Water Quality Portal Data Tools

Britt L. Dean

Jillian Adair

Shelly Thawley

Kiki Schneider

US EPA- OWOW-WRAPD

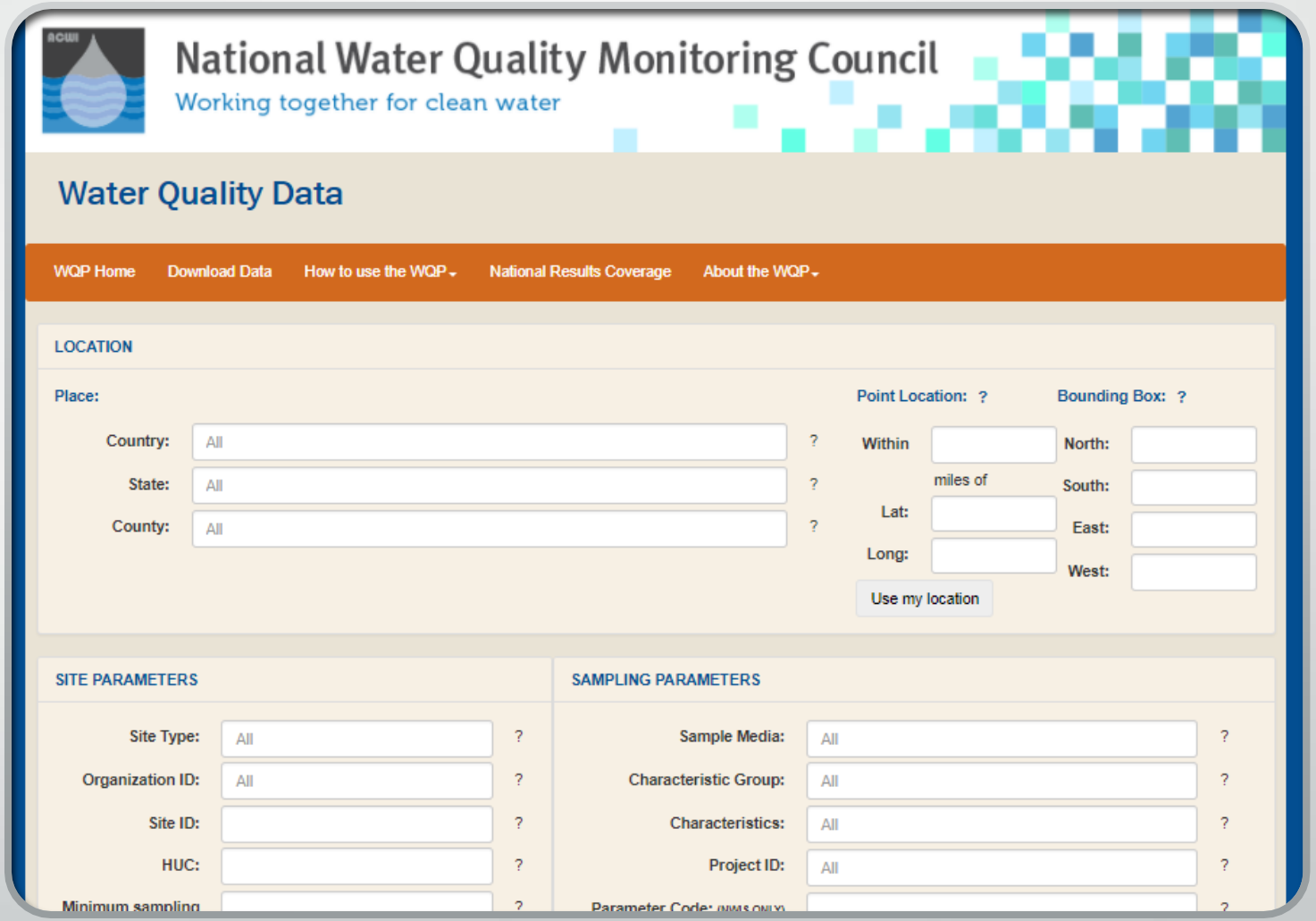
Water Data Integration Branch

Water Quality Portal Tools Session Agenda

- Intro to the WQP tools
- Data Discovery Tool:
  - Demonstration
  - ***New*** QAQC Module
- Data Analysis Tool:
  - Introduction
  - Demonstration
  - Identified Gaps and Future Enhancements
- Future of Tools?

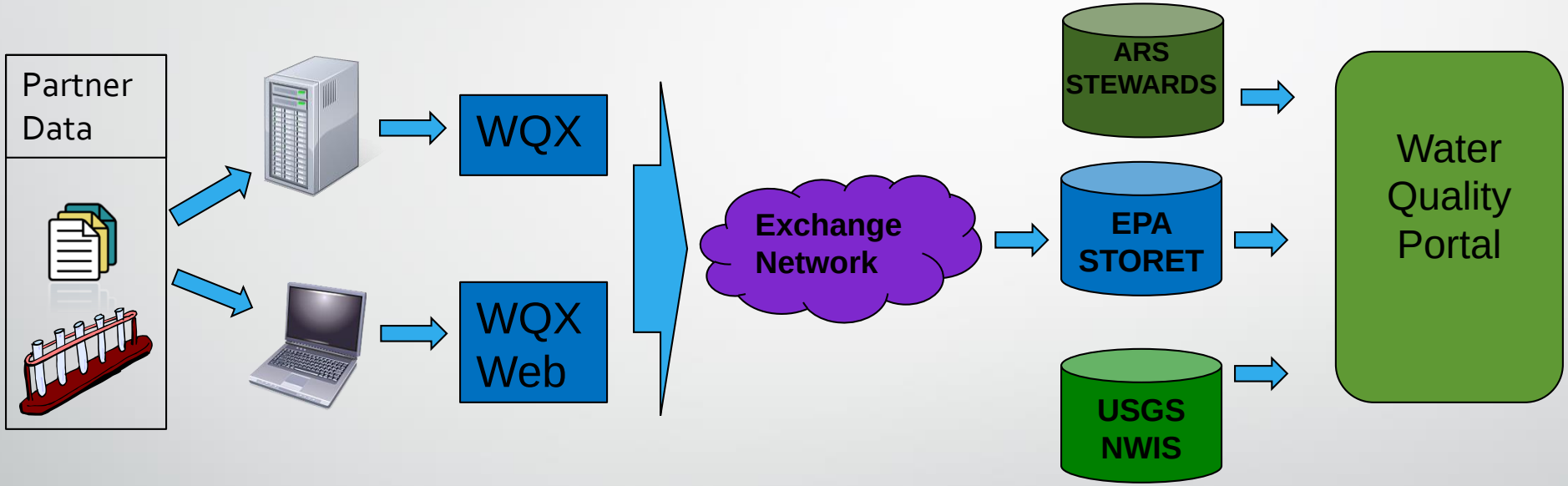
What is the Water Quality Portal?

- One-stop shop for water quality monitoring data (both data shared with EPA through WQX (over 400 partners) and USGS data)
- 349 million water quality results collected at close to 1 million monitoring locations
- Data use common terminology across USGS and EPA data (WQX format)
- Partnership led by the National Water Quality Monitoring Council



https://waterqualitydata.us

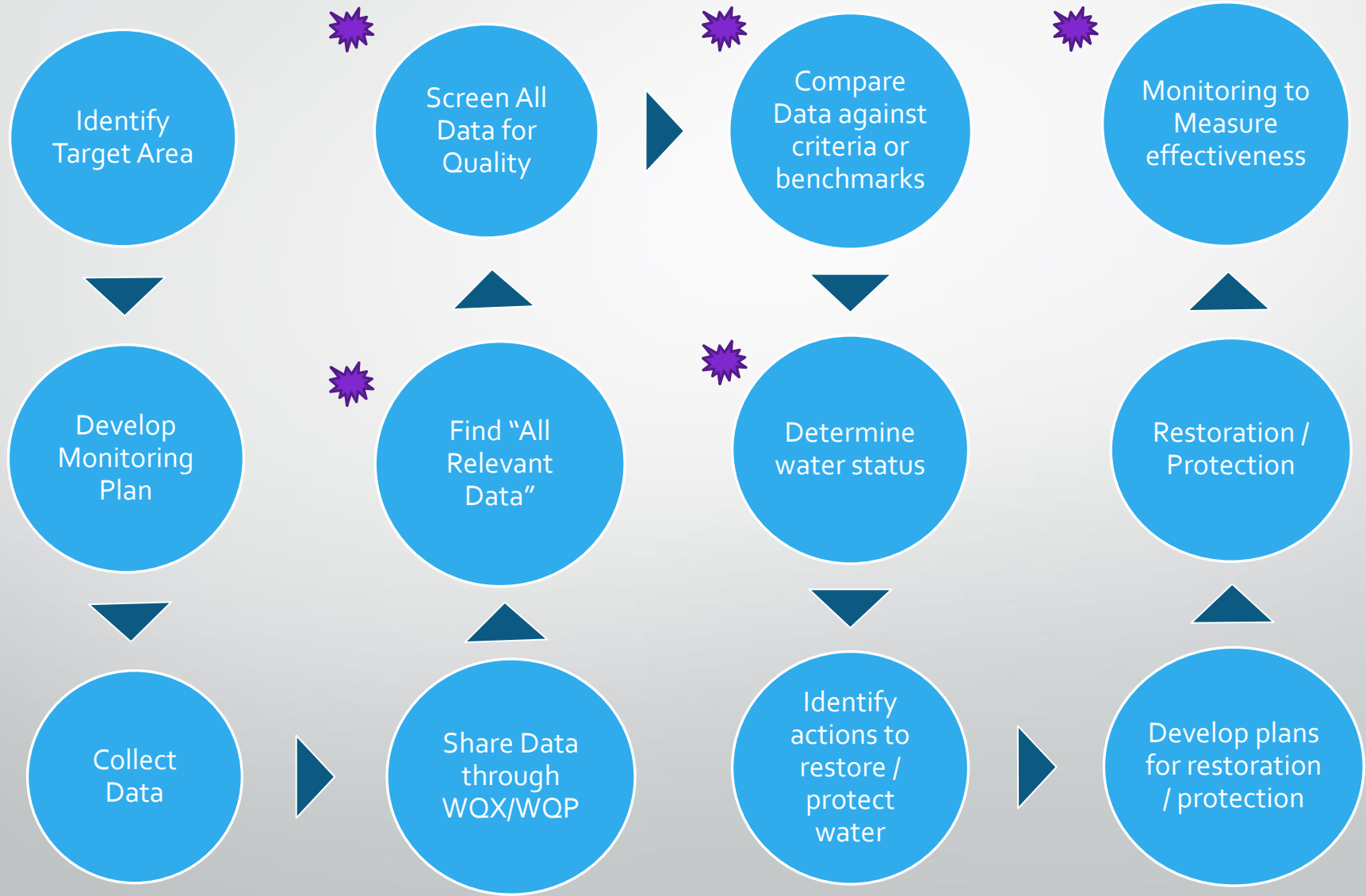
The role of WQX in Data Sharing



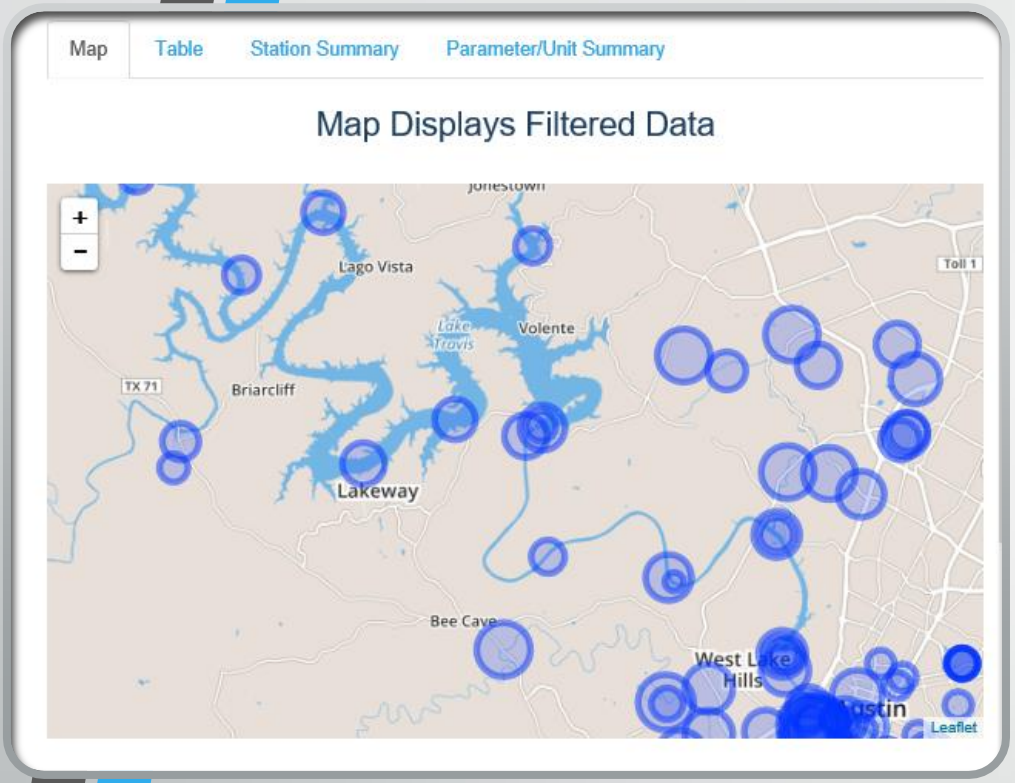
For more information on the portal see: https://waterqualitydata.us

State Water Quality Assessment and Protection Process

= Tools aid process



The Water Quality Portal Tools



Keys to Success in Development:

- Shiny web app- User interface
- R-portable- No need to download R to your machine
- Built using Water Quality Portal offerings via web services
- USGS dataretrieval package
- User testing and feedback

Functionality:

- Data Discovery-find a dataset of interest, QA checks and process out 'bad data', and deal with non-detects
- Data Analysis- assess water quality monitoring data against water quality criteria, including Acute/Chronic thresholds, Trend analysis, and Metals criteria calculations

WQP Data Discovery Tool Demonstration

QA/QC Enhancements to Data Discovery Tool

	A	B	C	D	G	H	I	J	K	L	M
1	DATA FIELDS FROM WATER QUALITY PORTAL				DATA DISCOVERY TOOL TRANSFORMATIONS / DATA SCREENS						
2	methods pulled from the results table				TRANSFORMATIONS / DATA SCREENS						
3											
4	ActivityMediaName	CharacteristicName	ResultSampleFractionText	ResultMeasure.MeasureUnitCo	apply	Characteristic	Unit	unitsConvMult	SampleFraction	qcMin	qcMax
5											
6	Activity Media	Characteristic	Sample Fraction	Units	Apply QAQC	Characteristic QAQC	Units QAQ	Units Conv Mu	Sample Frac	QC Min	QC Ma
14	Water	Chlorophyll a, uncorrected for pheophytin		mg/m3	TRUE	Chlorophyll a	µg/L	1	--	0	2000
15	Water	Chlorophyll a		mg/m2	TRUE	Chlorophyll a	mg/m2	1	--	0	NA
16	Water	Chlorophyll a, corrected for pheophytin	Total	ug/l	TRUE	Chlorophyll a, corrected	µg/L	1	--	0	2000
17	Water	Chlorophyll a (probe)		ug/l	TRUE	Chlorophyll a, probe	µg/L	1	--	0	2000
18	Water	Chlorophyll b	Total	ug/l	TRUE	Chlorophyll b	µg/L	1	--	0	2000
19	Water	Chlorophyll b	Total	mg/l	TRUE	Chlorophyll b	µg/L	1.00E+03	--	0	2
20	Water	Pheophytin a	Total	ug/l	TRUE	Pheophytin a	µg/L	1	--	0	2000
21	Water	Pheophytin a	Extractable	ug/l	TRUE	Pheophytin a	µg/L	1	--	0	2000
22	Water	Barometric pressure		mmHg	TRUE	Barometric pressure	mmHg	1	--	660	825
23	Water	Barometric pressure	Total	mmHg	TRUE	Barometric pressure	mmHg	1	--	660	825
24	Water	Barometric pressure		psi	TRUE	Barometric pressure	mmHg	51.7149	--	12.76	15.95
25	Water	Barometric pressure	Dissolved	mmHg	TRUE	Barometric pressure	mmHg	1	--	660	825
26	Water	Biochemical oxygen demand, standard conditions		mg/l	TRUE	Biochemical oxygen demand, 5 day	mg/L	1	--	0	50
27	Water	Biochemical oxygen demand, standard conditions	Total	mg/l	TRUE	Biochemical oxygen demand, 5 day	mg/L	1	--	0	50
28	Water	Biochemical oxygen demand, standard conditions	Dissolved	mg/l	TRUE	Biochemical oxygen demand, 5 day	mg/L	1	--	0	50
29	Water	Chemical oxygen demand	Total	mg/l	TRUE	Chemical oxygen demand	mg/L	1	--	0	50
30	Water	Chemical oxygen demand		mg/l	TRUE	Chemical oxygen demand	mg/L	1	--	0	50
31	Water	Specific conductance		uS/cm	TRUE	Specific conductance	µS/cm	1	--	20	1E+06
32	Water	Specific conductance	Total	uS/cm	TRUE	Specific conductance	µS/cm	1	--	20	1E+06
33	Water	Specific conductance	Total	umho/cm	TRUE	Specific conductance	µS/cm	1	--	20	1E+06
34	Water	Specific conductance		umho/cm	TRUE	Specific conductance	µS/cm	1	--	20	1E+06

Methods Table

Questions?

Github link: https://github.com/USEPA/Water-Quality-Data-Discovery-Tool

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

Interoperable Watersheds Network (Currents): http://54.210.62.171/

Britt Dean

202-566-1020

Dean.britt@epa.gov

Shelly Thawley

202-566-1214

Thawley.michelle@epa.gov

Jillian Adair

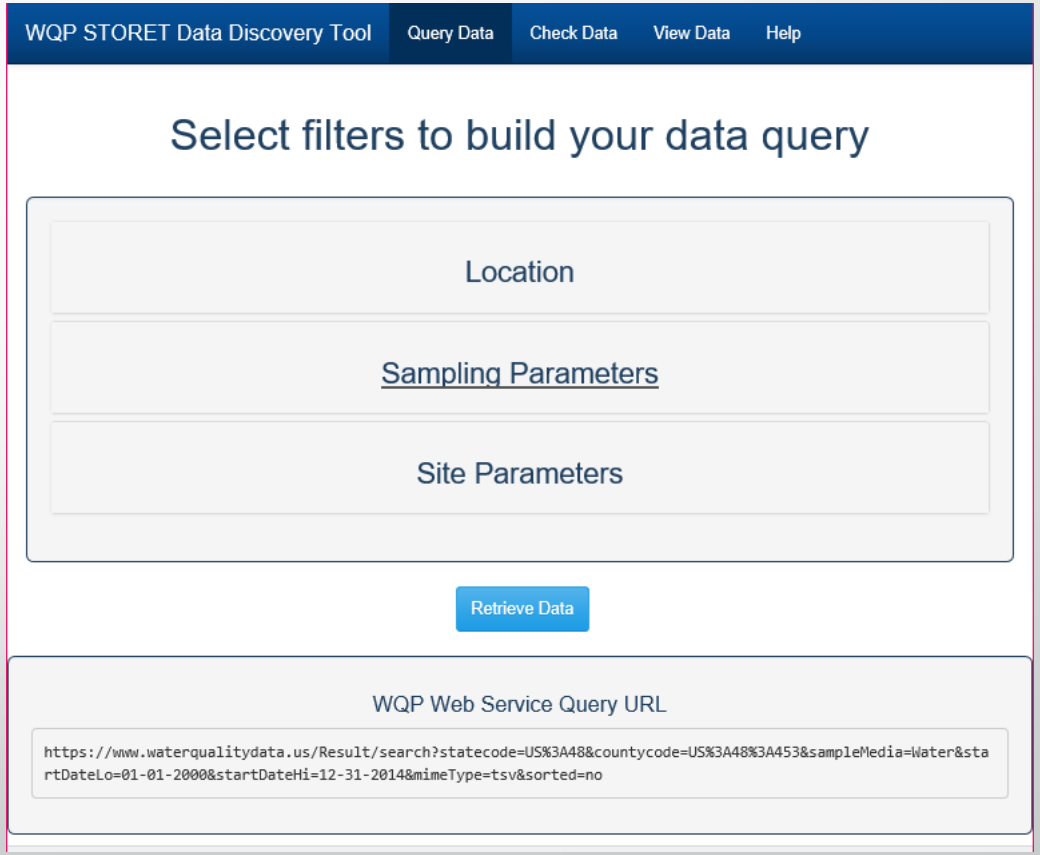
215-814-5713

Adair.jillian@epa.gov

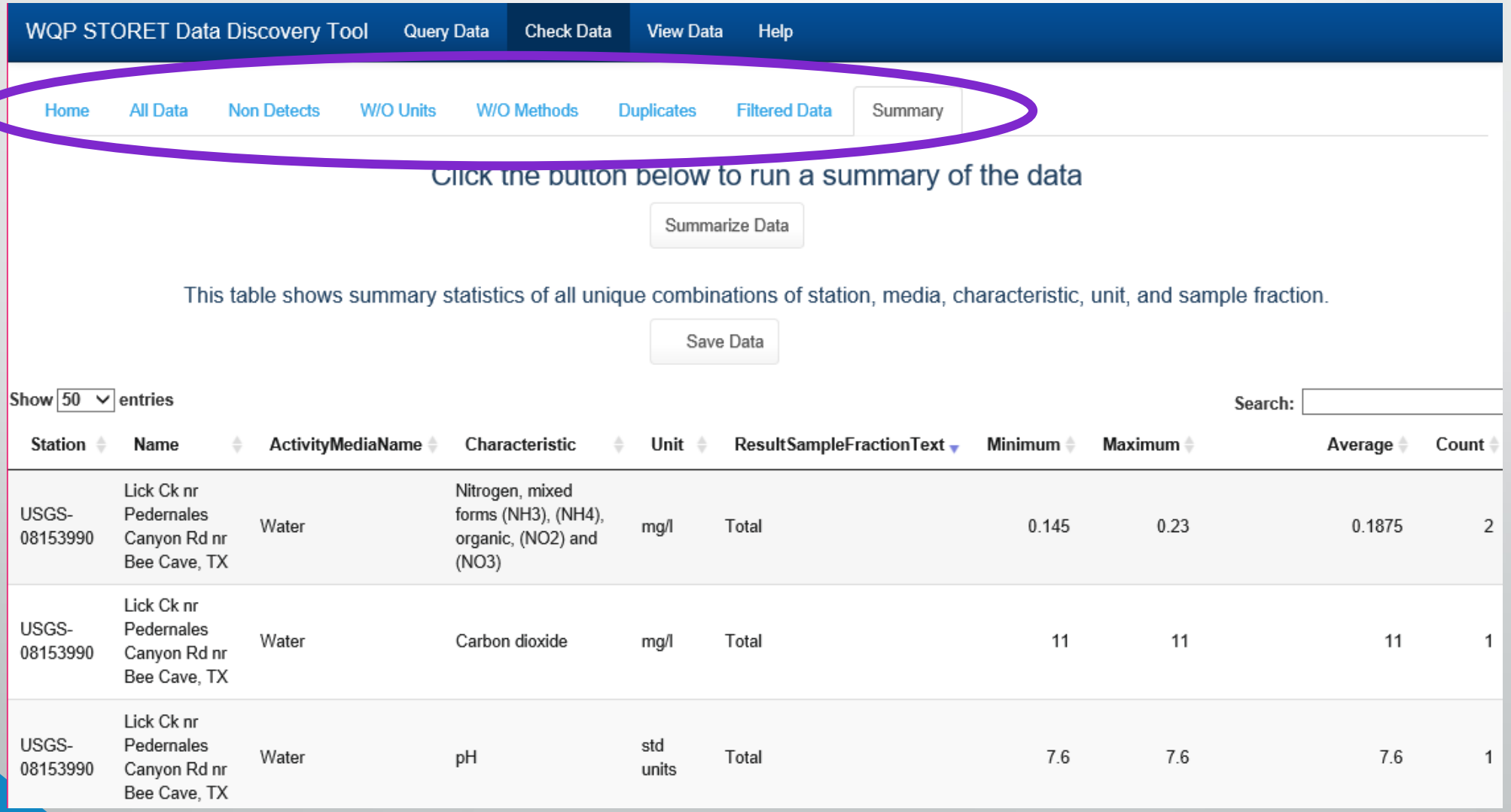
Extra, Extra!

Data Discovery Tool Capabilities: Finding Data

- Dynamic Data Source
- Data Retrieval Package enables live querying, packaging, and delivery of data from Portal
- Set up original query to portal- data is then imported into tool

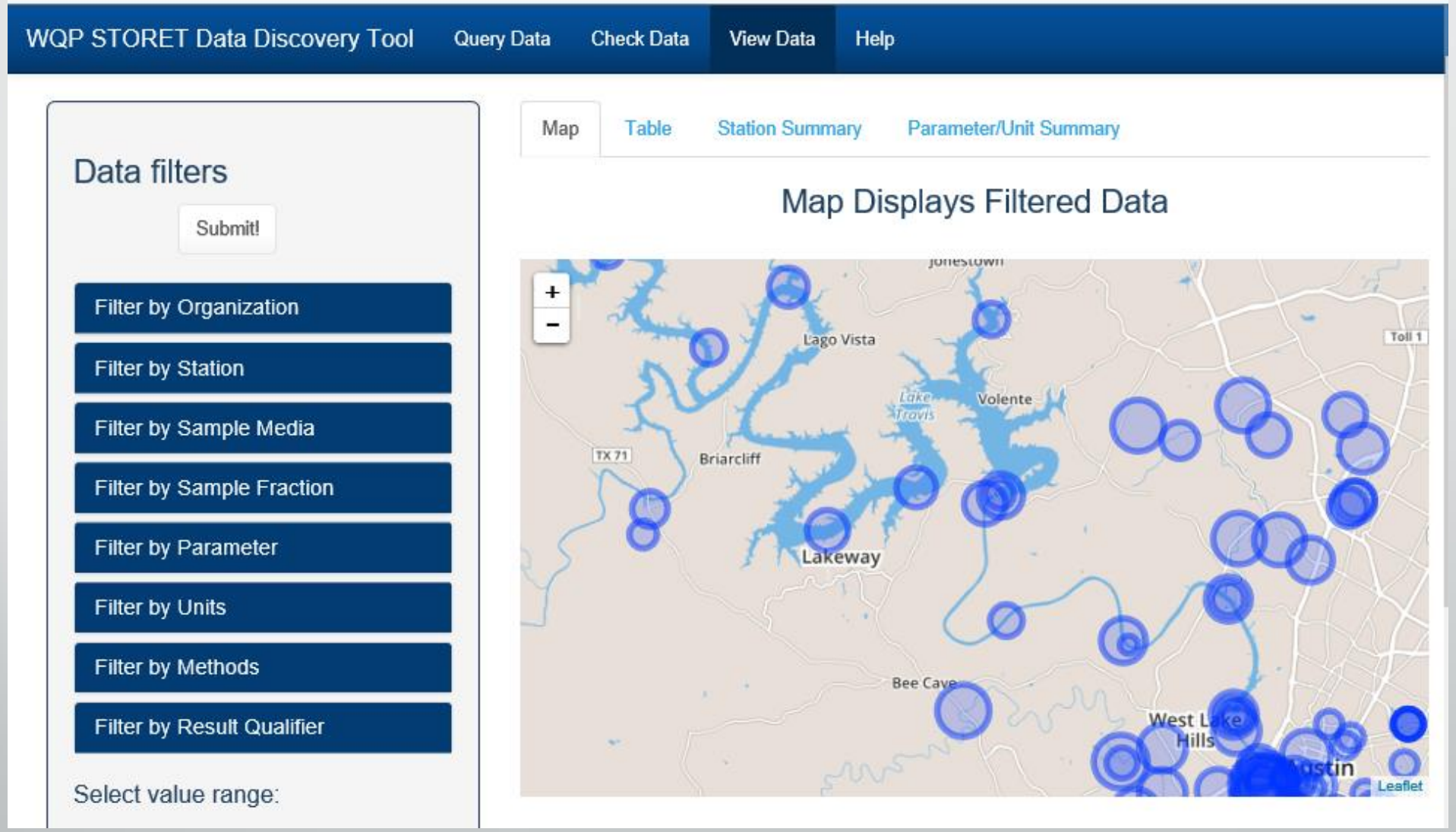


Data Discovery Tool Capabilities: Processing out unwanted Data

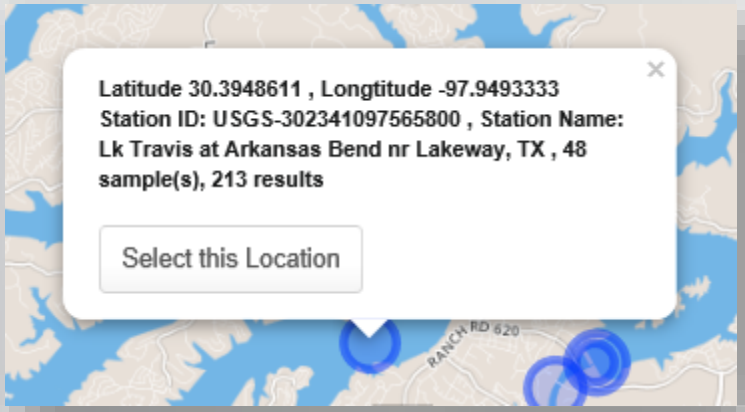


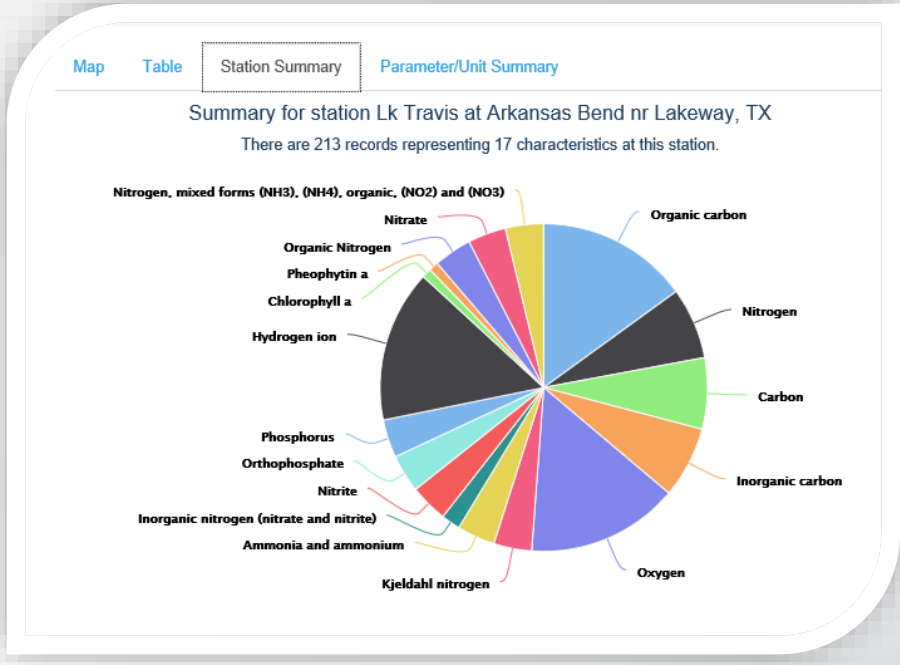
Station	Name	ActivityMediaName	Characteristic	Unit	ResultSampleFractionText	Minimum	Maximum	Average	Count
USGS-08153990	Lick Ck nr Pedernales Canyon Rd nr Bee Cave, TX	Water	Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3)	mg/l	Total	0.145	0.23	0.1875	2
USGS-08153990	Lick Ck nr Pedernales Canyon Rd nr Bee Cave, TX	Water	Carbon dioxide	mg/l	Total	11	11	11	1
USGS-08153990	Lick Ck nr Pedernales Canyon Rd nr Bee Cave, TX	Water	pH	std units	Total	7.6	7.6	7.6	1

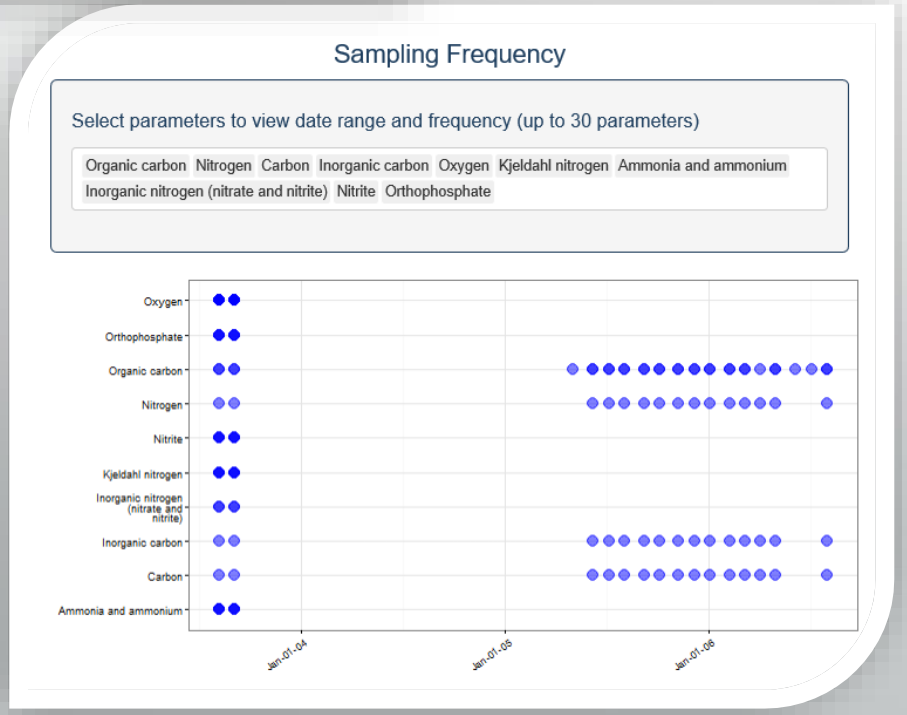
Data Discovery Tool Capabilities: Exploring the dataset

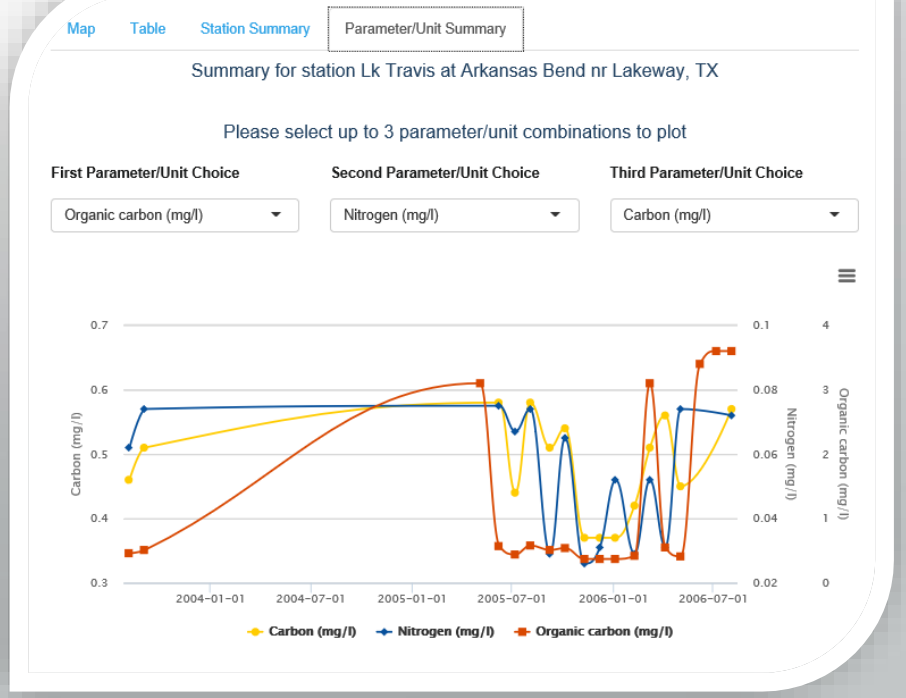


Data Discovery Tool Capabilities: Exploring the Data set









WQP STORET Data Discovery Tool | Query Data | Check Data | View Data | Help

Save Data

Click to download a .tsv file containing the data being viewed.

Save Data

Data filters

Submit!

- Filter by Organization
- Filter by Station
- Filter by Sample Media
- Filter by Sample Fraction
- Filter by Parameter
- Filter by Units
- Filter by Methods
- Filter by Result Qualifier

Select value range:

Map | Table | Stati

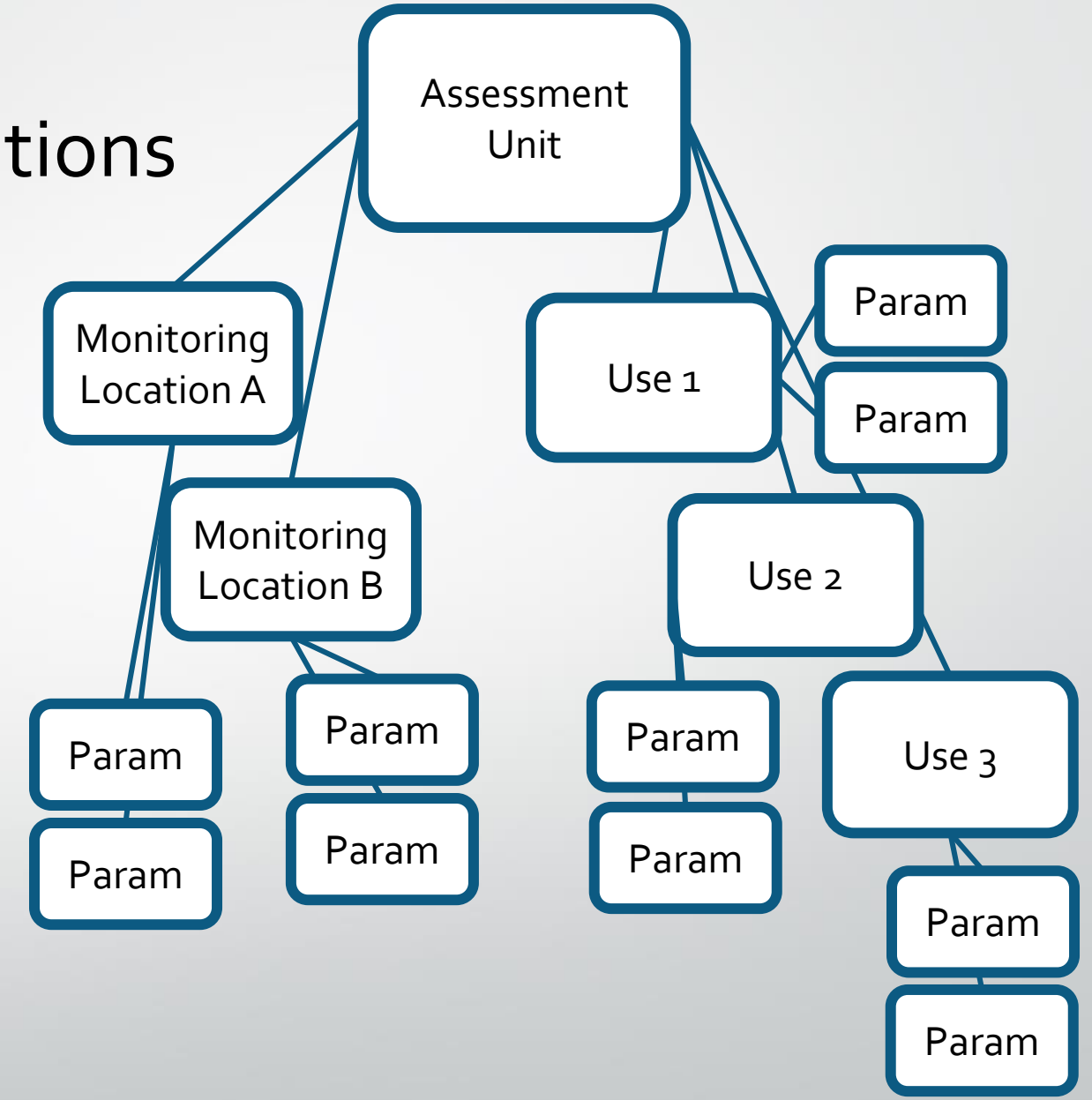
Show 100 entries Search:

Station	Name	Organization	Characteristic	Result	Unit	Method	Activity
NALMS-5735	Austin at City Park	NALMS	Depth, Secchi disk depth	1	m		2007-06-
NALMS-5735	Austin at City Park	NALMS	pH	7	None		2007-06-
NALMS-5735	Austin at City Park	NALMS	Dissolved oxygen (DO)	6	ppm		2007-06-
NALMS-5735	Austin at City Park	NALMS	Temperature, water	22	deg C		2007-06-
NALMS-5813	Travis	NALMS	Depth, Secchi disk depth	1.4	m		2008-07-
NALMS-5813	Travis	NALMS	Temperature, water	28	deg C		2008-07-

- Criteria
- Metals Analysis
- Analysis
- Map
- Station summary
- Assessment Unit Summa

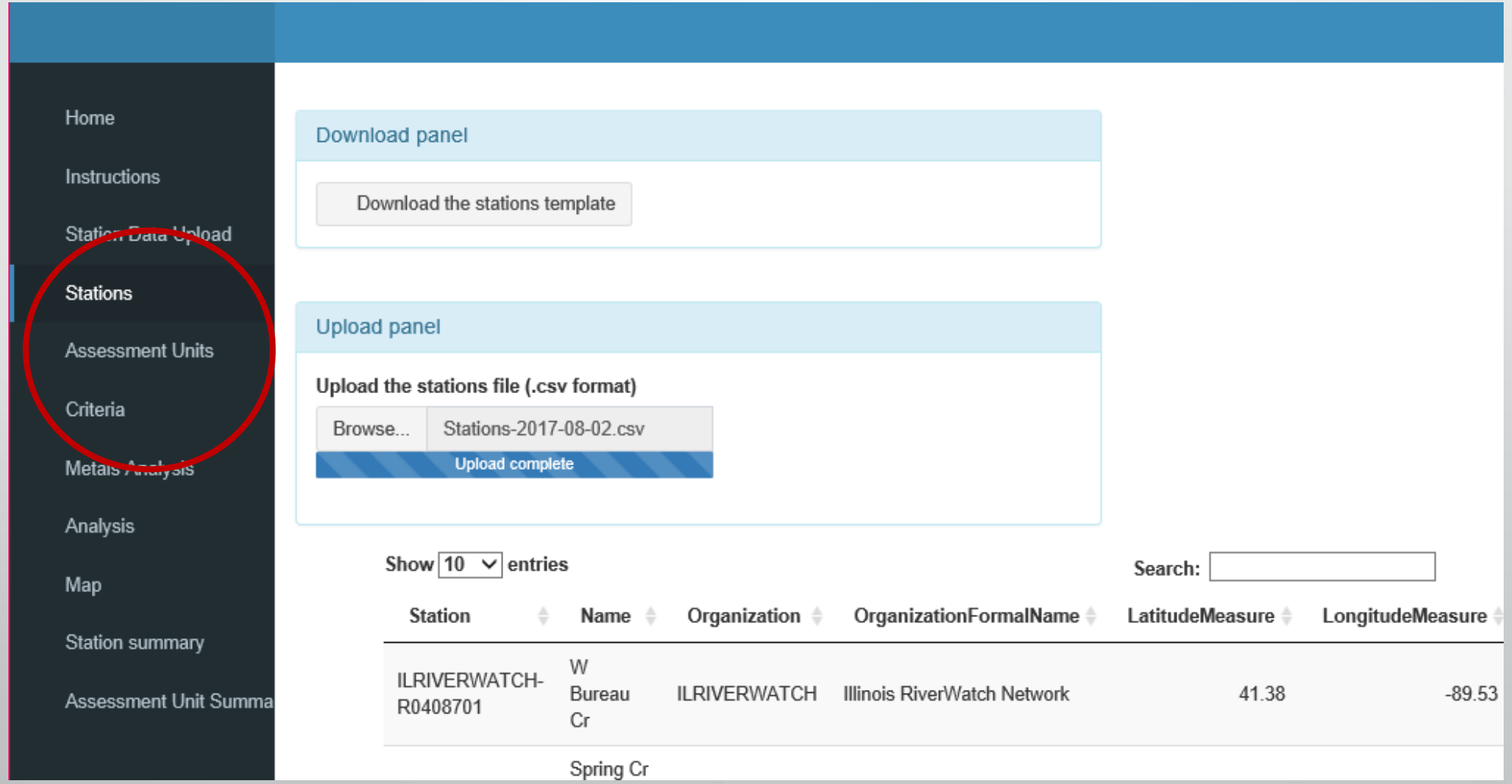
Data Analysis Tool: Development Considerations

- Variance between organizations:
  - statistics used and way they are applied
  - tool has to be flexible in the areas that vary, but not so flexible as to be complicated to use
- How to input many to many relationships:
  - User interface for inputting parameters for statistical tests OR Batch upload of a flat file with criteria enumerated
  - Some tests lend themselves more to a UI and some in flat files



Data Analysis Tool Capabilities: Inputting Data and Criteria

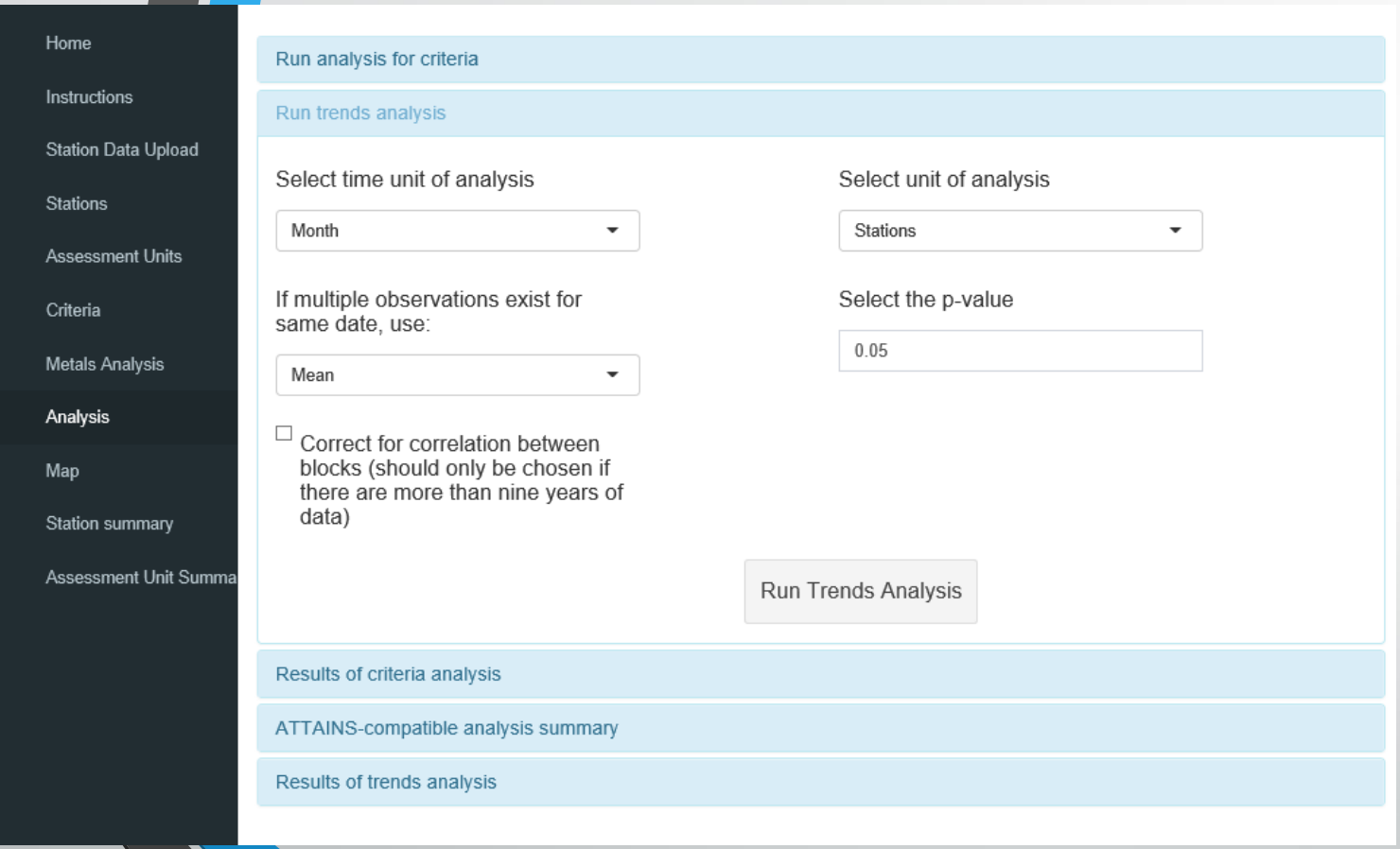
Series of flat files relating stations, to assessment units, to designated uses, and water quality criteria



Data Analysis Tool Capabilities: Assessing against criteria

	PARM	Units	USE_OR_CLASS	Media	Sample_Fraction	Criterion	Limit	Comparison
2	Conductivity	uS/cm	Aquatic Life	Water	Total	Upper	1000	GT
3	Dissolved Oxygen (DO)	mg/l	Aquatic Life	Water	NA	Lower	4	LT
4	Escherichia Coli	MPN/100	Primary Contact Recreation	Water	NA	Geomean	100	GT
5	Fluoride	mg/l	Aquatic Life	Water	Dissolved	Upper	1.4	GT
6	Fluoride	mg/l	Aquatic Life	Water	Total	Upper	1.4	GT
7	Iron	ug/l	Aquatic Life	Water	Total	Upper	1000000	GT
8	Iron	ug/l	Aquatic Life	Water	Dissolved	Upper	1000008	GT
9	Iron	mg/kg	Aquatic Life	Sediment	Total	Upper	1000008	GT
10	Kjeldahl Nitrogen	mg/l	Aquatic Life	Water	Total	Upper	4	GT
11	Kjeldahl Nitrogen	mg/l as N	Aquatic Life	Water	Dissolved	Upper	4	GT
12	Kjeldahl Nitrogen	mg/kg	Aquatic Life	Water	Sediment	Upper	3	

Data Analysis Tool Capabilities: Detecting Trends



- Detects trends in long term data sets with seasonality
- Considerations for setting parameters of test:
  - Length of data
  - Frequency of data
  - #Seasonal Bins
  - P-value to denote significance
- Test output in different format than other criteria (i.e exceedance y/n vs reported tau, p-value, trend y/n, significance y/n)

Data Analysis Tool Capabilities: Calculating metals criteria (beta)

- Metals criteria for Aquatic Life, to compare WQ data against is hardness dependent
  - Bioavailability of metal dependent on hardness of water
  - Spreadsheet of coefficients
  - Default hardness for now
  - Future enhancements to include sweep of dataset for all hardness related characteristics (i.e. ph, Calcium, Magnesium, Hardness, etc.)

	A	B	D	F	G	H	I	J	K
1	Metal	Media	USE_OR_CLASS	Unit	Criteria.Type	m	b	Hardness_ default	CF
2	Arsenic	Water	NA	ug/L	Freshwater Acute	NA	NA	3	1
3	Arsenic	Water	NA	ug/L	Freshwater Chronic	NA	NA	4	1
4	Arsenic	Water	NA	ug/L	Saltwater Acute	NA	NA	NA	NA
5	Arsenic	Water	NA	ug/L	Saltwater Chronic	NA	NA	NA	NA
6	Cadmium	Water	NA	ug/L	Freshwater Acute	0.9789	-3.866	2	NA
7	Cadmium	Water	NA	ug/L	Freshwater Chronic	0.7977	-3.909	2.5	NA
8	Cadmium	Water	NA	ug/L	Saltwater Acute	NA	NA	NA	NA
9	Cadmium	Water	NA	ug/L	Saltwater Chronic	NA	NA	NA	NA
10	Chromium III	Water	NA	ug/L	Freshwater Acute	0.819	3.7256	5	0.316
11	Chromium III	Water	NA	ug/L	Freshwater Chronic	0.819	0.6848	1	0.86
12	Chromium III	Water	NA	ug/L	Saltwater Acute	NA	NA	NA	NA
13	Chromium III	Water	NA	ug/L	Saltwater Chronic	NA	NA	NA	NA
14	Chromium VI	Water	NA	ug/L	Freshwater Acute	NA	NA	4.5	0.982
15	Chromium VI	Water	NA	ug/L	Freshwater Chronic	NA	NA	6	0.962
16	Chromium VI	Water	NA	ug/L	Saltwater Acute	NA	NA	NA	NA
17	Chromium VI	Water	NA	ug/L	Saltwater Chronic	NA	NA	NA	NA
18	Copper	Water	NA	ug/L	Freshwater Acute	NA	NA	3.5	0.96
19	Copper	Water	NA	ug/L	Freshwater Chronic	NA	NA	2	0.96
20	Copper	Water	NA	ug/L	Saltwater Acute	NA	NA	NA	NA
21	Copper	Water	NA	ug/L	Saltwater Chronic	NA	NA	NA	NA
22	Lead	Water	NA	ug/L	Freshwater Acute	1.273	-1.46	4	NA

Exercise 1: Data Download

If you don't already have the tool, you can find it at: https://www.epa.gov/waterdata/water-quality-portal-data-discovery-tool

1. Download a data set of interest. It could be data from your organization (if it has been published through WQX) or a watershed or county that you're interested in.
2. Use the tool to download the data.
    1. Open the tool by double clicking on DiscoveryTool.bat
    2. On the download form, choose your organization, or county, or watershed(you may need to add a date filter to limit the size of the data, it must be less than 200,000 records).
    3. You will first get a count of records that will be retrieved, then click 'import'. This may take some time.

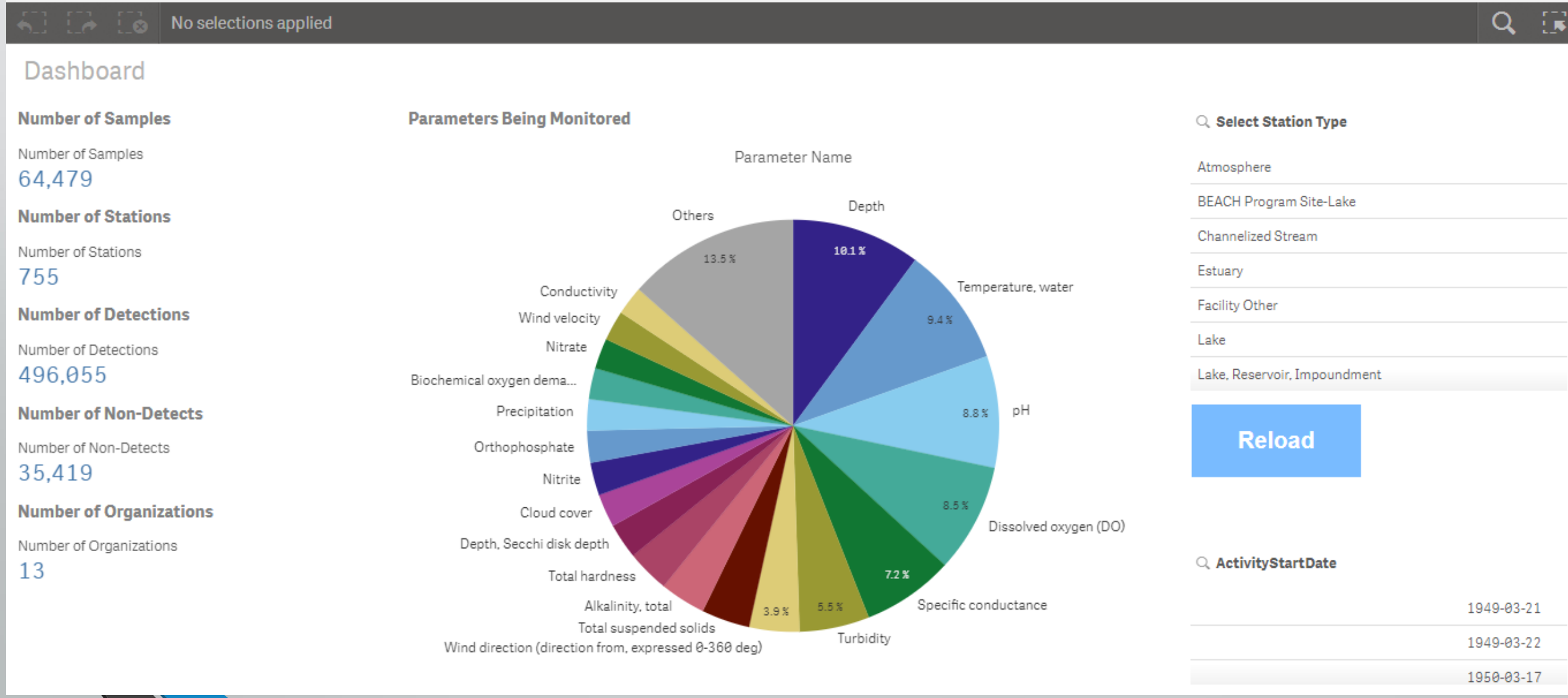
Exercise 2: Check Data

1. Click on the Check Data Tab
2. Look for duplicates, data without methods, make a decision on what to do with non-detects
3. Create a data summary and save to a file
    1. Open that file in Excel and explore the data. Look for anything that might be out of place (i.e. a max or min that doesn't make sense). Note the location or station.

Exercise 3: View Data

1. Click on the View Data Tab
2. Answer some questions that you might be interested in
   - Where do you have concentrations of samples?
   - Where isn't being sampled?
   - Who else might be collecting data (if you only downloaded your organization's data, you can't answer this question)
   - Are monitoring locations correct?
3. Explore some parameters
   - Find where those parameters are sampled
   - Look for places where that parameter is above a certain level (or below a level).
4. Filter down to a data set of interest and export that data out
5. Do a line chart of 1-3 parameters for a station and export out a graphic and paste in a Word Document

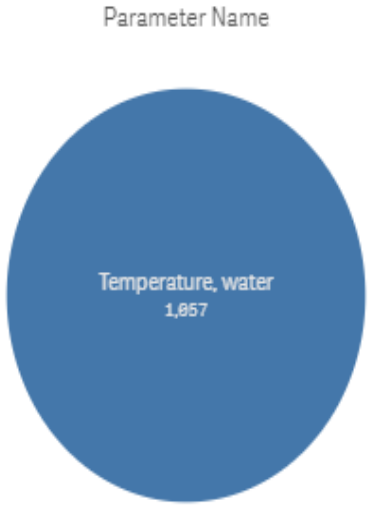
One More Thing



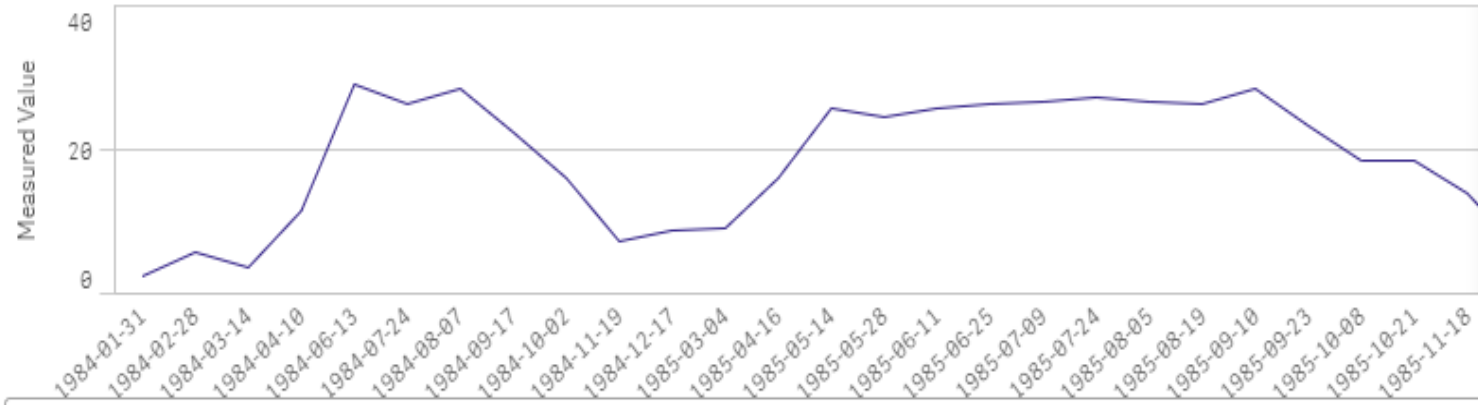
CharacteristicN... Temperature, water | MonitoringLoca... ANA14

Data Evaluation

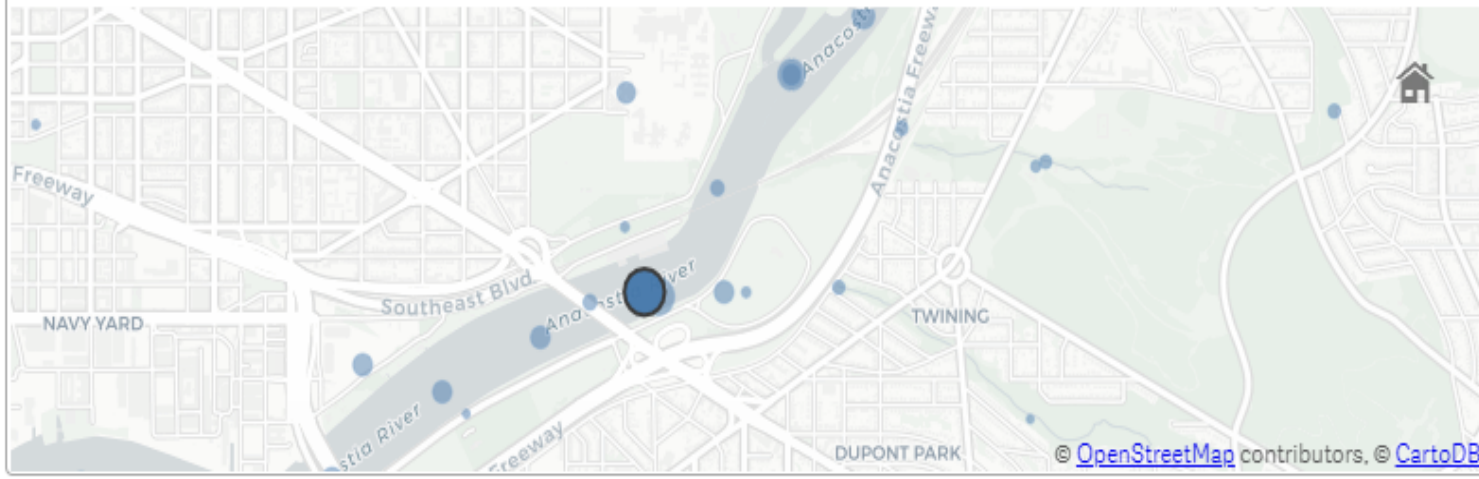
Parameters Being Monitored



Concentrations of Time

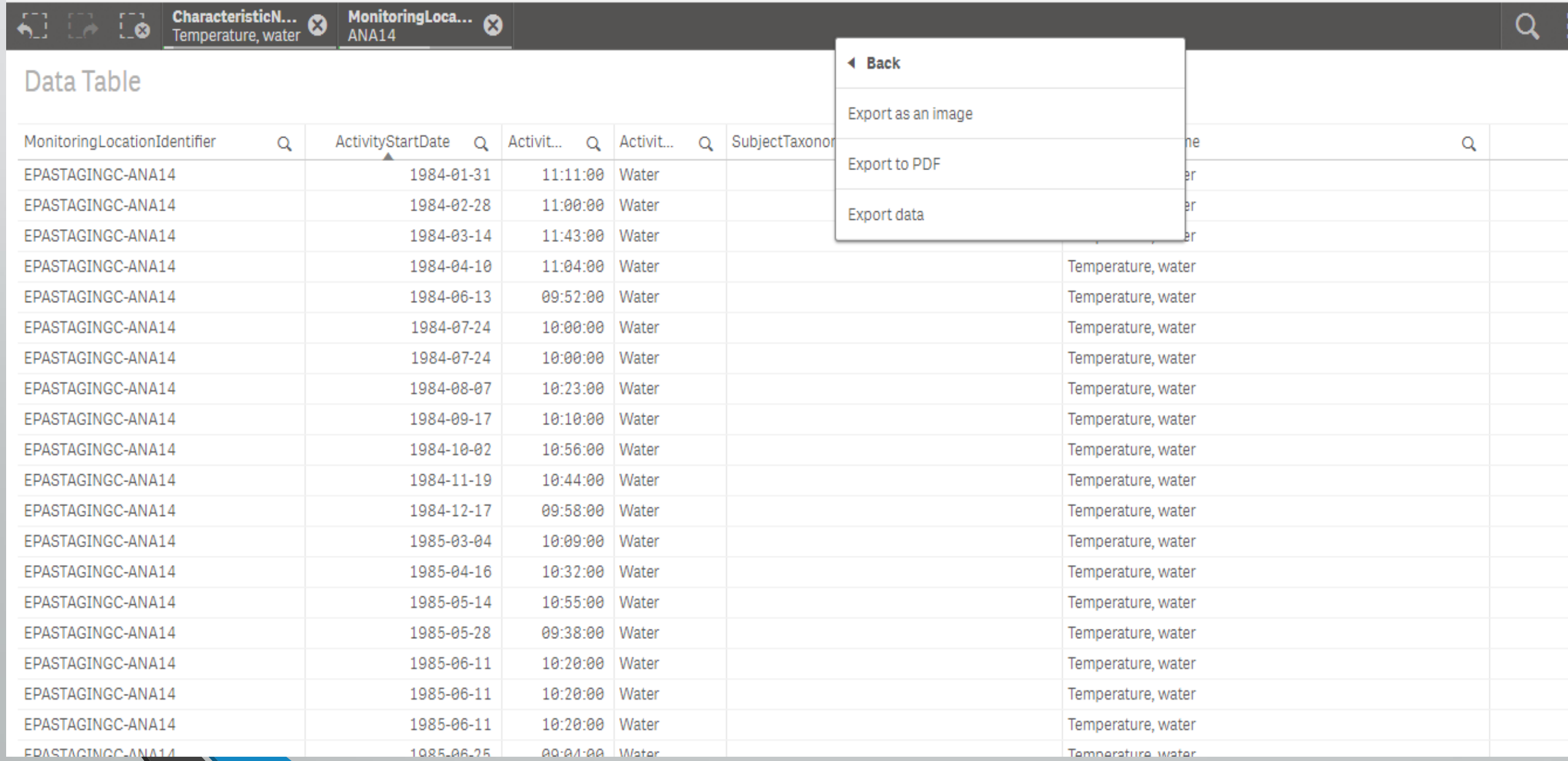


Map of Stations



Stations

MonitoringLocationIdentifier	MonitoringLocationName
EPASTAGINGC-ANA14	ANA14



Data Table

MonitoringLocationIdentifier	ActivityStartDate	Activit...	Activit...	SubjectTaxonor	ne
EPASTAGINGC-ANA14	1984-01-31	11:11:00	Water		er
EPASTAGINGC-ANA14	1984-02-28	11:00:00	Water		er
EPASTAGINGC-ANA14	1984-03-14	11:43:00	Water		er
EPASTAGINGC-ANA14	1984-04-10	11:04:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-06-13	09:52:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-07-24	10:00:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-07-24	10:00:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-08-07	10:23:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-09-17	10:10:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-10-02	10:56:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-11-19	10:44:00	Water		Temperature, water
EPASTAGINGC-ANA14	1984-12-17	09:58:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-03-04	10:09:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-04-16	10:32:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-05-14	10:55:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-05-28	09:38:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-06-11	10:20:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-06-11	10:20:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-06-11	10:20:00	Water		Temperature, water
EPASTAGINGC-ANA14	1985-06-25	09:04:00	Water		Temperature, water

How do you get data from the portal

- Simple user interface available at https://waterqualitydata.us
- Provides data back in multiple formats (Excel, tab separated, comma separated, KML, and WQX format)

Water Quality
Portal Demo