TADA

🇺🇸 An official website of the United States government Here's how you know

United States
Environmental Protection
Agency

Search EPA.gov

Environmental Topics   **Laws & Regulations**   **Report a Violation**   **About EPA**

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

| 1. Load | 2. Overview | 3. Flag | 4. Filter | 5. Censored Data | 6. Harmonize and Calculate | 7. Explore | 8. Review |

Option A: Use example data

Use example data

[▼]

📥 Load

Option B: Query the Water Quality Portal (WQP)

Use the fields below to download a dataset directly from WQP. Fields with '(s)' 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.

Option B: Query the Water Quality Portal (WQP)

Use the fields below to download a dataset directly from WQP. Fields with '(s)' 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.

Date Range

Start Date

2024-09-04

End Date

2024-09-04

Location Information

State

Select state

County (pick state first)

Select county

Hydrologic Unit

e.g. 020700100103

Monitoring Location ID(s)

Start typing or use drop down menu

Metadata Filters

Organization(s)

Project(s)

Site Type(s)

Sample Media ⓘ

Water   water

Characteristic Group

Characteristic(s)

☁ Run Query

Option C: Upload dataset

Upload a dataset from your computer. This upload feature only accepts data in .xls and .xlsx formats. The file can be a **fresh** TADA dataset or a **working** TADA dataset that you are returning to the app to iterate on. Data must also be formatted in the EPA Water Quality eXchange (WQX) schema to leverage this tool. You may reach out to the WQX helpdesk at WQX@epa.gov for assistance preparing and submitting your data to the WQP through EPA's WQX.

Browse...	No file selected

Optional: Upload Progress File

Upload a progress file from your computer. This upload feature only accepts data in the .RData format. The TADA Shiny application keeps track of all user selections, and makes a .RData file available for download at any time. If you saved a progress file you generated during a previous use of the TADA Shiny application, then it can be uploaded here and used to automatically parameterize the TADA Shiny app with the same selections. This file can be used to regenerate a dataset with the same decisions as before, or can be used to apply the same user selctions to a new dataset

Browse...	No file selected

TADA Working Summary

Total Results in Dataset: 0

Total Results Flagged for Removal: 0

Total Results Retained: 0

TADA Working Summary

Total Results in Dataset: 0

Total Results Flagged for Removal: 0

Total Results Retained: 0

Total Sites in Dataset: 0

Total Sites Flagged for Removal: 0

Total Sites Retained: 0

⬇ Download Working Dataset (.zip)

⬇ Download Final Dataset (.zip)

DISCLAIMER

🇺🇸 An official website of the United States government   Here's how you know

![EPA Logo] United States
Environmental Protection
Agency

Search EPA.gov

Environmental Topics   **Laws & Regulations**   **Report a Violation**   **About EPA**

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

| 1. Load | 2. Overview | 3. Flag | 4. Filter | 5. Censored Data | 6. Harmonize and Calculate | 7. Explore | 8. Review |

Option A: Use example data

Use example data

Tribal (132k results)   ▲

Nutrients Utah (15k results)
Shepherdstown (34k results)
Tribal (132k results)

Option B: Query the Water Quality Portal (WQP)

Use the fields below to download a dataset directly from WQP. Fields with '(s)' 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.
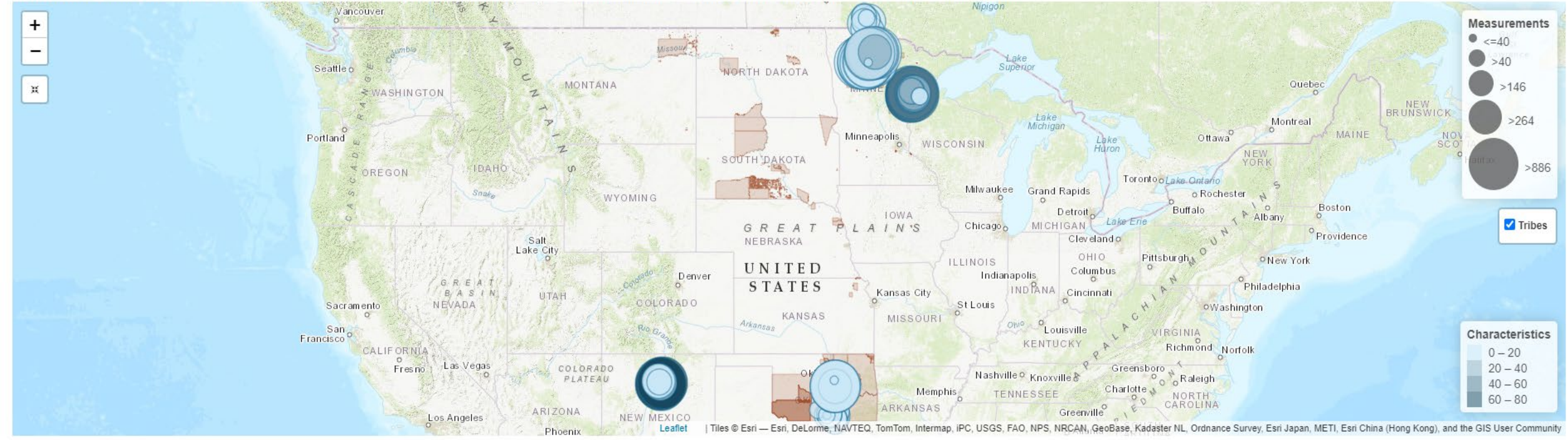
TADAShiny

owshiny-dev.app.cloud.gov/tada-dev/

Suggested Sites | TimeReporting | Workplace Region... | 6WQ-PT Home | ATTAINS Login | Arkansas | Louisiana | New Mexico | Oklahoma | Texas | Tribes | EPA | Travel | Sharepoint Files | WDIB | Academy | R6RB | ARMS | All Bookmarks

Data Overview

Note: This page shows maps and figures using the **original** dataset uploaded to this TADAShiny session. If you'd like to see updated figures after working in other tabs, please press the 'Refresh' button.

⟳ Refresh

Your dataset contains **134,050** unique results from **227** monitoring location(s) and **6** unique organization(s).

Your dataset, mapped: Zoom in and click on sites of interest. A pop up will appear that shows the number of measurements, characteristics, and visits at each site.
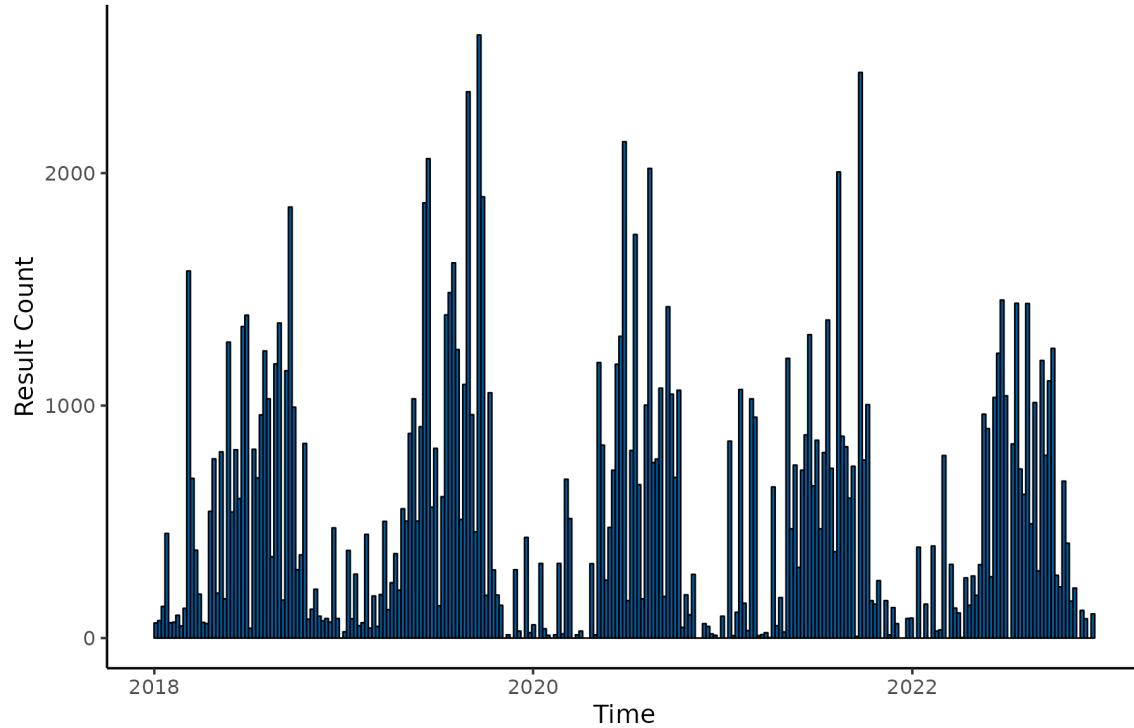

Measurements
- <=40
- >40
- >146
- >264
- >886

☑ Tribes

Characteristics
- 0 – 20
- 20 – 40
- 40 – 60
- 60 – 80

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

Results collected per week over date range queried

Results per Characteristic

DEPTH ▮ 2812

TADAShiny

owshiny-dev.app.cloud.gov/tada-dev/

Suggested Sites · TimeReporting · R6 EPA Workplace Region... · 6WQ-PT Home · ATTAINS Login · Arkansas · Louisiana · New Mexico · Oklahoma · Texas · Tribes · EPA · Travel · All Bookmarks

Results collected per week over date range queried

Results per Characteristic

Characteristic	Results Count
DEPTH	2812
WATER APPEARANCE (TEXT...	2918
SPECIFIC CONDUCTANCE	3241
DISSOLVED OXYGEN SATUR...	3618
ALKALINITY, TOTAL	3630
TURBIDITY	6160
CONDUCTIVITY	15613
PH	18241
DISSOLVED OXYGEN (DO)	18598
TEMPERATURE, WATER	19015
ALL OTHERS	40204

Organizations in dataset

The table below shows the organizations that collected data in your dataset and the number of measurements collected by each. Notice the third column, 'Rank'. This editable column is present because sometimes organizations unintentionally upload the same dataset multiple times to the WQP. For example, USGS will collect data at the request of state agencies. The USGS 'copy' results is uploaded to NWIS and made available in the portal, and the state agency's 'copy' of the results is uploaded to WQX. This rank provides the necessary info needed to flag and select representative result from groups of duplicative uploads based on date, characteristic and result value/unit, and proximity to other sites. Double click in a cell in the 'Rank' column to edit the hierarchy

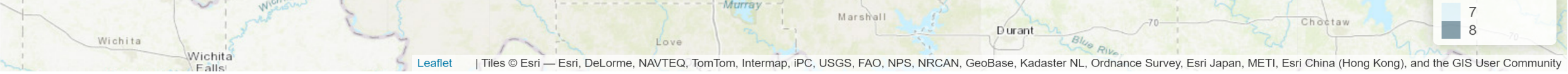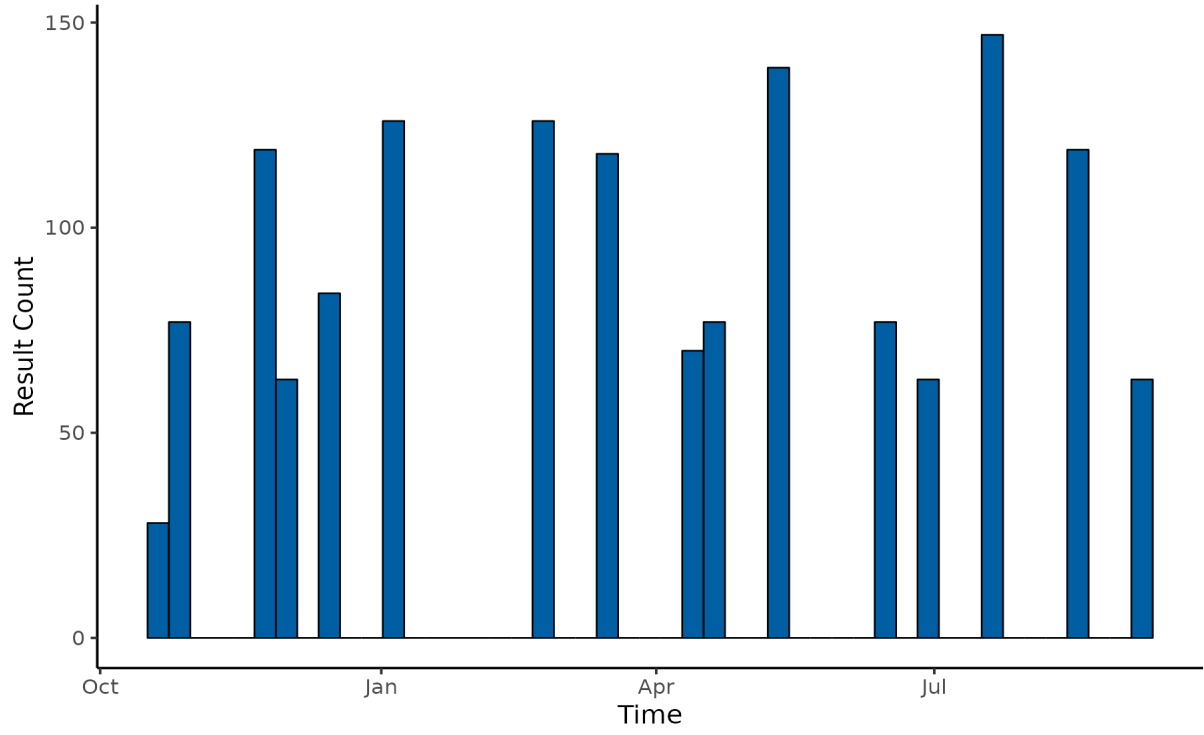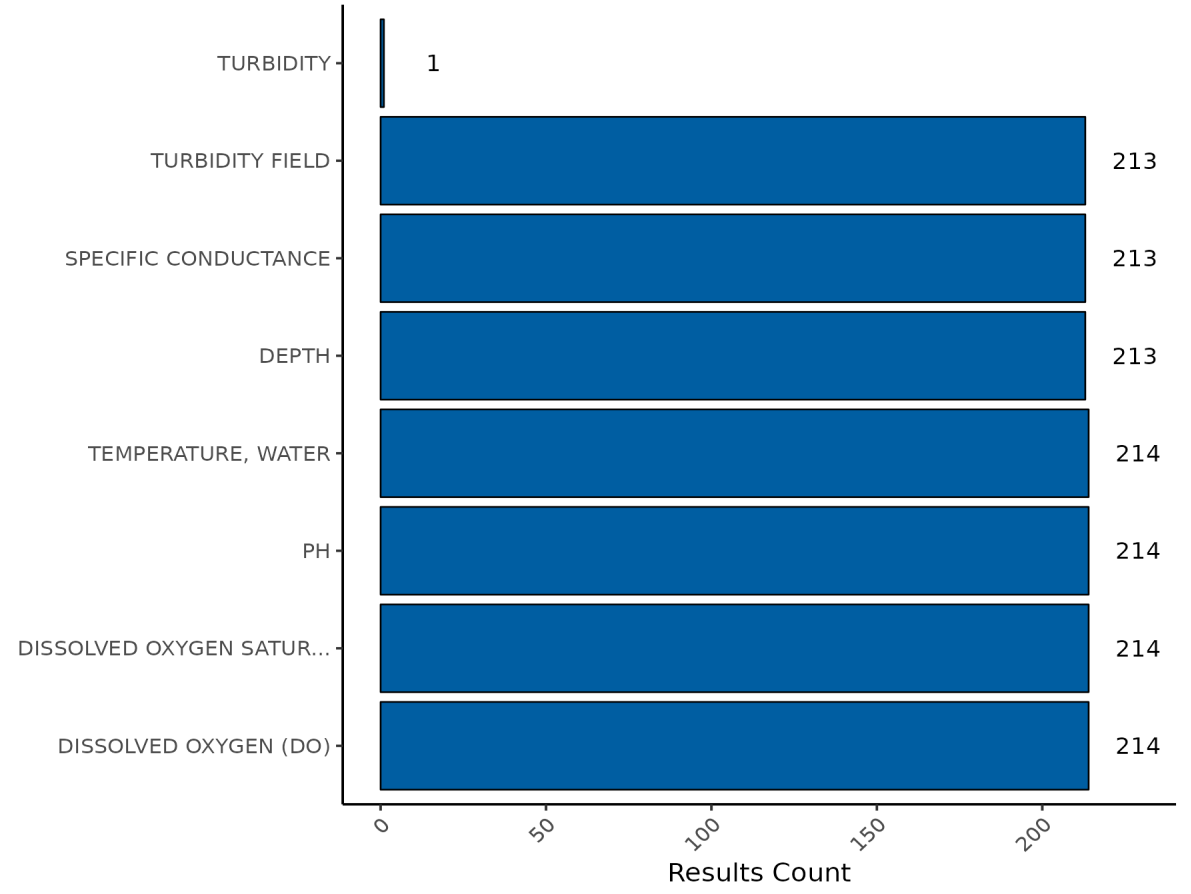## Results collected per week over date range queried

Results per Characteristic

Organizations in dataset

The table below shows the organizations that collected data in your dataset and the number of measurements collected by each. Notice the third column, 'Rank'. This editable column is present

Organizations in dataset

The table below shows the organizations that collected data in your dataset and the number of measurements collected by each. Notice the third column, 'Rank'. This editable column is present because sometimes organizations unintentionally upload the same dataset multiple times to the WQP. For example, USGS will collect data at the request of state agencies. The USGS 'copy' of the results is uploaded to NWIS and made available in the portal, and the state agency's 'copy' of the results is uploaded to WQX. This rank provides the necessary info needed to flag and select one representative result from groups of duplicative uploads based on date, characteristic and result value/unit, and proximity to other sites. Double click in a cell in the 'Rank' column to edit the hierarchy of organizations and Ctrl-Enter to save those changes in the table: the default ranks organizations by the number of measurements in the dataset. Using the state vs USGS data example, if the state agency's organization name has a higher rank (ex. ranked #1) than USGS (ex. ranked #2), its result will be selected over the USGS upload of the sample, and the USGS version will be flagged for removal.

Show 6 entries

Organization Name	Results Count	Rank - Double Click to Edit, Ctrl-Enter to Save
Red Lake DNR	87187	1
Fond du Lac Band of Chippewa (MN)	23972	2
Sac and Fox Nation (Tribal)	9943	3
Pueblo Of Tesuque (Tribal)	6901	4
Chickasaw Nation Environmental Service (Tribal)	4946	5
Pueblo of Pojoaque	1101	6

Showing 1 to 6 of 6 entries

Previous 1 Next

Showing 1 to 6 of 6 entries

Previous 1 Next

TADA Working Summary

Total Results in Dataset: 134,050

Total Results Flagged for Removal: 2,575

Total Results Retained: 131,475

Total Sites in Dataset: 227

Total Sites Flagged for Removal: 0

Total Sites Retained: 227

⬇ Download Working Dataset (.zip)

⬇ Download Final Dataset (.zip)

DISCLAIMER

Portal Data Discovery and Cleaning

| 1. Load | 2. Overview | 3. Flag | 4. Filter | 5. Censored Data | 6. Harmonize and Calculate | 7. Explore | 8. Review |

Flag data for potential issues

Click the button below to run a series of tests that check for quality control issues or data formats not compatible with TADA. When the tests are finished running, a table will appear below. Each row describes an evaluation test, reports the number of results affected, and contains a switch users may toggle on/off to decide whether to flag results for removal. However, evaluation tests marked as **Required** have permanently 'ON' light blue switches that cannot be changed. **Recommended** tests are automatically switched 'ON' (darker blue), and **Optional** tests are automatically switched 'OFF' (gray).

Run Tests

Convert depth units (Optional)

Depth units in the dataset are automatically converted to **meters** upon data retrieval. Click the radio buttons below to convert depth units to feet, inches, or back to meters.

○ feet   ○ inches   ● meters

TADA Working Summary

Total Results in Dataset: 134,050

Total Results Flagged for Removal: 0

Total Results Retained: 134,050

owshiny-dev.app.cloud.gov/tada-dev/

Portal Data Discovery and Cleaning

| 1. Load | 2. Overview | 3. Flag | 4. Filter | 5. Censored Data | 6. Harmonize and Calculate | 7. Explore | 8. Review |

Flag data for potential issues

Click the button below to run a series of tests that check for quality control issues or data formats not compatible with TADA. When the tests are finished running, a table will appear below. Each row describes an evaluation test, reports the number of results affected, and contains a switch users may toggle on/off to decide whether to flag results for removal. However, evaluation tests marked as **Required** have permanently 'ON' light blue switches that cannot be changed. **Recommended** tests are automatically switched 'ON' (darker blue), and **Optional** tests are automatically switched 'OFF' (gray).

Run Tests

Flag reason	Results affected	Required/Optional	Switch 'on' to flag for removal
Result value is not numeric or NA and no detection limit value is provided (see TADA.ResultMeasureValueDataTypes.Flag)	3821	Required	n/a
Result value is not numeric or NA and the detection limit value cannot be interpreted because there is a conflict between the detection condition text and detection limit type provided by the data submitter [or the detection limit type is not in WQX domain tables (USGS/NWIS-specific)] (see TADA.CensoredData.Flag)	115	Required	n/a
Suspect characteristic and speciation combination (see TADA.MethodSpeciation.Flag)	0	Recommended	⬤ (on)
Suspect characteristic and fraction combination (see TADA.SampleFraction.Flag)	304	Recommended	⬤ (on)
Suspect characteristic and unit combination (see TADA.ResultUnit.Flag)	796	Recommended	⬤ (on)

Suspect characteristic and unit combination (see TADA.ResultUnit.Flag)	796	Recommended	🔵 On
ActivityTypeCode indicates result value is a QC replicate, duplicate, or blank (see TADA.ActivityType.Flag)	1421	Recommended	🔵 On
MeasureQualifierCode indicates result value is suspect (see TADA.MeasureQualifierCode.Flag)	55	Recommended	🔵 On
Metadata indicates duplicative uploads of the same results within a single organization (see TADA.SingleOrgDup.Flag)	1755	Optional	🔵 On
Metadata indicates duplicative uploads of the same results by multiple organizations (see TADA.ResultSelectedMultipleOrgs)	0	Optional	⚪ Off
Uncommon or suspect characteristic and analytical method combination (see TADA.AnalyticalMethod.Flag)	1019	Optional	⚪ Off
Quality Assurance Project Plan (QAPPApprovedIndicator) indicates "N" (not approved) or NA	95562	Optional	⚪ Off
Quality Assurance Project Plan (QAPP) document is not available via the ProjectFileUrl (see TADA.QAPPDocAvailable)	134050	Optional	⚪ Off
Result value(s) above the national threshold for a given characteristic, possibly indicating non-sensical value(s) (see TADA.ResultValueAboveUpperThreshold.Flag)	6597	Optional	⚪ Off
Result value(s) below the national threshold for a given characteristic, possibly indicating non-sensical value(s) (see TADA.ResultValueBelowLowerThreshold.Flag)	110	Optional	⚪ Off
Coordinates are outside of the United States (see TADA.InvalidCoordinates.Flag)	0	Optional	⚪ Off
Coordinates are imprecise (less than three decimal digits) (see TADA.InvalidCoordinates.Flag)	2286	Optional	⚪ Off

| Coordinates are imprecise (less than three decimal digits) (see TADA.InvalidCoordinates.Flag) | 2286 | Optional | |

Convert depth units (Optional)

Depth units in the dataset are automatically converted to **meters** upon data retrieval. Click the radio buttons below to convert depth units to feet, inches, or back to meters.

○ feet ○ inches ● meters

TADA Working Summary

Total Results in Dataset: 134,050

Total Results Flagged for Removal: 4,274

Total Results Retained: 129,776

Total Sites in Dataset: 227

Total Sites Flagged for Removal: 0

Total Sites Retained: 227

⬇ Download Working Dataset (.zip)

⬇ Download Final Dataset (.zip)

DISCLAIMER

Select field to filter on:

Fields are listed in the table below, along with the number of unique values present in that field. These counts do not include unique values from results flagged for removal. Click on a field name and a new table will appear below showing the counts associated with each unique value in the selected field.

Fields to select from:

🔘 key   ⚪ most   ⚪ all

Fields	Count
ActivityCommentText	805
ResultCommentText	362
SubjectTaxonomicName	278
TADA.ComparableDataIdentifier	189
TADA.CharacteristicName	141
ResultAnalyticalMethod.MethodName	70
ProjectName	30
ProjectDescriptionText	18
SampleCollectionMethod.MethodName	15

ResultCommentText	362
SubjectTaxonomicName	278
TADA.ComparableDataIdentifier	189
TADA.CharacteristicName	141
ResultAnalyticalMethod.MethodName	70
ProjectName	30
ProjectDescriptionText	18
SampleCollectionMethod.MethodName	15
LaboratoryName	9
DetectionQuantitationLimitTypeName	7
TADA.ResultSampleFractionText	7
OrganizationFormalName	6
MonitoringLocationTypeName	5
TADA.MeasureQualifierCode.Def	5
StateCode	4
ActivityTypeCode	3
TADA.ActivityMediaName	3
ActivityMediaSubdivisionName	

ActivityMediaSubdivisionName | 3

Filter by 'OrganizationFormalName'

In this table, you may either exclude selected values, or ONLY include selected values and exclude all other non-selected values. Use the buttons at the bottom of this table to make your decisions. Note that once you select a filtering type (Exclude or Include), the other filtering type button is disabled for that field. **Note:** If any results are NA, they will be represented by a blank (empty) row in this table.

Value	Count
Chickasaw Nation Environmental Service (Tribal)	4730
Fond du Lac Band of Chippewa (MN)	23128
Pueblo of Pojoaque	1101
Pueblo Of Tesuque (Tribal)	6791
Red Lake DNR	85789
Sac and Fox Nation (Tribal)	9936

Include Only Selected Values   **Exclude Selected Values**

Selected filters

Your exclude/include filters are documented below. You can easily reset one or more filter decisions by clicking on rows of interest and then hitting the 'Remove Selected Filters' button. Alternat... you can reset/remove all filters on this page by clicking the 'Reset All Filters' button.
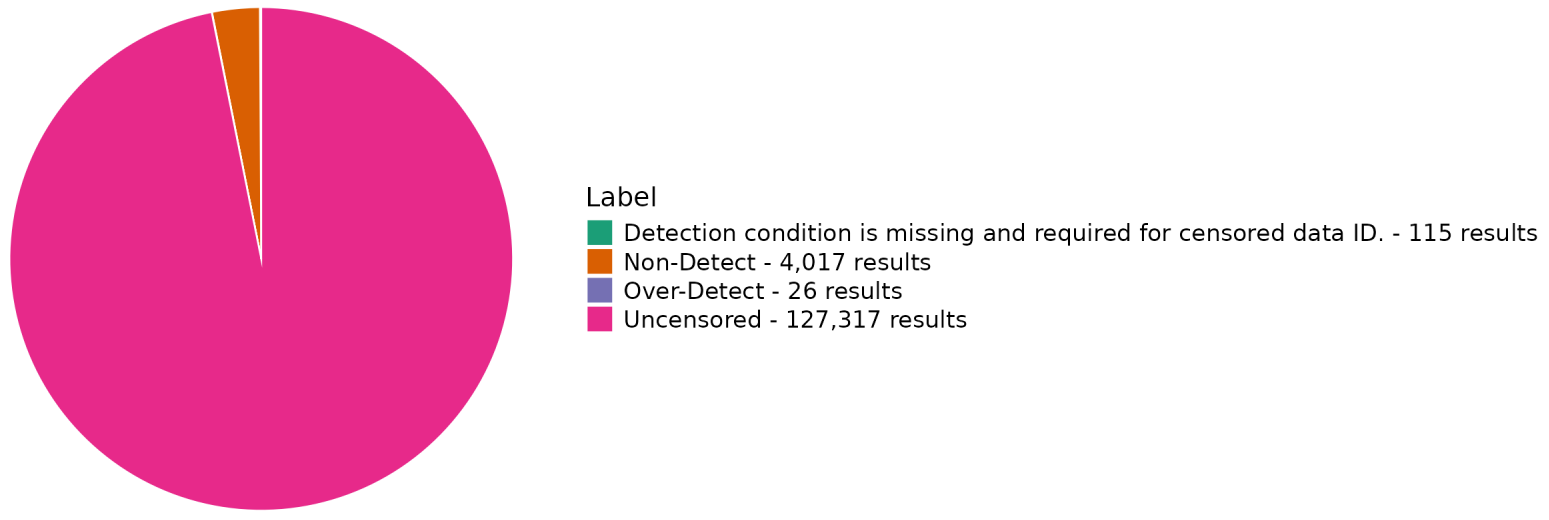
1. Load   2. Overview   3. Flag   4. Filter   **5. Censored Data**   6. Harmonize and Calculate   7. Explore   8. Review

Censored Data Categories

TADAdataRetrieval assigns each result in your dataset to non-detect, over-detect, other, or uncensored. The pie chart below displays the relative proportions of results in each category. Please note that detection limit data with conflicts or data quality issues are not displayed in this pie chart or handled in the methods below.

Number of Results per Censored Data Category

Label
- Detection condition is missing and required for censored data ID. - 115 results
- Non-Detect - 4,017 results
- Over-Detect - 26 results
- Uncensored - 127,317 results

Handle Censored Data Using Simple Methods

Use the drop down menus below to pick a simple method for handling non-detects and over-detects in the dataset. When you press 'Apply Methods to Dataset', a table will appear below with the 10 detection limit results, showing their initial values and estimated values.

Handle Censored Data Using Simple Methods

Use the drop down menus below to pick a simple method for handling non-detects and over-detects in the dataset. When you press 'Apply Methods to Dataset', a table will appear below with the first 10 detection limit results, showing their initial values and estimated values.

Non-Detect Handling Method

Multiply detection limit by x ▼

Multiplier (x)

0.5

Over-Detect Handling Method

No change ▼

Apply Methods to Dataset

Consider More Complex Censored Data Handling Methods

Use the picker list below to select grouping columns to create summary table. The summary table shows the number of non- and over-detects in each group, the total number of results in each group, the number of detection limit types (censoring levels) and the percentage of the dataset that is censored. These numbers are then used to suggest a potential statistical censored data method to use. Currently, the user must perform more complex analyses outside of TADAShiny.

Select Grouping Columns for Summarization

TADA.ComparableDataIdentifier

ID and Summarize Censored Data

Handle Censored Data Using Simple Methods

Use the drop down menus below to pick a simple method for handling non-detects and over-detects in the dataset. When you press 'Apply Methods to Dataset', a table will appear below with the first 10 detection limit results, showing their initial values and estimated values.

Non-Detect Handling Method

Multiply detection limit by x ▾

Multiplier (x)

0.5

Over-Detect Handling Method

Multiply detection limit by x ▾

Multiplier (x)

0.5

[Apply Methods to Dataset]   [Undo Method Application]

ResultIdentifier	TADA.CharacteristicName	Original Detection Limit Value	Original Unit	Estimated Detection Limit Value	Estimated Unit
STORET-1023801466	ORTHOPHOSPHATE	0.003	mg/L	1.5	UG/L
STORET-1023801467	TOTAL PHOSPHORUS, MIXED FORMS	0.2	mg/L	100	UG/L
STORET-1023801468	NITRATE + NITRITE	0.02	mg/L	0.01	MG/L
STORET-1023801470	AMMONIA	0.1	mg/L	0.05	MG/L
STORET-1023801477	ORTHOPHOSPHATE	0.003	mg/L	1.5	UG/L
STORET-1023801478	TOTAL PHOSPHORUS, MIXED FORMS	0.2	mg/L	100	UG/L
STORET-1023801479	NITRATE + NITRITE	0.02	mg/L	0.01	MG/L
STORET-1023801481	ZINC	5	ug/L	2.5	UG/L
STORET-1023801483	NICKEL	0.5	ug/L	0.25	UG/L
STORET-1023801485	COPPER	1	ug/L	0.5	UG/L

Select Grouping Columns for Summarization

TADA.ComparableDataIdentifier

[ID and Summarize Censored Data]

CSV Show [10] entries

TADA.ComparableDataIdentifier	Location_Count	Measurement_Count	Non_Detect_Count	Non_Detect_Pct	Non_Detect_Lvls	Over_Detect_Count	Over_Detec
PH_NA_NA_NONE	206	18233	0	0	0	0	
TEMPERATURE, WATER_NA_NA_DEG C	206	19008	0	0	0	0	
DISSOLVED OXYGEN (DO)_NA_NA_MG/L	200	18598	0	0	0	0	
TURBIDITY_NA_NA_NTU	170	5748	0	0	0	0	
PHOSPHORUS_TOTAL_NA_UG/L	167	2456	17	0.6921824104234527	2	0	
CONDUCTIVITY_NA_NA_US/CM	113	15390	0	0	0	0	
ALKALINITY, TOTAL_NA_NA_UG/L	107	2958	0	0	0	0	
CHLORIDE_TOTAL_NA_UG/L	104	762	10	1.312233958005249	3	0	
DEPTH, SECCHI DISK DEPTH_NA_NA_M	88	1619	0	0	0	0	
DISSOLVED OXYGEN							