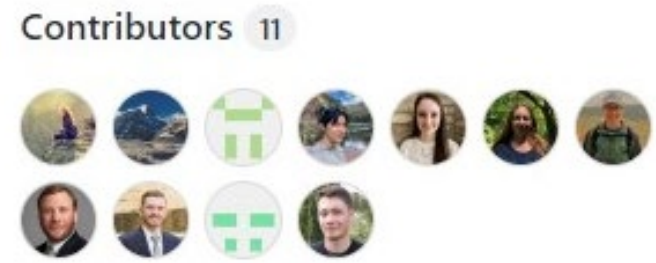
Involving end users in the development process

Agile & Community Development via GitHub

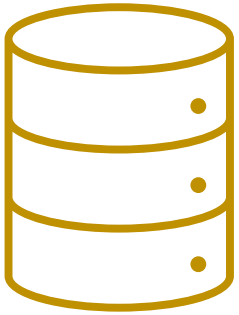


- GitHub: setting stage for community participation (two public repositories)
  - Community can be anyone
  - EPA/ORISE initial development
  - Contract to support subject experts with varying R skills

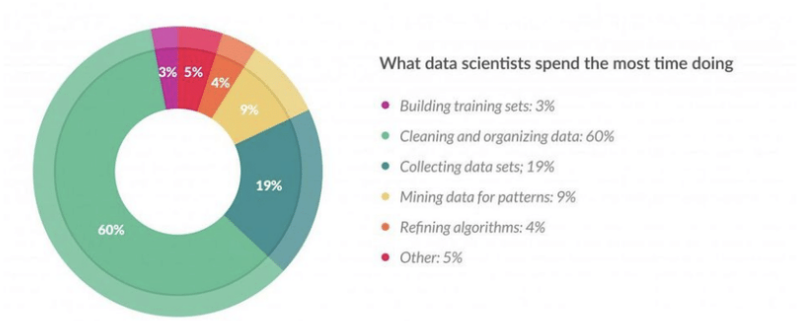




https://usepa.github.io/TADA/articles/CONTRIBUTING.html



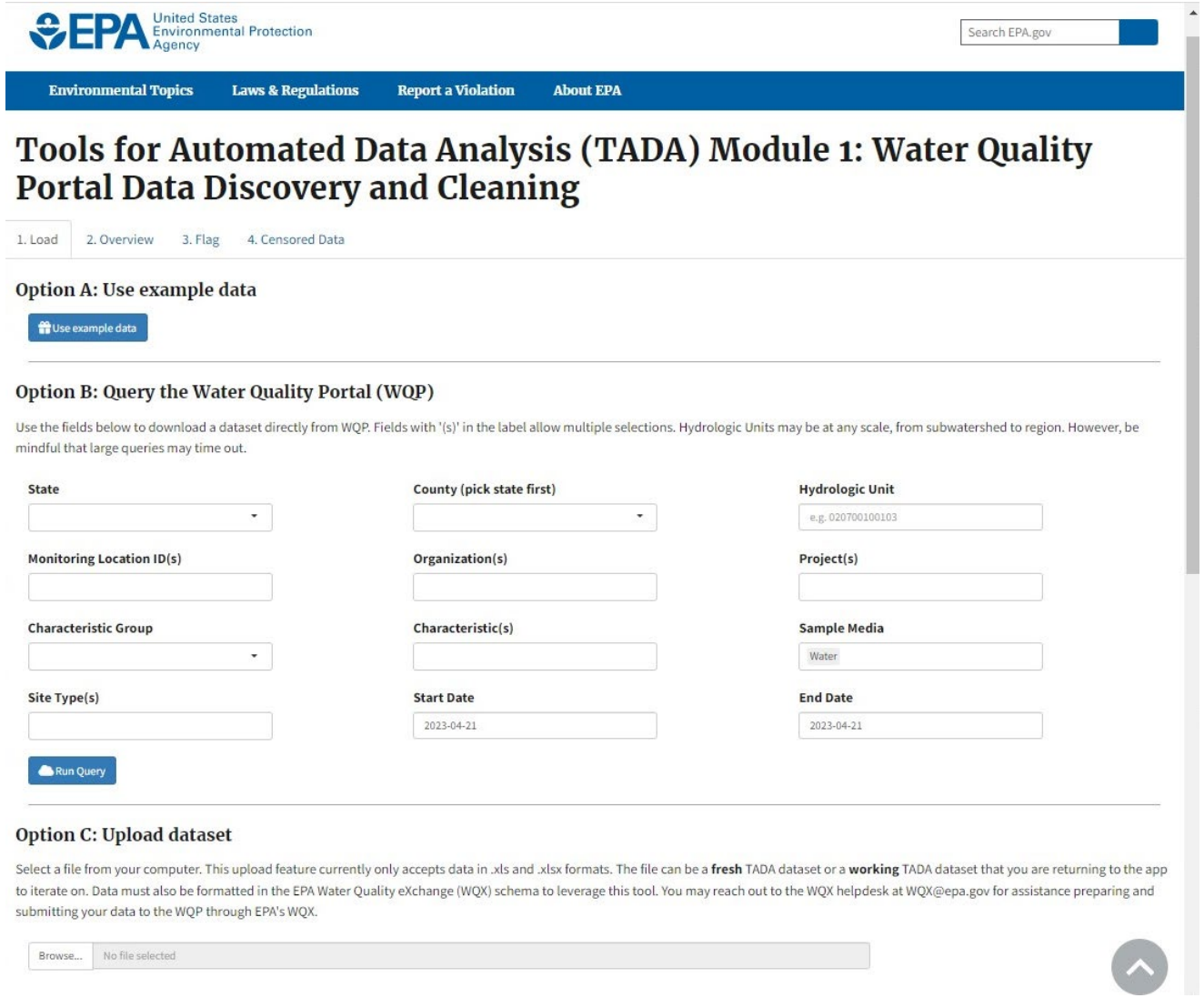
Data preparation *accounts for about 80% of the work of data scientists*



Module 1: Data Discovery and Wrangling

In Progress

Finding readily available data



Water Quality Portal (WQP)

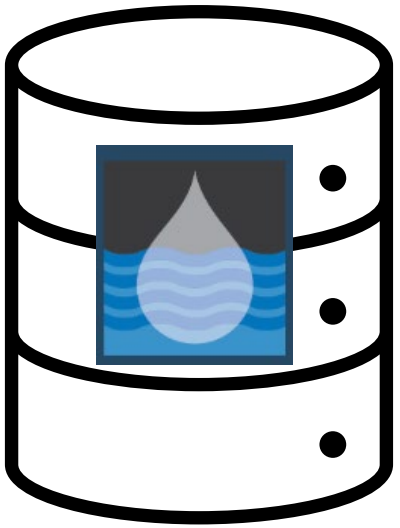
NATIONAL WATER QUALITY MONITORING COUNCIL

EPA United States Environmental Protection Agency

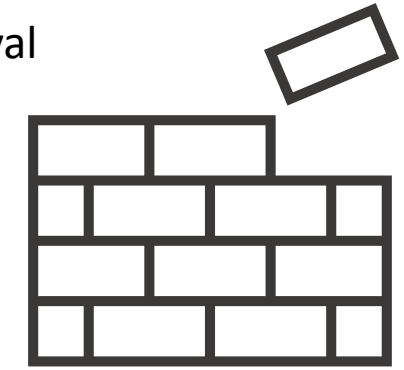
USGS science for a changing world

WQX WATER QUALITY EXCHANGE

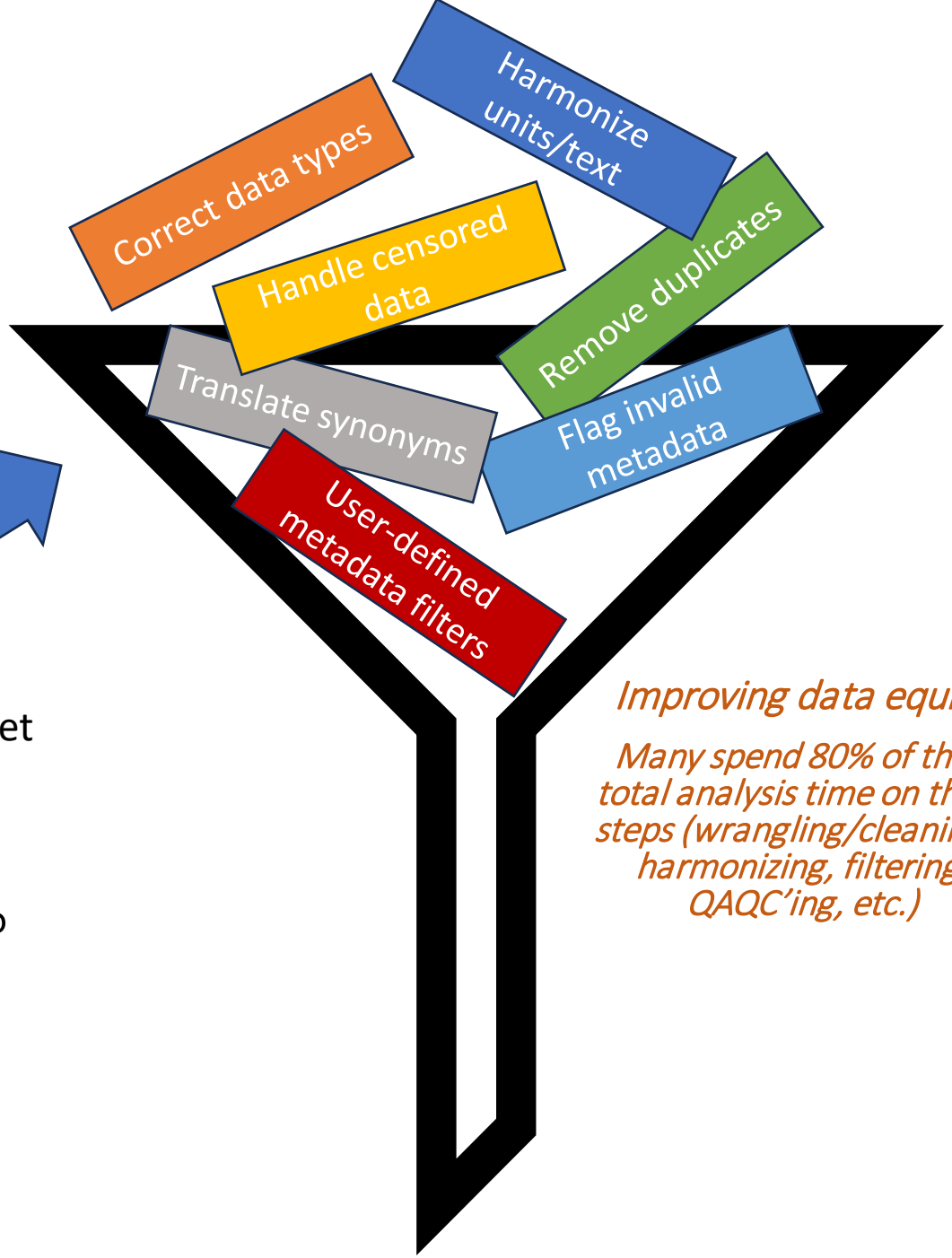
dataRetrieval USGS R



USGS's dataRetrieval functions



Constructing a unified dataset containing key metadata



Improving data equity

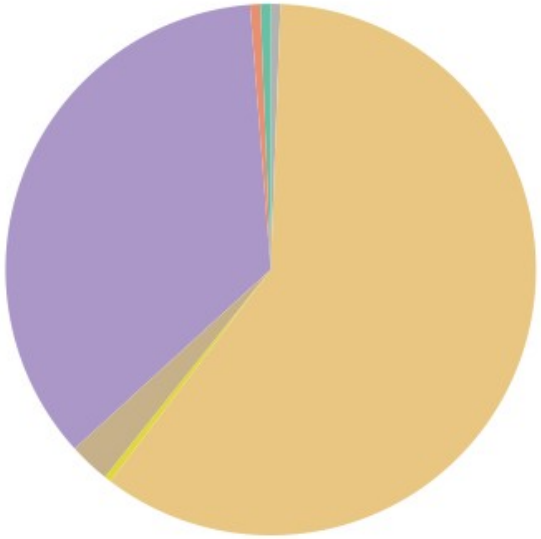
Many spend 80% of their total analysis time on these steps (wrangling/cleaning – harmonizing, filtering, QAQC'ing, etc.)

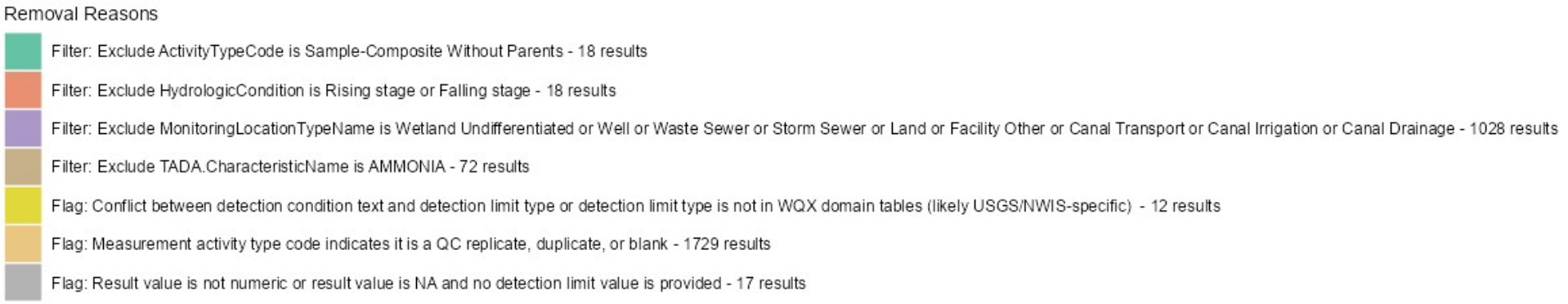
Are the data of sufficient quality for my analysis?

What harmonization and formatting steps are needed to use the data?

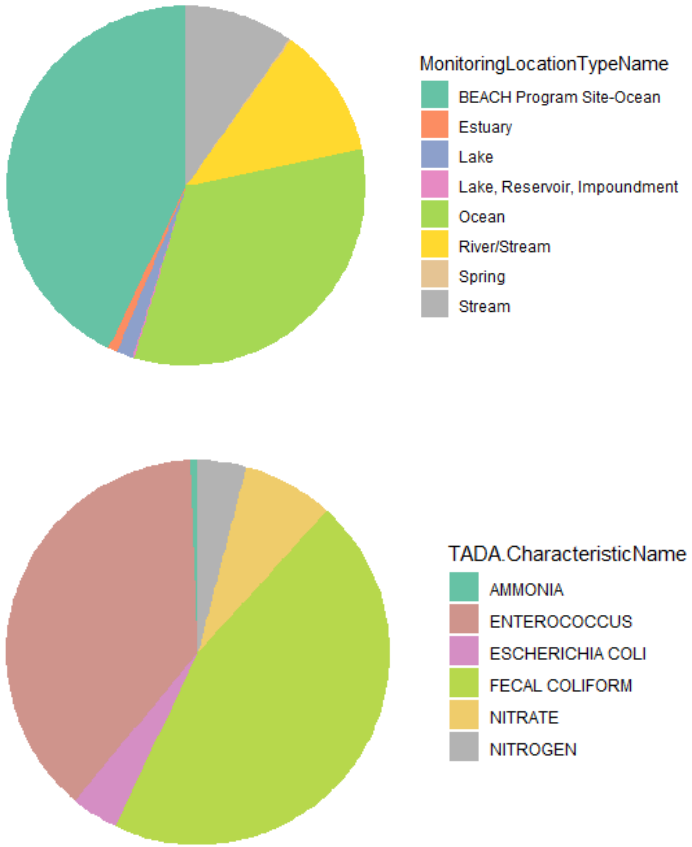
Summary of User Decisions

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

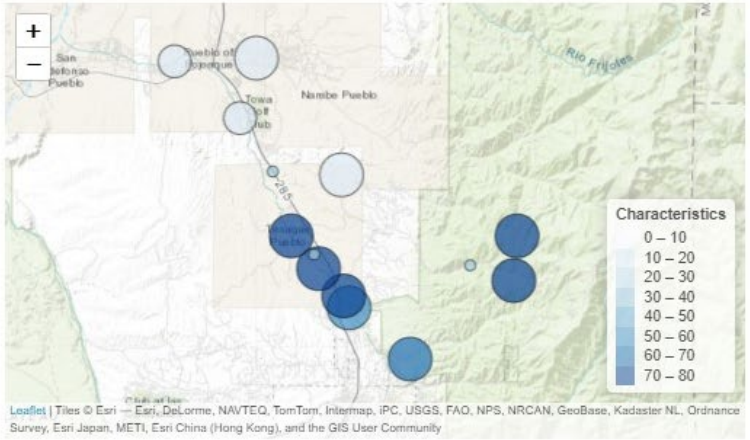




Data Visualization



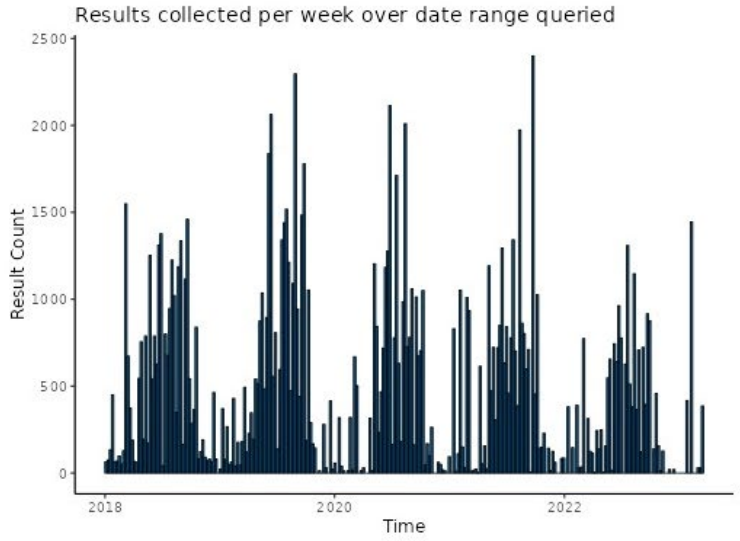
Your dataset contains **131,106** unique results from **221** monitoring location(s) and **6** unique organization(s).

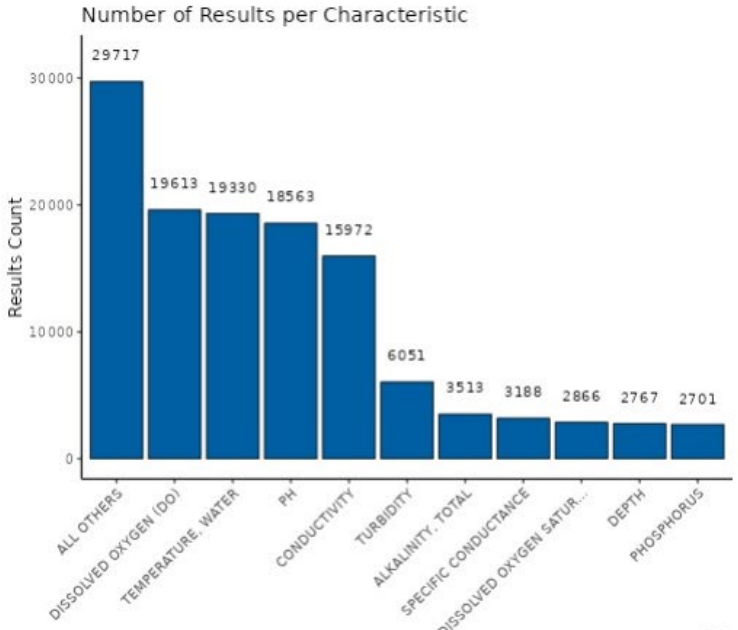


Show 10 entries

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



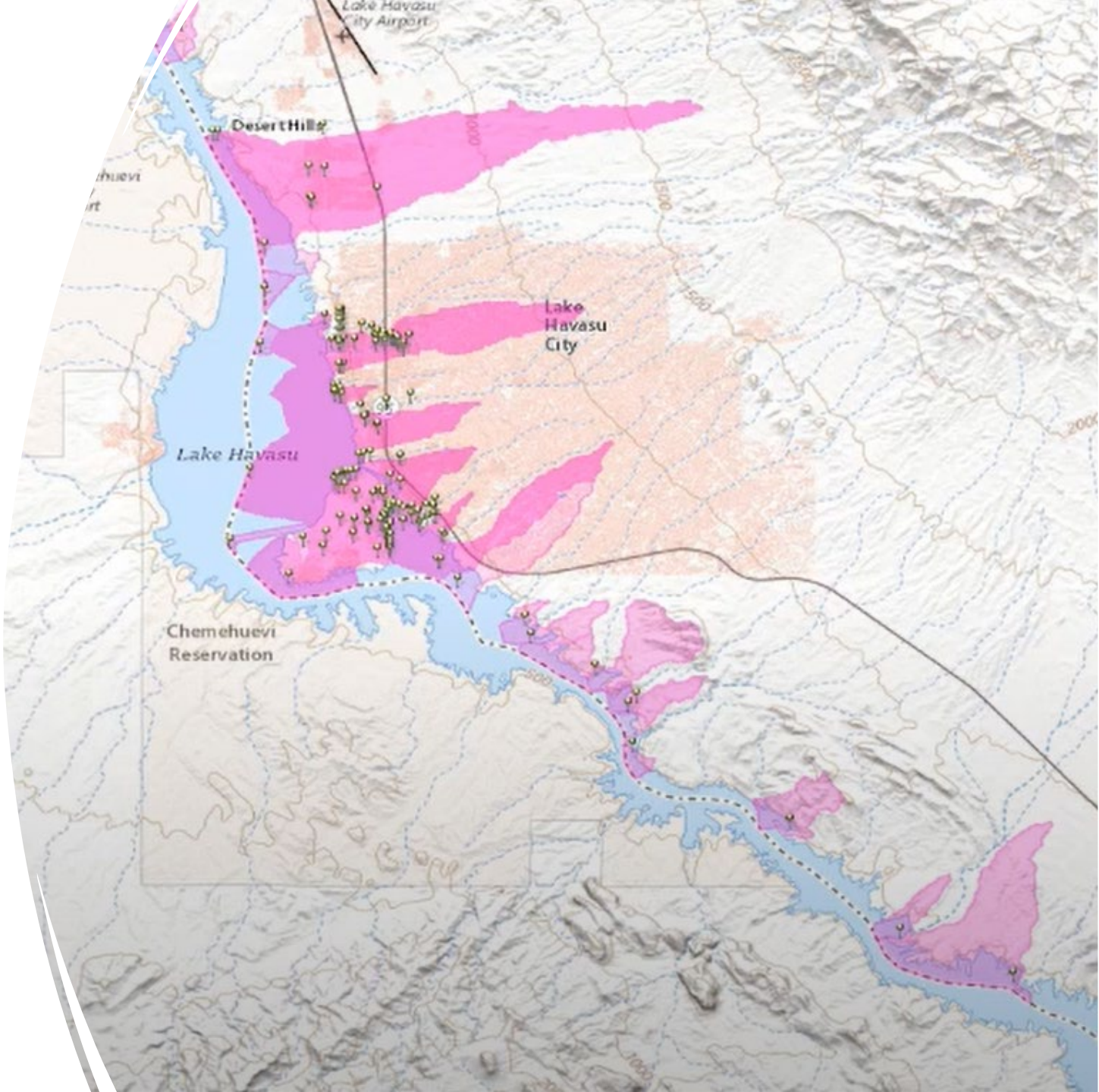


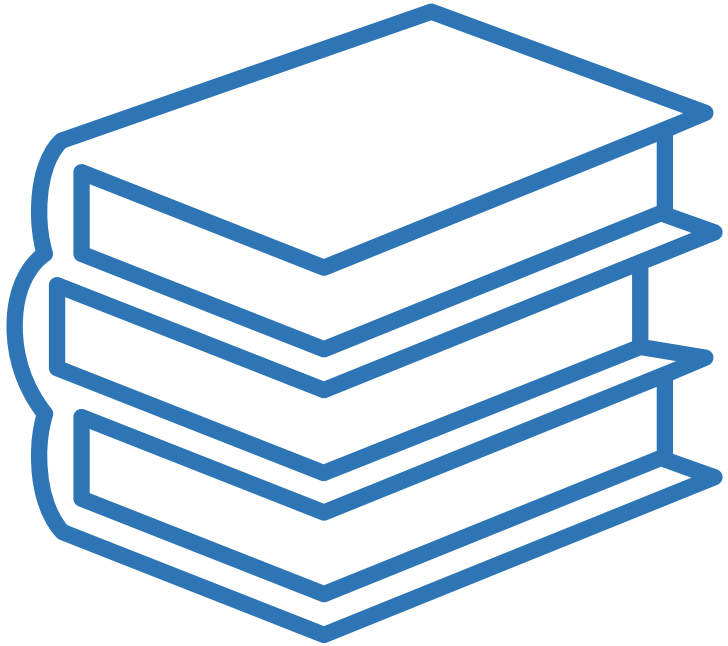
Module 2: Spatial Aggregation & Use Assignments

Vision

Associating Water Quality Criteria, Assessment Units & Uses with WQP Stations

- Integrate ATTAINS





Module 3: Assessment Criteria and Methodologies

Vision

TADA Working Group: Identifying requirements and 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

Integrate Criteria Search Tool

https://www.epa.gov/wqs-tech/state-specific-water-quality-standards-effective-under-clean-water-act-cwa

Under development

- Does not include narrative standards, duration and frequency, or methodologies
- This compilation is continuously updated as EPA approves new or revised WQS

Users can:

- Query by parameter, application or criteria magnitude value
  - Direct application = designated uses
  - Indirect application = specific waterbody, all other waters, or a class of waters of the parameter
- Query within a state or across all states
- Find National Recommended Water Quality Criteria (304A)
- More detailed webpages for each State, Territory, or Authorized Tribe
- Download as a "flat" spreadsheet file to support custom searches and analyses
- Find source document linking the criterion to EPA-approved state regulation

Module 4: Bringing it all together

Vision

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

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

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

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

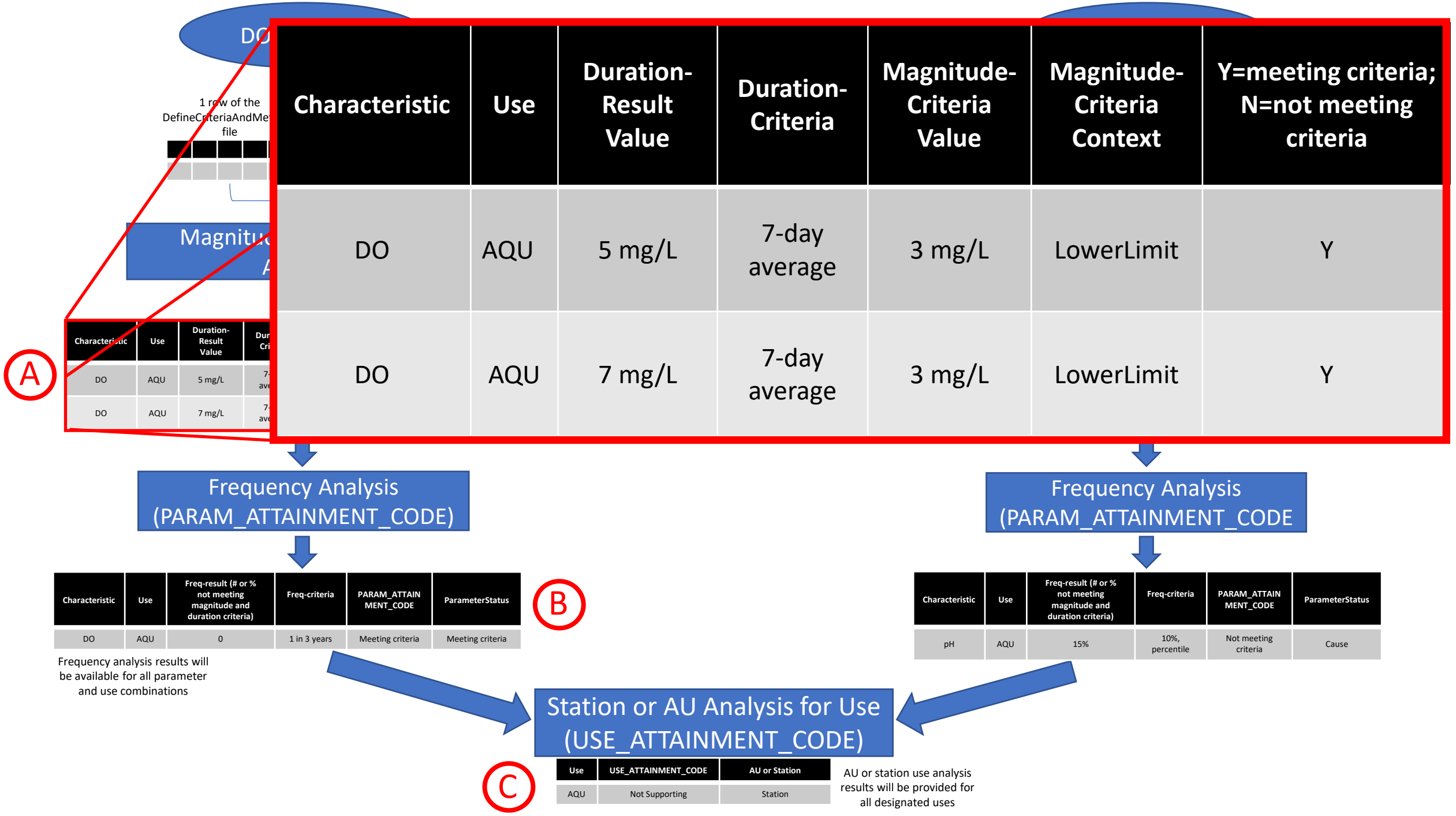
Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

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


DO
1 row of the DefineCriteriaAndMe file
Magnitu 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
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)
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
B
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

DO

1 row of the DefineCriteriaAndMe file

Magnitud A

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

A

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
DO	AQU	0	1 in 3 years	Meeting criteria	Meeting criteria

B

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

Frequency Analysis (PARAM_ATTAINMENT_CODE

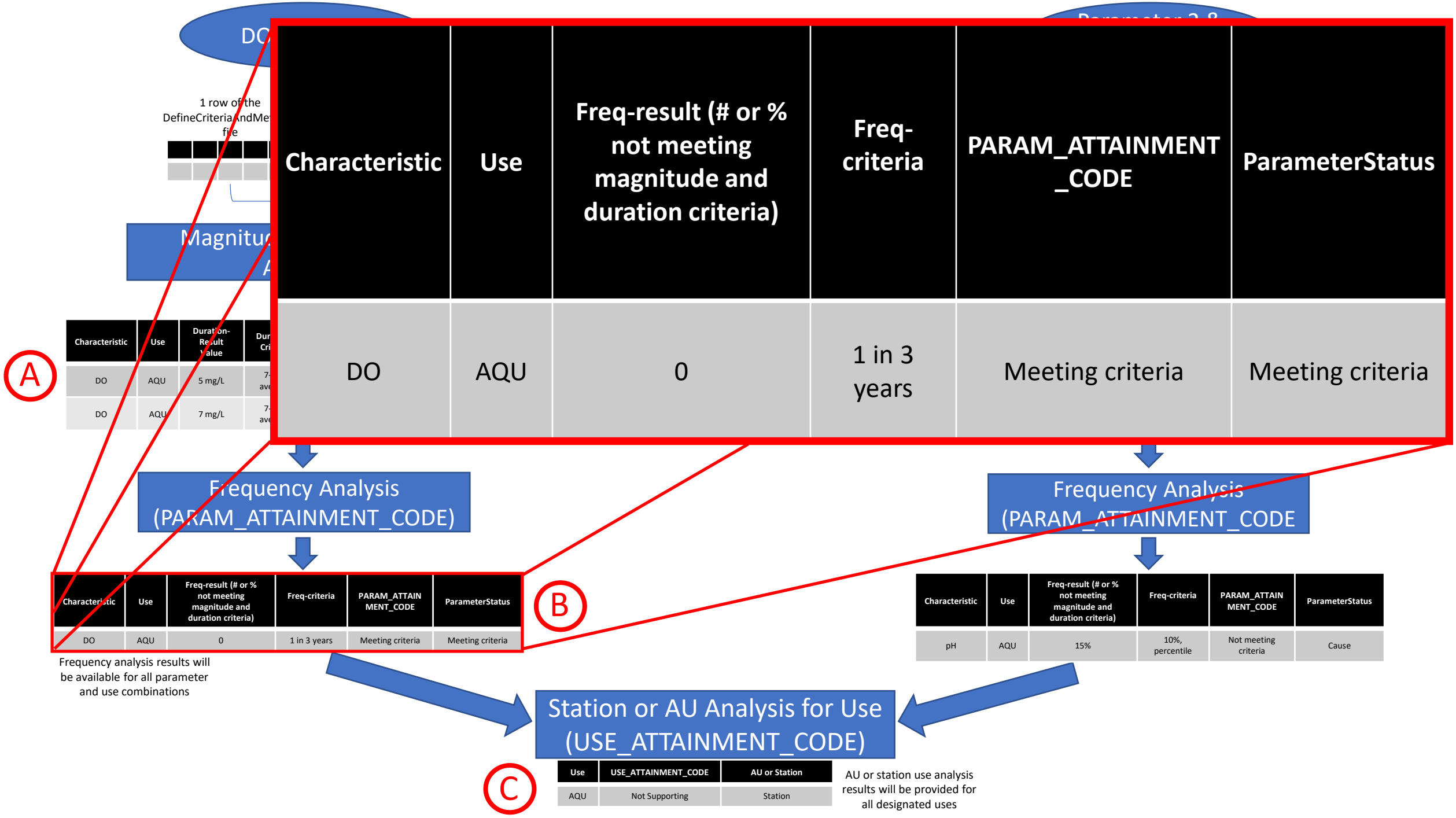
Characteristic	Use	Freq-result (# or % not meeting magnitude and duration criteria)	Freq-criteria	PARAM_ATTAINMENT_CODE	ParameterStatus
pH	AQU	15%	10%, percentile	Not meeting criteria	Cause

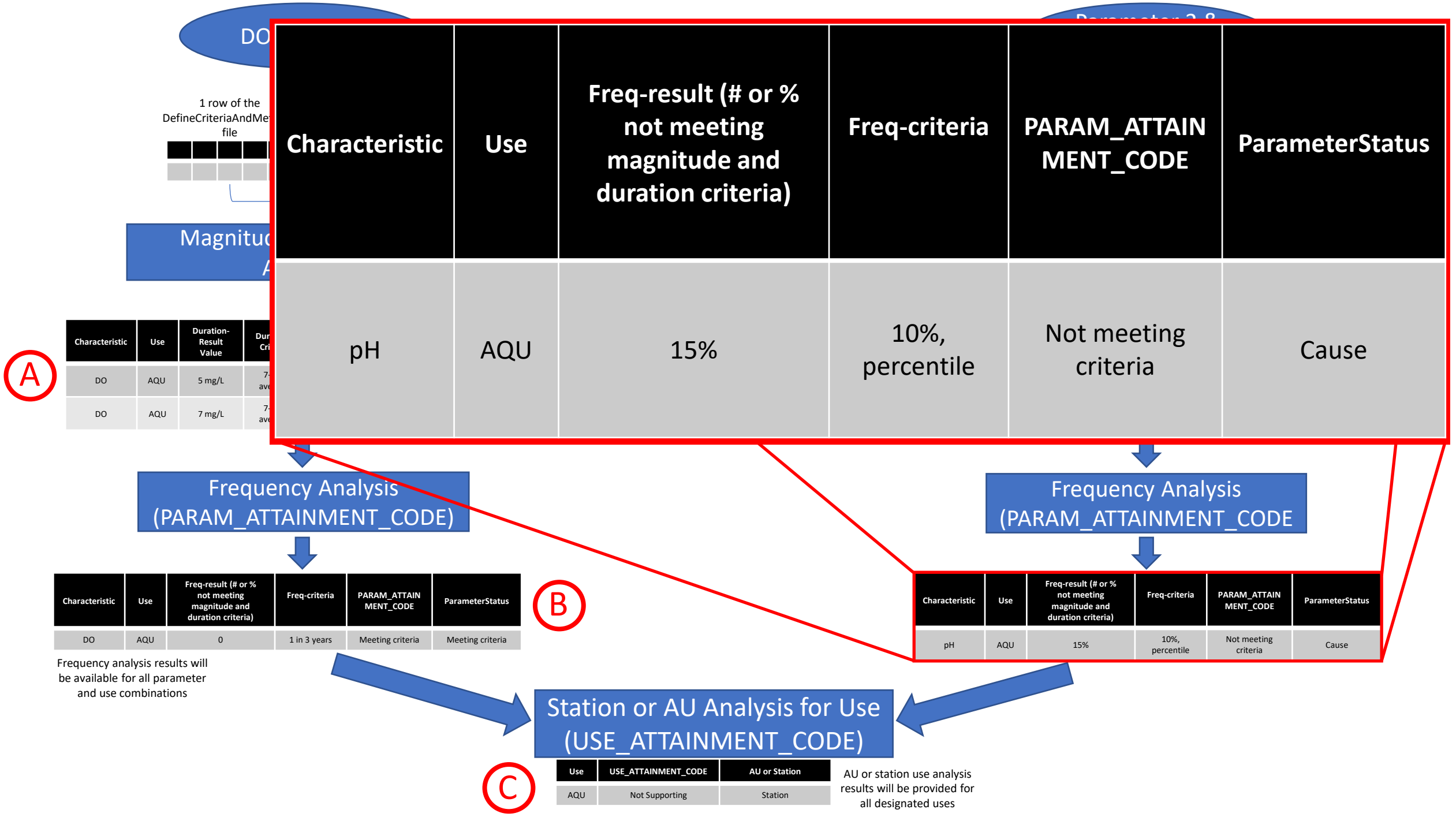
Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

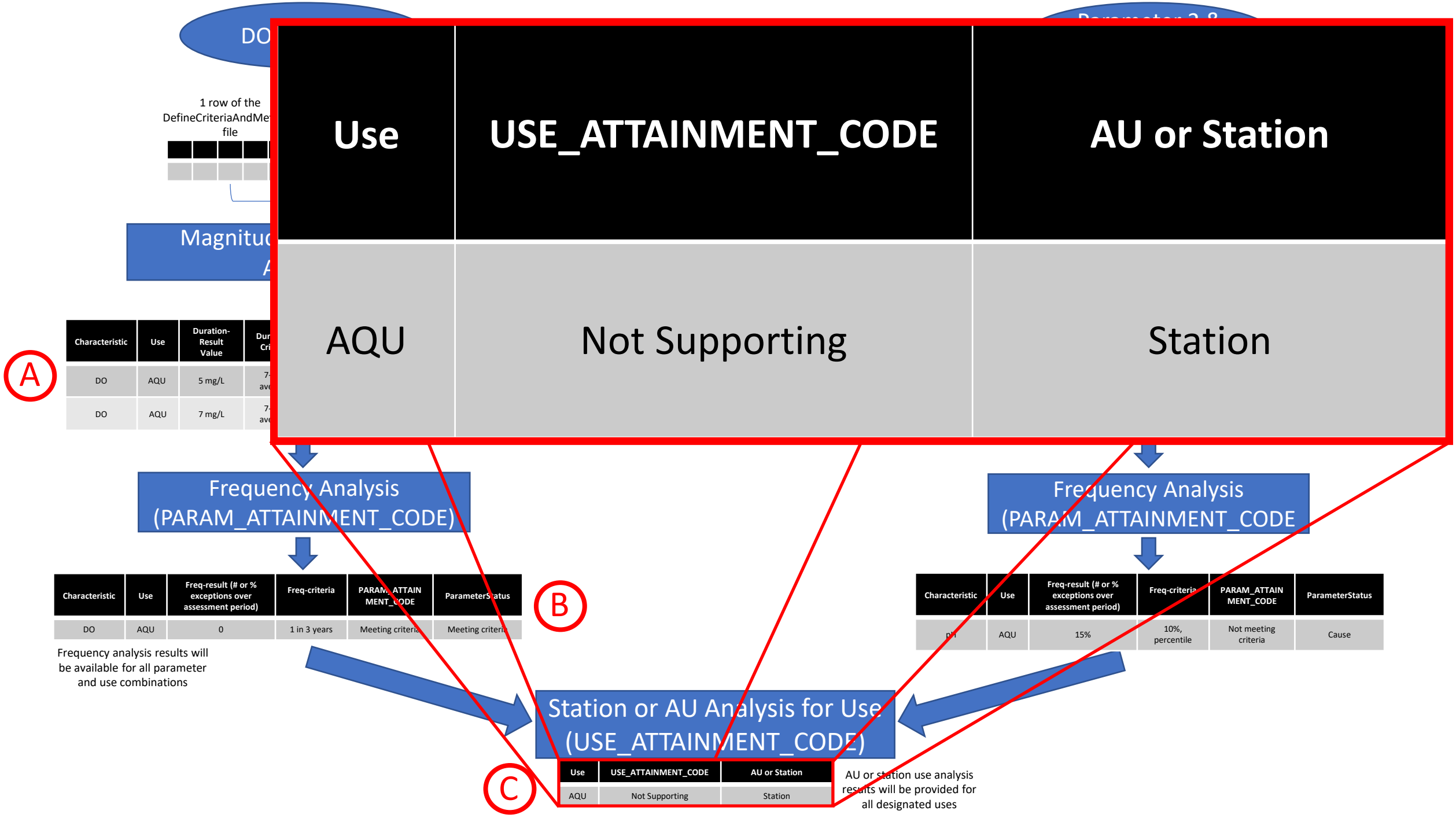
C

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 fine
Magnitu A
Characteristic
Use
Duration-Result Value
Dur Cri
DO
AQU
5 mg/L
DO
AQU
7 mg/L
A
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 (PARAM_ATTAINMENT_CODE)
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
B
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 results will be available for all parameter and use combinations
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
Magnitud
A
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
A
Characteristic
Use
Duration-Result Value
DO
AQU
5 mg/L
DO
AQU
7 mg/L
Frequency Analysis (PARAM_ATTAINMENT_CODE)
Frequency Analysis (PARAM_ATTAINMENT_CODE
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
B
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
Frequency analysis results will be available for all parameter and use combinations
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

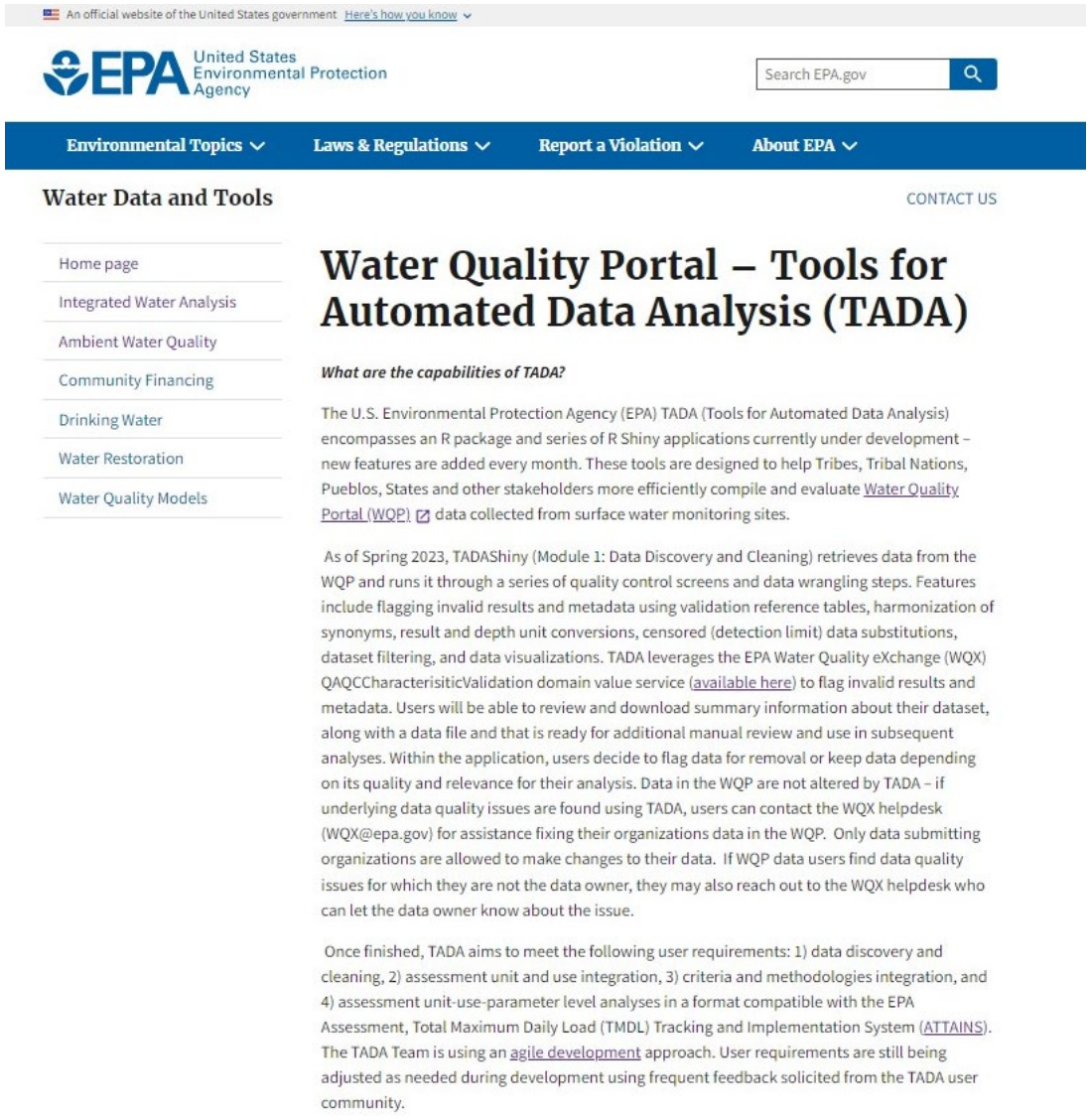

DO
1 row of the DefineCriteriaAndMe file
Magnitud
A
Use
USE_ATTAINMENT_CODE
AU or Station
AQU
Not Supporting
Station
A
Characteristic
Use
Duration-Result Value
DO
AQU
5 mg/L
DO
AQU
7 mg/L
Frequency Analysis (PARAM_ATTAINMENT_CODE)
Frequency Analysis (PARAM_ATTAINMENT_CODE
Characteristic
Use
Freq-result (# or % exceptions over assessment period)
Freq-criteria
PARAM_ATTAINMENT_CODE
ParameterStatus
DO
AQU
0
1 in 3 years
Meeting criteria
Meeting criteria
B
pH
AQU
15%
10%, percentile
Not meeting criteria
Cause
Frequency analysis results will be available for all parameter and use combinations
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

Module 1

- Gathered requirements, at refining stage
- Developed *most* functions, at refining stage
- Aiming to wrap up this fall/winter
- Potential barriers/slow downs
  - Testing
  - Staff turnover

Modules 2-4

- Gathered requirements, at refining stage
- Started proof of concept, not coding yet
- Potential barriers/slow downs
  - Formatting water quality standard information (related CST IoW project)
    - CST does not include narrative standards, duration and frequency, or methodologies
  - System crosswalk development and maintenance (parameters, designated uses)
  - Collaboration between OW and OST, regional coordinators, and states/territories/tribes



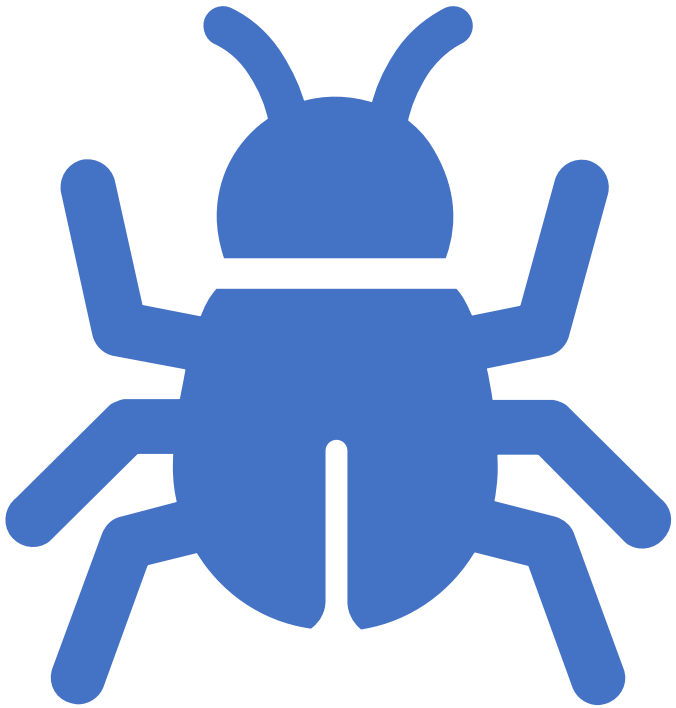
Current TADA Products

- Different tools for different users
  - R Package (coders)
  - R Shiny Application (non-coders)
- User Guides on GitHub Pages
- EPA TADA Website
- Inventory of Open Source and/or Publicly Available Tools that Use WQP Data/Services
- R and R Shiny Learning Resources for Water Community, Collaborative Effort Between TADA Working Group & North American Lake Management Society (NALMS)
- TADA Master List of Requirements (four modules)

Other tools that are compatible with WQX/WQP

- WQX laid the groundwork for EPA's How's My Waterway (HMW)
  - Would not be possible without WQX data model and standards that drive it all
- USGS's EGRET R package: Exploration and Graphics for RivEr Trends (EGRET): An R-package for the analysis of long-term changes in water quality and streamflow, including the water-quality method Weighted Regressions on Time, Discharge, and Season (WRTDS)
- Delaware WQP: Features of this site include a map interface of all water quality monitoring sites, graphs of water quality data, historical data files, and a water quality conditions search tool
- U.S. EPA Freshwater Explorer – specific conductivity
- Utah's irTools and wqTools R packages: A set of useful functions that are commonly used and re-used in water quality analyses (ex. converting units, loading calculations, and downloading data). These packages live on the Utah Division of Water Quality's GitHub.
- https://nalms.shinyapps.io/Shiny_for_Water_Resources/
- ContDataSumViz, for summarizing and visualizing QC'd continuous sensor data

..and more! Please reach out if you are interested in learning about others!



For Today's Training

- TADA Package Vignette: https://usepa.github.io/TADA/articles/TADAModule1.html
- TADA Shiny App: https://github.com/USEPA/TADAShiny
- Feedback Form: https://forms.gle/BKeb9PxgcYgNK9mb9

Thank you for listening!

Email:

mullin.cristina@epa.gov

LinkedIn:
https://www.linkedin.com/in/cristina-mullin-95953099/