TADA (Tools for Automated Data Analysis)

Cristina Mullin, PhD
Water Data Integration Branch
EPA Office of Water

TADA Team
Cristina Mullin, Team Lead
Shelly Thawley
Elise Hinman, PhD (ORISE Fellow)
Trip Hook (Skills Marketplace, EPA OLEM)
Katie Healy (Skills Marketplace, EPA ARD)
TADA Working Group (~100 volunteers)
Agenda

• Presentation (~25 min)
• Hands-on/Q&A (~65 min)
EPA Water Data Integration Branch (WDIB)
Products Supporting the Clean Water Act

Internet of Water Principles
(Data Standards and Interoperability)
• Water Quality Data (WQX/WQP)
• Hydrography (NHDPlus)
• Assessment Decisions (ATTAINS)

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

Monitor Waters
106, 319

WQX/WQP
(Share Water Data)
106 grant req.

How’s My Waterway
(Share Water Conditions)

ATTAINS
(Report Water Conditions)
106/305(b)/303(d)

Findable
Accessible
Interoperable
Broader Impacts

Centralized tools (like TADA and HMW) have potential to greatly reduce total government costs

- State and tribal agencies
- EPA regions
- EPA HQ
- USGS, other federal agencies
- Applicable even beyond water

Improving data equity

- TADA makes the WQX QAQC service available on the WQP side
- Organizations can use TADA to help find and address data quality issues and then fix them in WQX/WQP

Well received by enthusiastic, engaged user community (open source/collaborative effort from start = more use product)

More efficient assessments may lead to more time for other things, or more waters being assessed! (Or anything else of importance to entity!)
Developing **TADA** as an **R package** and series of user-friendly web apps (R Shiny)

- Open-source
- Efficient, reproducible workflows
- Community driven requirements
  - Accessible
  - Standardized but flexible
  - Highly customizable
- Designed to work with Water Quality Portal data
  - Companion to USGS’s **dataRetrieval** package

Tools vary in...
- **Scope**
- **Application**
- **Stage of development**
“Serve as a hub for an open-source water quality community”

**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.
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)
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
Module 1: Data Discovery and Wrangling

In Progress
Finding readily available data

Water Quality Portal (WQP)

Tools for Automated Data Analysis (TADA) Module 1: Water Quality Portal Data Discovery and Cleaning

Option A: Use example data

Option B: Query the Water Quality Portal (WQP)

Use the fields below to download a dataset directly from WQP. Fields with "x" in the label allow multiple selections. Hydrologic Units may be at any scale, from subwatershed to region. However, be mindful that large queries may time out.

State
Monitoring Location ID(s)
Characteristic Group
Site Type(s)
County (pick state first)
Organization(s)
Characteristic(s)
Start Date
End Date
Hydrologic Unit
Project(s)
Sample Media

Option C: Upload dataset

Select a file from your computer. This upload feature currently only accepts data in xls and xlsx formats. The file can be a fresh TMDA dataset or a working TMDA dataset that you are returning to the app to continue. Additional enhancements and validation might be required. You may reach out to the WQP Helpdesk at WQP@epa.gov for assistance preparing and submitting your data to the WQP through EPA's WQP.

Search EPA.gov
Are the data of sufficient quality for my analysis?

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

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

Constructing a unified dataset containing key metadata
Summary of User Decisions

<table>
<thead>
<tr>
<th>TADA.Remove TADA.RemovalReason</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
</tr>
<tr>
<td>TRUE</td>
</tr>
<tr>
<td>TRUE</td>
</tr>
<tr>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
</tr>
<tr>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
</tr>
<tr>
<td>TRUE</td>
</tr>
</tbody>
</table>

Removal Reasons
- Filter: Exclude ActivityTypeCode is Sample-Composite Without Parents - 18 results
- Filter: Exclude HydrologicCondition is Rising stage or Falling stage - 18 results
- 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 - 1028 results
- Filter: Exclude TADA.CharacteristicName is AMMONIA - 72 results
- Flag: Conflict between detection condition text and detection limit type or detection limit type is not in WQX domain tables (likely USGS/NWS-specific) - 12 results
- Flag: Measurement activity type code indicates it is a QC replicate, duplicate, or blank - 1729 results
- Flag: Result value is not numeric or result value is NA and no detection limit value is provided - 17 results
Data Visualization

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

<table>
<thead>
<tr>
<th>Organization/EntityName</th>
<th>Result_Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chickasaw Nation Environmental Service</td>
<td>4580</td>
</tr>
<tr>
<td>Fond du Lac Band of Chippawa (MN)</td>
<td>20178</td>
</tr>
<tr>
<td>Pueblo Of Tesuque</td>
<td>6706</td>
</tr>
<tr>
<td>Pueblo of Pojoaque</td>
<td>1181</td>
</tr>
<tr>
<td>Red Lake DNR</td>
<td>81734</td>
</tr>
<tr>
<td>Sac and Fox Nation (Tribal)</td>
<td>3819</td>
</tr>
</tbody>
</table>

Showing 1 to 6 of 6 entries

Results collected per week over date range queried

Number of Results per Characteristic

Data

Visualization
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
**Identified Requirements/Priorities**

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

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

**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


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
Conceptual Example

Magnitude and Duration Analysis

**Characteristic**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Duration Result Value</th>
<th>Duration-Criteria Result</th>
<th>Magnitude Criteria</th>
<th>Yr-meeting criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO AQU</td>
<td>5 mg/L</td>
<td>7-day average</td>
<td>3 mg/L</td>
<td>LowerLimit</td>
<td>Y</td>
</tr>
<tr>
<td>DO AQU</td>
<td>7 mg/L</td>
<td>7-day average</td>
<td>3 mg/L</td>
<td>LowerLimit</td>
<td>Y</td>
</tr>
</tbody>
</table>

Frequency Analysis (PARAM_ATTAINMENT_CODE)

**Characteristic**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Param_result (# or % not meeting magnitude and duration criteria)</th>
<th>Frequency criteria</th>
<th>PARAM_ATTAINMENT_CODE</th>
<th>ParameterStatus</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO AQU</td>
<td>0</td>
<td>1 in 3 years</td>
<td>Meeting criteria</td>
<td>Meeting criteria</td>
<td></td>
</tr>
</tbody>
</table>

Magnitude and Duration Analysis

**Characteristic**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Duration Result Value</th>
<th>Duration-Criteria Result</th>
<th>Magnitude Criteria</th>
<th>Magnitude-Criteria Result</th>
<th>Yr-meeting criteria; N=not meeting criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH AQU</td>
<td>3</td>
<td>30-day Min:</td>
<td>6-8</td>
<td>Range</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>pH AQU</td>
<td>5</td>
<td>30-day Min:</td>
<td>6-8</td>
<td>Range</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

Frequency Analysis (PARAM_ATTAINMENT_CODE)

**Characteristic**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Param_result (# or % not meeting magnitude and duration criteria)</th>
<th>Frequency criteria</th>
<th>PARAM_ATTAINMENT_CODE</th>
<th>ParameterStatus</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH AQU</td>
<td>15%</td>
<td>10% percentile</td>
<td>Not meeting criteria</td>
<td>Cause</td>
<td></td>
</tr>
</tbody>
</table>

Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

<table>
<thead>
<tr>
<th>Use</th>
<th>USE_ATTAINMENT_CODE</th>
<th>AU or Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQU</td>
<td>Not Supporting</td>
<td>Station</td>
</tr>
</tbody>
</table>

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

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

Frequency analysis results will be provided for all designated uses.
## Frequency Analysis (PARAM_ATTAINMENT_CODE)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Duration - Result Value</th>
<th>Duration - Criteria</th>
<th>Magnitude - Criteria Value</th>
<th>Magnitude - Criteria Context</th>
<th>Y = meeting criteria; N = not meeting criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO AQU</td>
<td></td>
<td>5 mg/L</td>
<td>7-day average</td>
<td>3 mg/L</td>
<td>LowerLimit</td>
<td>Y</td>
</tr>
<tr>
<td>DO AQU</td>
<td></td>
<td>7 mg/L</td>
<td>7-day average</td>
<td>3 mg/L</td>
<td>LowerLimit</td>
<td>Y</td>
</tr>
</tbody>
</table>

**Station or AU Analysis for Use (USE_ATTAINMENT_CODE)**

<table>
<thead>
<tr>
<th>Use</th>
<th>USE_ATTAINMENT_CODE</th>
<th>AU or Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQU</td>
<td>Not Supporting</td>
<td>Station</td>
</tr>
</tbody>
</table>

**Frequency analysis results** will be available for all parameter and use combinations.
### Frequency Analysis (PARAM_ATTAINMENT_CODE)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Freq-result (# or % not meeting magnitude and duration criteria)</th>
<th>Param_Status</th>
<th>Parameters/Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO AQU</td>
<td>0</td>
<td>1 in 3 years</td>
<td>Meeting criteria</td>
<td>Meeting criteria</td>
</tr>
</tbody>
</table>

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

### Frequency Analysis (PARAM_ATTAINMENT_CODE)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Freq-result (# or % not meeting magnitude and duration criteria)</th>
<th>Parameters/Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH AQU</td>
<td>15%</td>
<td>10% percentile</td>
<td>Not meeting criteria</td>
</tr>
</tbody>
</table>

### Station or AU Analysis for Use (USE_ATTAINMENT_CODE)

<table>
<thead>
<tr>
<th>Use</th>
<th>USE_ATTAINMENT_CODE</th>
<th>AU or Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQU</td>
<td>Not Supporting</td>
<td>Station</td>
</tr>
</tbody>
</table>
Frequency analysis results will be available for all parameter and use combinations.

Station or AU Analysis for Use

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

Characteristics of the data frame:

- **Magnitude and Duration Analysis**
  - **DO and AQU**
  - **Criteria**
  - **Value**
  - **Context**
    - **Y** = not meeting criteria; **N** = meeting criteria

**Table: Frequency Analysis**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Frequency Result (Freq)</th>
<th>Magnitude Criteria</th>
<th>Duration Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DO AQU</strong></td>
<td>0</td>
<td>1 in 3 years</td>
<td>Meeting criteria</td>
<td>Meeting criteria</td>
</tr>
</tbody>
</table>

**Table: Parameter Status**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Frequency Result (Freq)</th>
<th>Magnitude Criteria</th>
<th>Duration Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pH AQU</strong></td>
<td>15%</td>
<td>10% percentile</td>
<td>Not meeting criteria</td>
<td>Cause</td>
</tr>
</tbody>
</table>

**Note:**
- **Freq-result (# or % not meeting magnitude and duration criteria)**
- **Freq-criteria**
- **PARAM_ATTAINMENT_CODE**
- **Parameter Status**

**Output:**
- **Assessment Determinations**
- **Parameter Status**

**Assessment:**
- Frequency analysis results will be available for all parameter and use combinations.
### Frequency Analysis

**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

**DO** | **AQU** | 5 mg/L | 7 day average | 3 mg/L | LowerLimit

**DO** | **AQU** | 7 mg/L | 7 day average | 3 mg/L | LowerLimit

---

**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

**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)*

**Use** | **USE_ATTAINMENT_CODE** | **AU or Station**
--- | --- | ---

**AQU** | **Not Supporting** | **Station**

---

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

Station or AU analysis results will be provided for all designated uses
### Magnitude and Duration Analysis

#### Use  
**USE_ATTAINMENT_CODE**  
**AU or Station**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Result (mg/L)</th>
<th>Duration (day average)</th>
<th>Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO</td>
<td>AQU</td>
<td>5</td>
<td>7</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>DO</td>
<td>AQU</td>
<td>7</td>
<td>7</td>
<td>N</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Frequency Analysis

**PARAM_ATTAINMENT_CODE**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Frequency (result)</th>
<th>Criteria (percentile)</th>
<th>Value (cause)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO</td>
<td>AQU</td>
<td>15%</td>
<td>10%percentile</td>
<td>Not meeting</td>
</tr>
</tbody>
</table>

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

**Station or AU Analysis for Use**

<table>
<thead>
<tr>
<th>Use</th>
<th>USE_ATTAINMENT_CODE</th>
<th>AU or Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQU</td>
<td>Not Supporting</td>
<td>Station</td>
</tr>
</tbody>
</table>

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

### Frequency Analysis

**PARAM_ATTAINMENT_CODE**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Use</th>
<th>Frequency (result)</th>
<th>Criteria (percentile)</th>
<th>Value (cause)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>AQU</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>AQU</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Use**

<table>
<thead>
<tr>
<th>Use</th>
<th>USE_ATTAINMENT_CODE</th>
<th>AU or Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQU</td>
<td>Not Supporting</td>
<td>Station</td>
</tr>
</tbody>
</table>

**AU or Station**

**Frequency Analysis**

**PARAM_ATTAINMENT_CODE**
Wrap Up: TADA Vision

QC'd data
Wrap Up: TADA Vision

- QC'd data
- Assessment unit overlay with monitoring locations
- Assign beneficial uses
Wrap Up: TADA Vision

- QC'd 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
Wrap Up: TADA Vision

QC'd 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
Wrap Up: TADA Vision

Power happens in the spreadsheets

Visualization/Review, Exploration, Customization
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
- TADA Master List of Requirements (four modules)
TADA Project Timeline

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

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

<table>
<thead>
<tr>
<th>September 2020</th>
<th>January 2021</th>
<th>August 2021</th>
<th>June 2023</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assemble project team, begin TADA conceptualization and management</td>
<td>TADA Working Group start date (two tracks – 2 meeting per month)</td>
<td>TADA Working Group (single track – growing quickly)</td>
<td>Host regular meetings with open-source water quality data analysis community to deploy tool</td>
</tr>
</tbody>
</table>

https://commons.wikimedia.org/wiki/File:Waterfallvsagile.jpg
For Today’s Training

- TADA Package Vignette: https://usepa.github.io/TADA/articles/TADAModule1.html
- TADA Shiny App: https://github.com/USEPA/TADAShiny
- Bug Form: https://forms.gle/PoTCXEeVAaTnEgLs6
- Feedback Form: https://forms.gle/MAcieQUCqsqywMLY7
Thank you for attending!

- Contact Cristina: mullin.Cristina@epa.gov
- We want your feedback! We are always looking for users to test our products and features
  - Bug/Error Report
  - Feedback Form
  - Submit Issue on GitHub
    - TADA R Package
    - TADA Shiny
EXTRA SLIDES BEYOND
Water Quality Portal – Tools for Automated Data Analysis (TADA)

What are the capabilities of TADA?

The U.S. Environmental Protection Agency (EPA) TADA (Tools for Automated Data Analysis) encompasses an R package and series of R Shiny applications currently under development – new features are added every month. These tools are designed to help Tribes, Tribal Nations, Pueblos, States and other stakeholders more efficiently compile and evaluate Water Quality Portal (WQP) data collected from surface water monitoring sites.

As of Spring 2023, TADASHiny (Module 1: Data Discovery and Cleaning) retrieves data from the WQP and runs it through a series of quality control screens and data wrangling steps. Features include flagging invalid results and metadata using validation reference tables, harmonization of synonyms, result and depth unit conversions, censored (detection limit) data substitutions, dataset filtering, and data visualizations. TADA leverages the EPA Water Quality eXchange (WQX) QualityCharacteristicValidation domain value service (available here) to flag invalid results and metadata. Users will be able to review and download summary information about their dataset, along with a data file and that is ready for additional manual review and use in subsequent analyses. Within the application, users decide to flag data for removal or keep data depending on its quality and relevance for their analysis. Data in the WQP are not altered by TADA – if underlying data quality issues are found using TADA, users can contact the WQX helpdesk (WQX@epa.gov) for assistance fixing their organizations data in the WQP. Only data submitting organizations are allowed to make changes to their data. If WQP data users find data quality issues for which they are not the data owner, they may also reach out to the WQX helpdesk who
This R package can be used to compile and evaluate Water Quality Portal (WQP) data for samples collected from surface water monitoring sites on streams and lakes. It can be used to create applications that support water quality programs and help states, tribes, and other stakeholders efficiently analyze the data.

- [usepa.github.io/TADA/](usepa.github.io/TADA/)
Welcome to TADA: Tools for Automated Data Analysis!

We encourage you to read this package's CONTRIBUTING, LICENSE, and README files (you are here).

Tools for Automated Data Analysis, or TADA, is a draft R package being developed to help States, Tribes, Tribal Nations, Pueblos, and other stakeholders more efficiently compile and evaluate Water Quality Portal (WQP) data collected from surface water 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 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! More information on how to contribute can be found in the CONTRIBUTING file. This file explains how users can contribute to the R package by submitting a pull request or issue to request a change or provide feedback. We hope to build a collaborative community dedicated to this effort where contributors can discover, share and build the package functionality over time.

More about the TADA Project

Installation

You can install and load the most recent version of the TADA R Package from GitHub by running:

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

Water Quality Portal
Contributing
2023-06-08
Source: vignettes/CONTRIBUTING.Rmd

Contribute to TADA!
We encourage you to read this project's CONTRIBUTING policy (you are here), its LICENSE, and its README.

We're so glad you're thinking about contributing to an EPA open source project! If you're unsure about anything, just ask — or submit your issue or pull request anyway. The worst that can happen is we'll politely ask you to change something. We appreciate all friendly contributions.

No matter who you are, if you spot an error, omission, or bug, you're welcome to open an issue in this repo!

TADA Working Group Mission
To share and develop R code for evaluating and visualizing Water Quality Portal (WQP) data more efficiently through 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.

Package Development
This article will walk through how to contribute to the TADA package via a pull request workflow. This is also not a complete guide to R package development (a comprehensive guide is R Packages), instead this is meant as more of a checklist for the general steps. Several references are included at the bottom for more information on R package development and git workflows.

What is GitHub?
GitHub is a third-party website that offers version-controlled repositories that developers and teams can use to host their code. It's a great way to collaborate with others on code, track changes, and keep a history of who made what changes. For more information, see the GitHub website.
Function reference

All functions

```
AboveNationalsQXUpperThreshold()
    Check Result Value Against WQX Upper Threshold

AggregatedContinuousData()
    Check for Aggregated Continuous Data

BelowNationalsQXLowerThreshold()
    Check Result Value Against WQX Lower Threshold

GetActivityTypeRef()
    Update Activity Type Reference Table

GetDetCondRef()
    Update Result Detection Condition Reference Table

GetDetLimitRef()
```
Reviewed by the TADA team. Once approved, updates are then merged into the develop branch. However, you are welcome to download any branch you’d like using the `ref` input in `install_github` (see code chunk above). This functionality is mainly only useful to TADA package developers/contributors.

The following code block ensures the additional packages needed to run the code in this RMarkdown document are loaded. However, users may also use the `package name:: package function` notation to avoid the list of `library()` calls.

```r
# Load tidyverse
if(!"tidyverse"%in%installed.packages()){
  install.packages("tidyverse")
}
library(tidyverse)
```

### Help pages

All TADA R package functions have their own individual help pages, listed on the Function reference page on the GitHub site. Users can also access the help page for a given function in R or RStudio using the following format (example below): `?TADA::[name of TADA function]`

```r
?TADA::TADAdataRetrieval
```
Install and setup

Users can install the TADA package from GitHub into their R library using the `remotes` package. Copy and paste the code below into your R or RStudio console to download and install.

TADA package relies on other packages, therefore you may be prompted in the console to update dependency packages that have more recent versions available. If you see this prompt, it is recommended to update all of them (enter 1 into the console).

```r
# Install TADA
if(!"remotes"%in%installed.packages()){
  install.packages("remotes")
}
remotes::install_github("USEPA/TADA", ref="develop")
library(TADA)
```

It’s that easy! The most stable branch for TADA right now is the develop branch. Contributors generally create their own branches based on develop, make some improvements, and then submit a pull request to be reviewed by the TADA Team. Once approved, updates are then merged into the develop branch. However, you are welcome to download any branch you’d like using the `ref` input in `install_github` (see code chunk above). This functionality is mainly only useful to TADA package developers/contributors.