



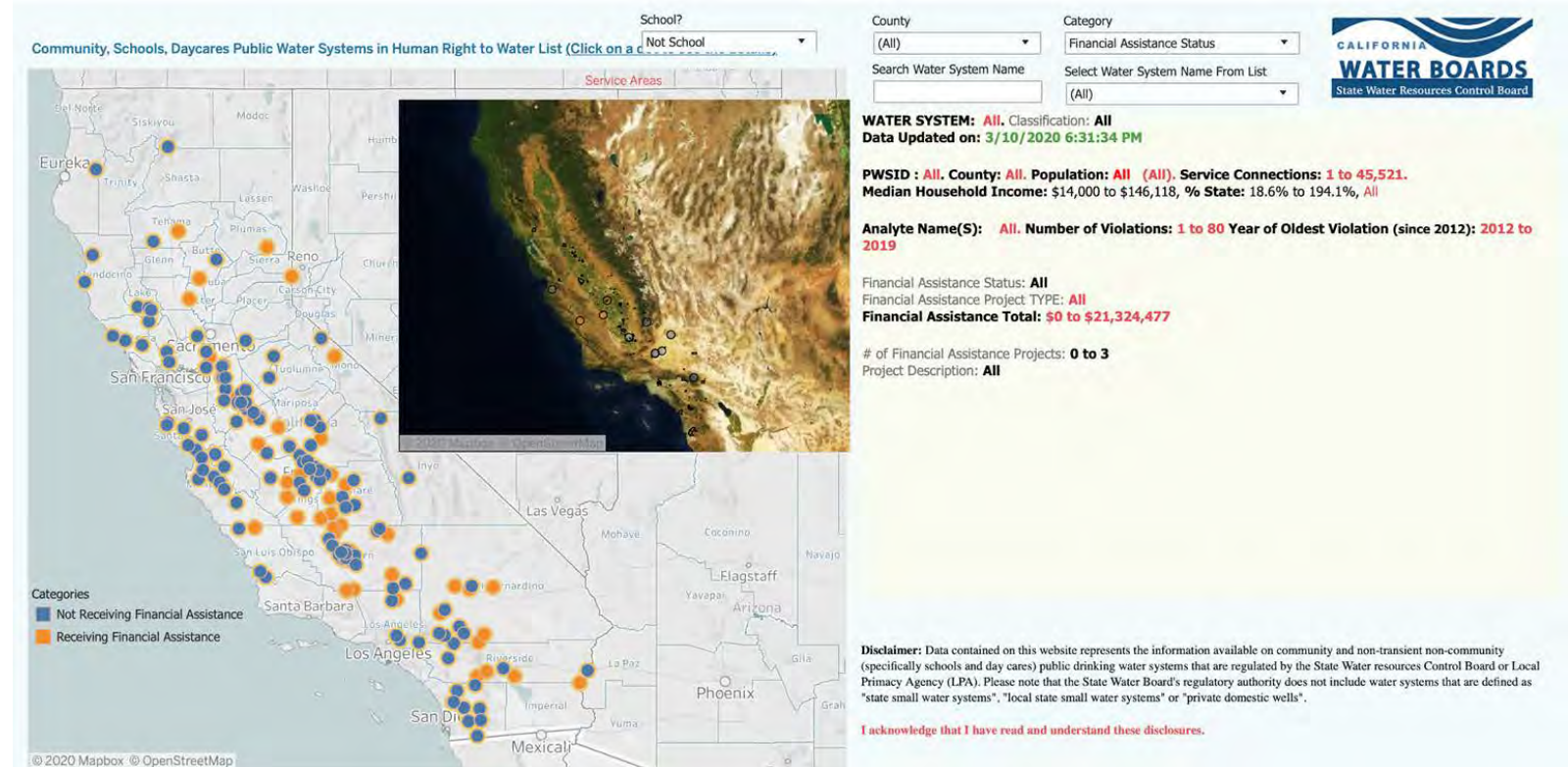
California State Water Resources Control Board  
DATA CENTER

# Water Data Federation: a CA Perspective

Greg Gearheart,  
California State Water Resources Control Board  
Office of Information Management & Analysis  
*New Mexico Water Data Workshop: April 28, 2020*

# Overview

- CA Water Data Ecosystem
- Open and Transparent Water Data Act
- Open Data Resolution
- Departments Efforts to Build Capacity / Infrastructure at WB
- Federation Strategy



# CA Water Boards

- 10 Boards (9 Regional WQ, 1 State WR)
- 68 ( $7*9 + 5*1 = 68$ ) Governor-appointed Board Members
- State Board has water quality, water rights, drinking water, funding and fiscal/administration duties



# Water Boards by the Numbers

- Water Quality, Water Rights and Drinking Water number (as of June 30, 2018)
  - **10** Boards, **43** office locations
  - **18** drinking water offices, **25** State and Regional Board office locations
  - **>45** programs, **~15** core regulatory programs
  - **2300** positions, **345** in supervisory / management positions
  - **\$736,000,000** annual operating budget
  - Regulate **~38,000** dischargers
  - Allocate **~34,000** water right holders
  - Plan/Assess **\$37,600,000** in monitoring
  - Fund **\$902,000,000** in local assistance funding and cleanup
  - Fund **\$10,600,000,000** in loans (as of June 30, 2018)
  - Fund **\$20,000,000** in penalties assessed in 2017

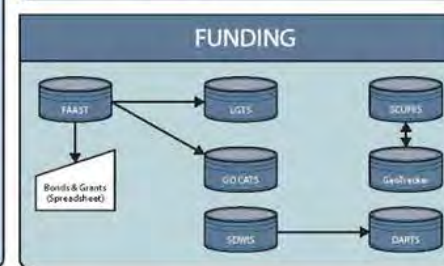
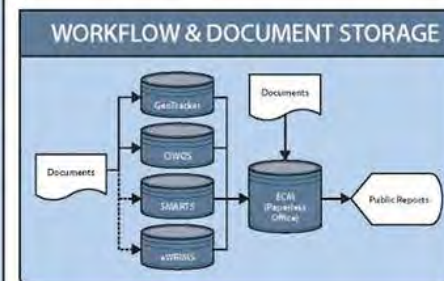
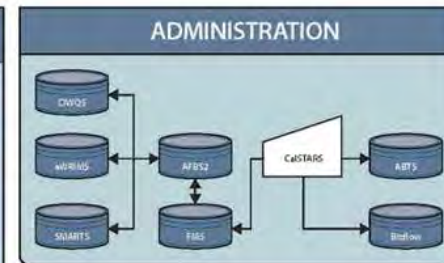
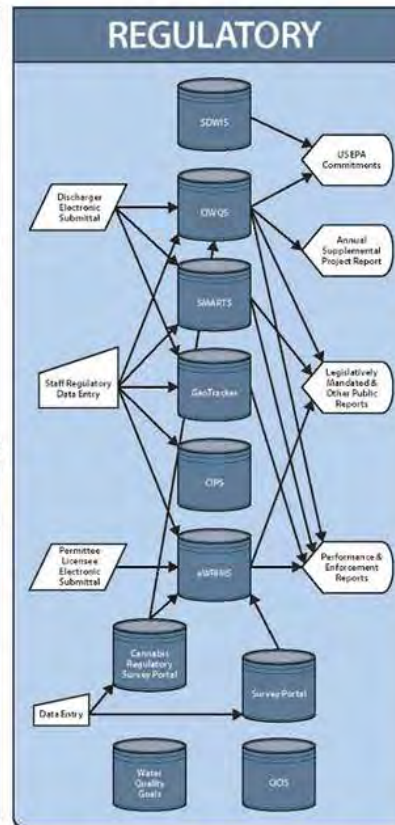
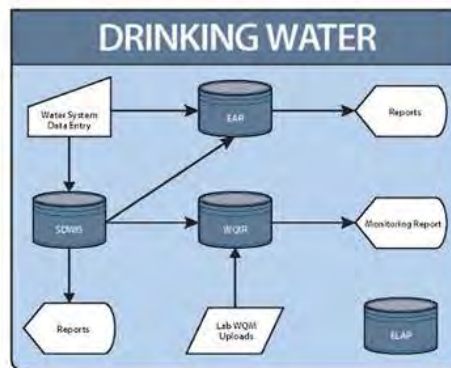
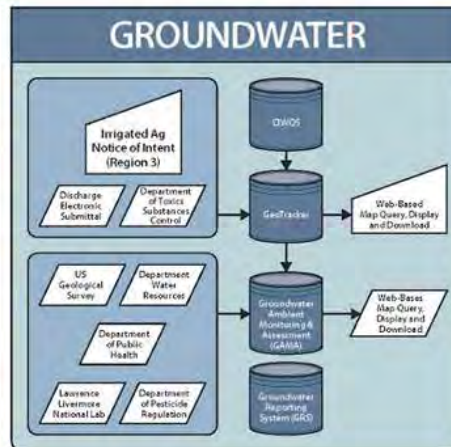
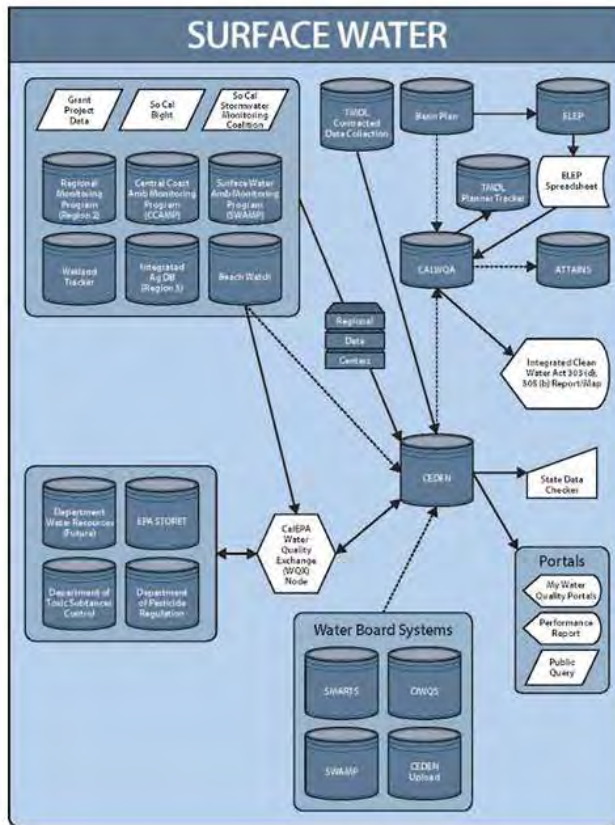
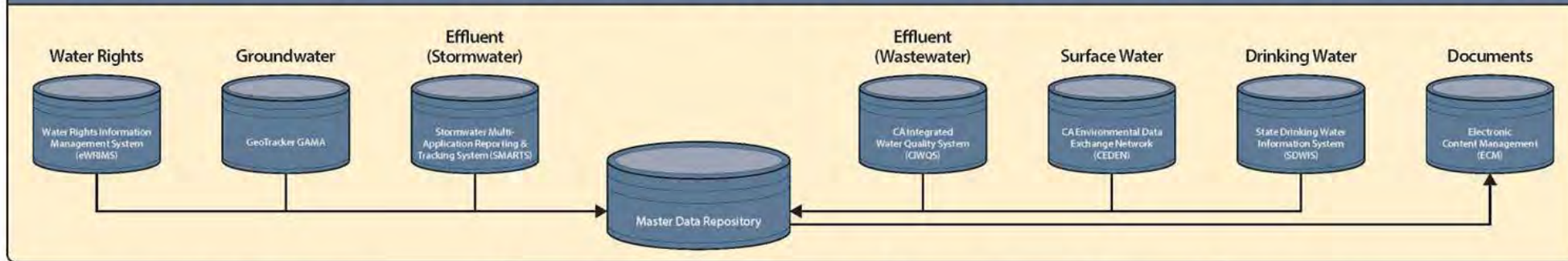
# Background

- September 2016 – **AB 1755: *The Open and Transparent Water Data Act***
  - Develop an integrated platform for existing water and ecological data
  - Make water-related data, tools, and applications developed using state funds publicly accessible
  - Promote principles of openness and interoperability (“making information accessible, discoverable, and usable by the public can foster entrepreneurship, innovation, and scientific discovery”)
- July 2018 – **State Water Resources Control Board Resolution 2018-0032: *Adopting Principles of Open Data as a Core Value and Directing Programs and Activities to Implement Strategic Actions to Improve Data Accessibility and Associated Innovation*** ([link](#))
  - Make Data Accessible (“Open First”): make all critical public data available in machine readable datasets with metadata and data dictionaries
  - Understand Data Quality and Integrity: ensure data are of known and acceptable quality; deploy practices to protect its integrity with standards and protocols

# Fix Data Sources / Architecture

Build Open Architecture and Infrastructure / Inventory our Datasets

# WATER BOARDS' ENTERPRISE TECHNOLOGY APPLICATIONS AND SYSTEMS INFRASTRUCTURE



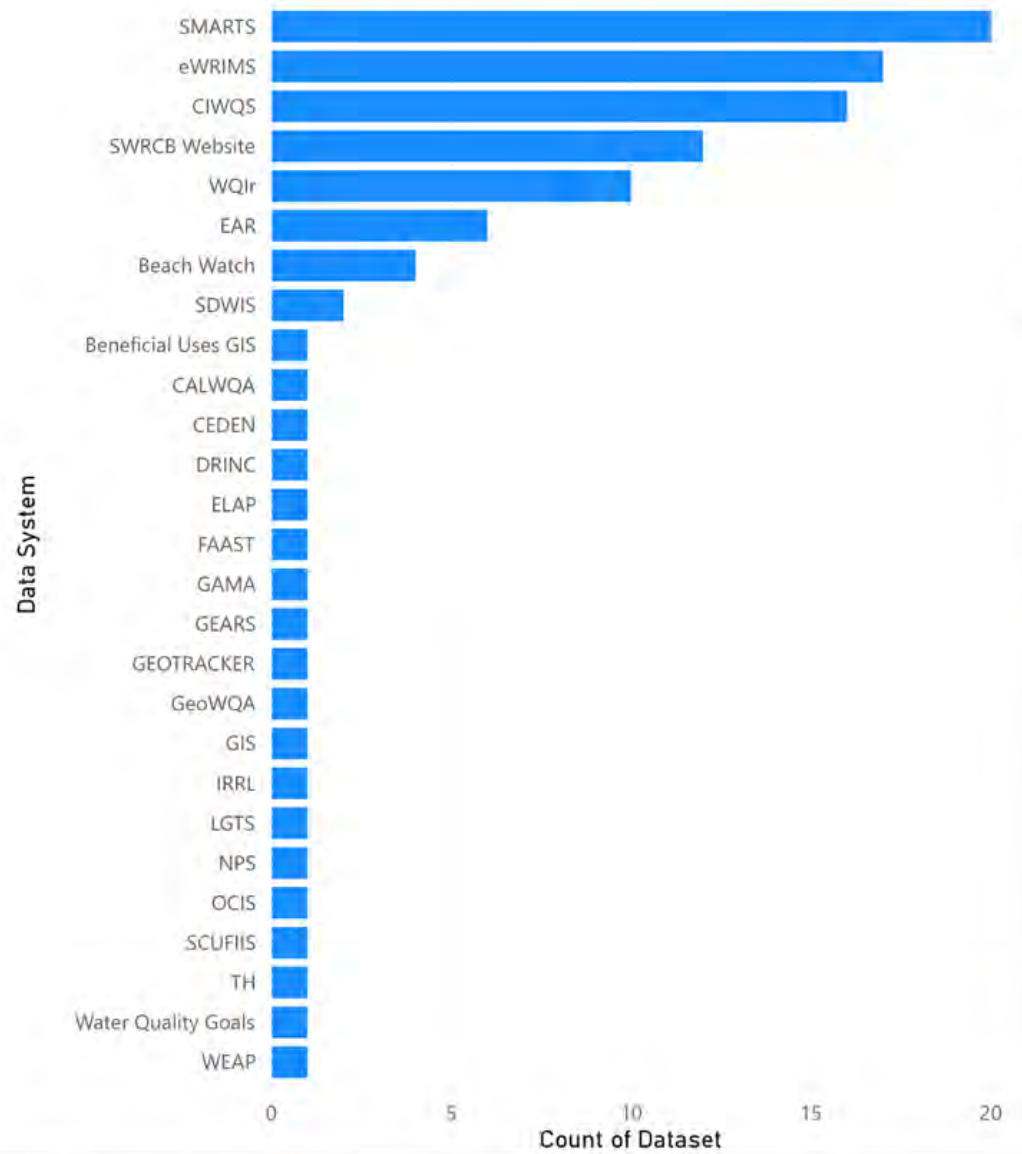




# Inventory of Water Board Data

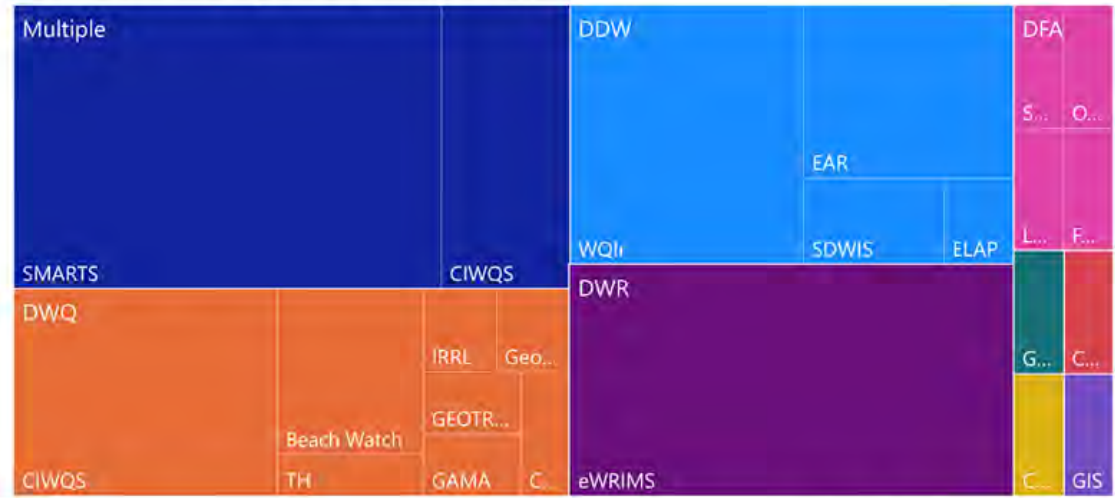
- Interviewed >100 staff and managers
- Assessed open data readiness
- Identified dataset opportunities
  - Observation data
    - Water Quality
    - Water Use / Resource
  - Regulatory data
  - Administrative data
- Over 100 datasets in current inventory

### Count of Dataset by Data System



Division	Application	Program	Count of Dataset
	and Excel files made using Jasper Reports java library for multiple non-java applications: Cost Recovery, SCUFIS, Bonds and Grants)		
	CIWQS Public Reporting		
DDW	EAR - Drinking Water Electronic Annual Report	DRINKING	6
DDW	ELAP - Environmental Laboratory Accreditation Program	ELAP	1
DWR	eWRIMS - Water Rights Information Management System Core	WATER RIGHTS	17
DFA	FAAST - Financial Assistance Application Submittal Tool	LOANS AND GRANTS	1
DWQ	GAMA	GROUNDWATER MONITORING	1
DWQ	Geo Tracker	GROUNDWATER CLEANUP	1
DIT	GeoWQA	SPATIAL	1
ORPP	Groundwater Extraction and Reporting System (GEARS)	SIGMA	1
DWQ	Integrated Report Upload Portal (IR Portal)-Survey Wizard	TMDL/BASIN PLANNING	1
<b>Total</b>			<b>106</b>

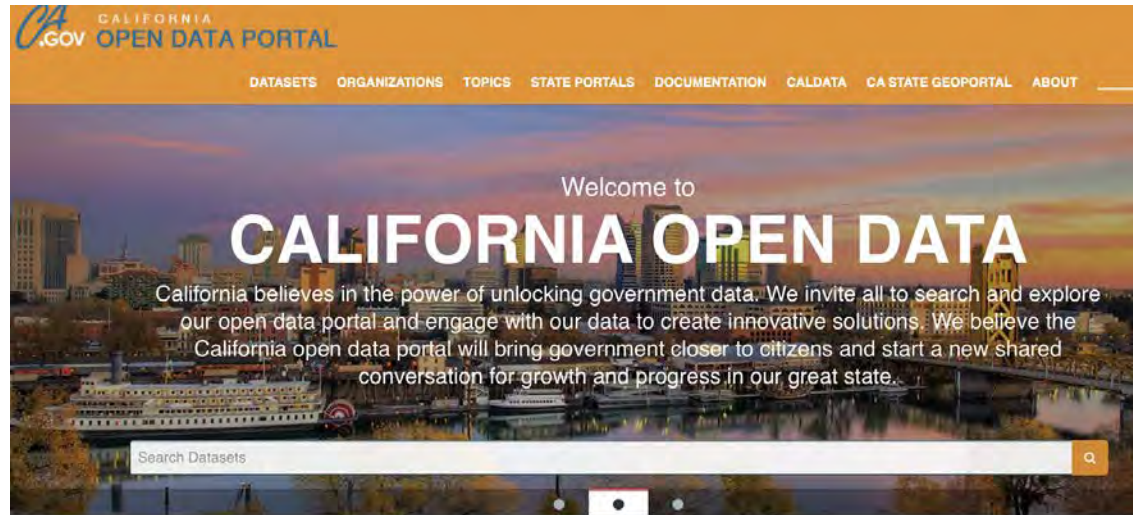
### Count of Dataset by Division and Data System



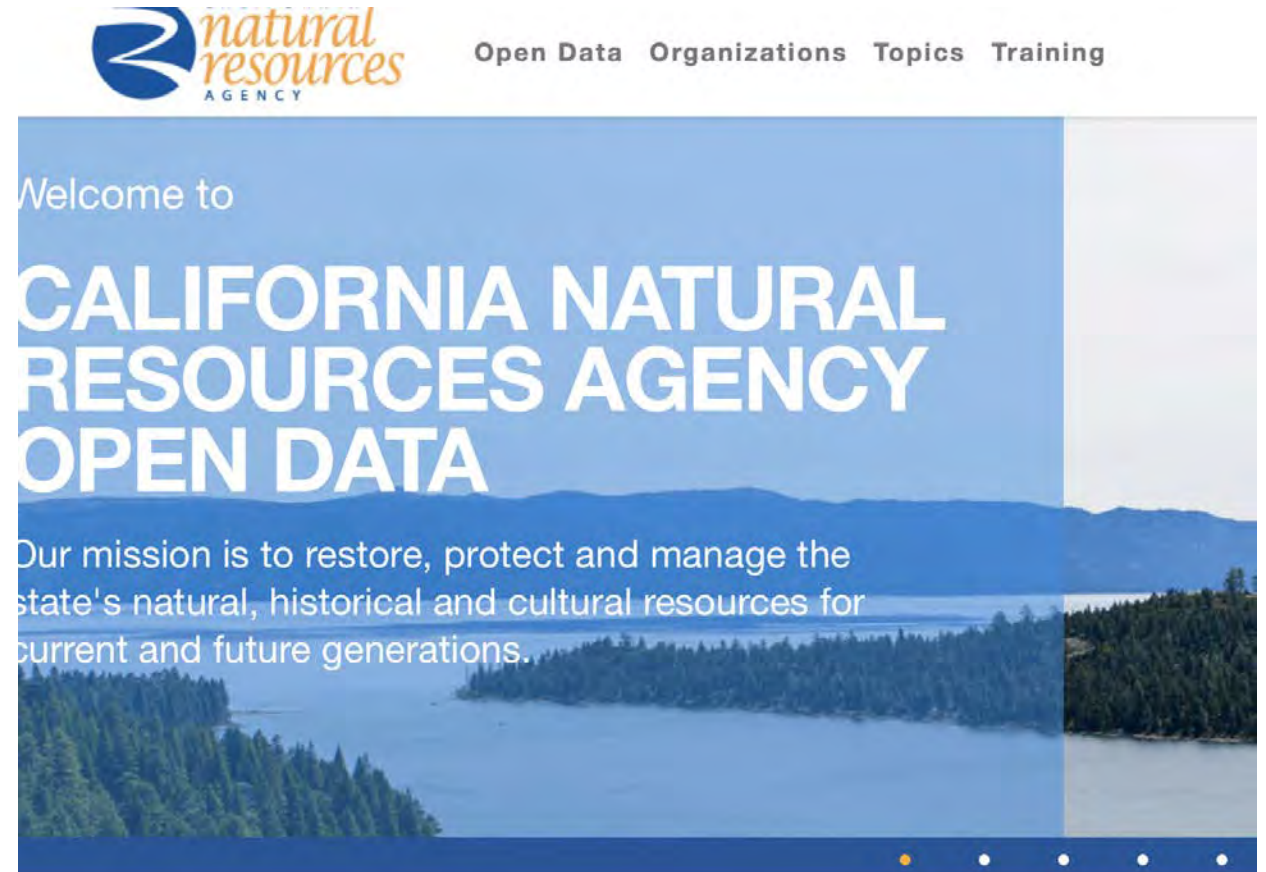
# Improve Data Flows → Open Data

- Enterprise “modernization” and “virtualization” efforts
- Tools to automate cleaning and loading data
- Open Data Portal(s)
  - [data.ca.gov](http://data.ca.gov) and [data.cnra.ca.gov](http://data.cnra.ca.gov)
- Internal Databases and Single Datasets → published to portals
- Metadata, naming conventions, discoverability, accessibility issues

# Two Open Data Portals in CA for Water Data



## Showcases



Get started by searching from **19200** datasets and maps from across California

## Topics

# CA Open Water Data Strategic Plan

## Goals:

1. Data are Sufficient
2. Data are Accessible
3. Data are Useful
4. Data are Used

Strategic Action Projects → focused on final statutory deadline of “federate with federal” datasets.

<https://water.ca.gov/ab1755>

# Federation Strategy Concept Paper (Draft)

- Stage 1 - Discoverability: Single source to search inventory
- Stage 2 - Discoverability: Open data platforms sharing inventories (no standard for metadata)
- Stage 3 - Discoverability: Tools assisting inventory discoverability
- Stage 4 - Discoverability: Open data platforms sharing inventories (using common standard for metadata)
- Stage 5 - Interoperability: Open data platforms making federal inventory discoverable (using common standard for metadata)
- Stage 6 - Interoperability: Optimize discoverability of inventory on other interfaces and platforms
- Stage 7 - Interoperability: Limited (pilot) inventory is interoperable and presented as new, integrated inventory
- Stage 8 - Interoperability: Limited (pilot) inventory is intrinsically interoperable
- Stage 9 - Interoperability: All core inventory is published as interoperable intrinsically

# Foster Engagement With Data

Encourage data use and internal / external collaboration

# Foster Engagement With Data

## Community / Civic Engagement

- Water Board Data Fair – March every other year
- Watershed Health Indicator and Data Science Symposium – June every year
- California Water Data Challenges – Every year (most recent trash and drinking water)
- “Local” hackathons and brigades

## Workforce Engagement

- WB Data Science Club
- CalData
- Recruitment Fairs / College Embed



A screenshot of the CalData website. The top navigation bar is blue with white text for "Submit Your Event", "About", "Contact Us", and "Settings". Below this is a white header with "Organization Title" on the left and navigation links for "Upcoming Events", "Past Events", "Contact", "Past Events Summary", and "Search" on the right. The main content area has a blue heading: "Bringing California government and people together to create solutions." followed by the sub-heading: "Engage with the State of California and help us find new ways of solving our government's most pressing challenges." Below this are three event cards. The first card is for the "March 30, 2018 - Water Board Data Fair and Data Summit" with a blue water drop icon containing a bar chart. The second card is for the "July 28 - 30, 2017 - Food and Agriculture Hackathon in Davis/Sacramento in July with a \$10k first prize" with a photo of people in a field. The third card is for the "June 29 - 30, 2017 - Annual Water Quality Health Indicator and Data Science Symposium" with a photo of a river and mountains. Each card includes a "read more" link at the bottom.





**Congratulations to the 2019 California Water Data Challenge participants**

#CAWATERDATA CHALLENGE

# California SAFE

## DRINKING WATER data challenge

the **POWER** of OPEN data...

\$1500 PRIZE for the MOST INNOVATIVE & IMPACTFUL SOLUTION

What are we GOING to LEAVE to OUR CHILDREN & GRANDCHILDREN if WE DON'T TAKE CARE of OUR DRINKING WATER

SUMMER 2018 Join now!

What is a DATA Challenge  
an OPPORTUNITY for INDIVIDUALS & TEAM-BASED Problem Solving

LEVERAGE DATA & TALENT to PROVIDE KNOWLEDGE & INSIGHTS

DIGITAL & MACHINE READABLE DATA MAKES more INNOVATIVE SOLUTIONS POSSIBLE!

BRINGING Talent to BEAR for INNOVATIVE SOLUTIONS for SAFE DRINKING WATER...



20 new datasets in machine-readable format

- Human Right to Water - Drinking Water Enforcement
- Monthly Water Production Reported by Water Suppliers in the Electronic Annual Reports (2011-2015)
- Drinking Water Watch - Public Water Systems Information
- Disadvantaged Communities Mapping and Land Use by Parcel

...in addition to over 1,000 datasets to explore!

JUNE 26, 2018  
GRAPHICS by steph@10xcollective.com @STEPHSCRIBES

# Foster Engagement With Data

## Sharing and Collaborating

- Organizational GitHub
- Open source tools / projects

The screenshot shows the GitHub profile for the California Water Board Data Center. At the top, it displays the organization's name and a brief description: "Open source resources from the California Water Resources Control Board and our partners. For more information or to join, contact waterdata@waterboards.ca.gov". Below this, there are navigation tabs for "Repositories" (23), "Packages", "People" (4), and "Projects". A prominent banner encourages users to "Grow your team on GitHub" with a "Sign up" button. Below the banner is a search bar for repositories and filters for "Type" and "Language". Two repository cards are visible: "Cannabis-Regional-Instream-Policies-" and "Datathon-Resources". The "Datathon-Resources" card includes a description: "Catalog of resources related to datathons which the CA Water Board sponsors or collaborates on. The site is at: https://cawaterboarddatacenter.github.io/Datathon-Resources/".

Datathon Resources

Home

Trash Projects

2019 Drinking Water Projects

Other Topics

## Datathon Resource Repository

This site is a catalog of projects and datasets relevant to datathons and similar events sponsored by or held in collaboration with the [California State Water Resources Control Board](#). Click on any of the tabs at the top of this page to find resources related to that topic area.

This site is maintained by the Water Board's Office of Information Management and Analysis, and will be continually updated as new resources are made available. If you'd like to suggest a resource to add, please send a message to [waterdata@waterboards.ca.gov](mailto:waterdata@waterboards.ca.gov).



# Data Driven Management

Water Quality Status Reports (2017, 2018 and 2019):

- [https://www.waterboards.ca.gov/resources/data\\_databases/wq\\_status\\_report.html](https://www.waterboards.ca.gov/resources/data_databases/wq_status_report.html)

Annual Performance Report

- [https://www.waterboards.ca.gov/about\\_us/performance\\_report\\_1819/index.html](https://www.waterboards.ca.gov/about_us/performance_report_1819/index.html)

FY2016/17 Water Boards Performance Report Story:

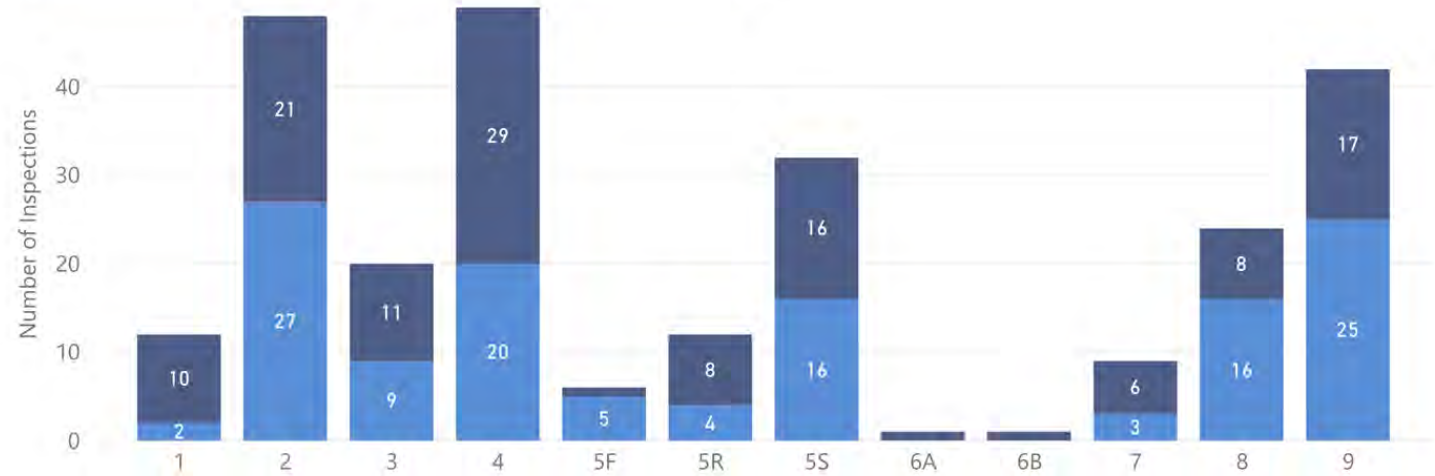
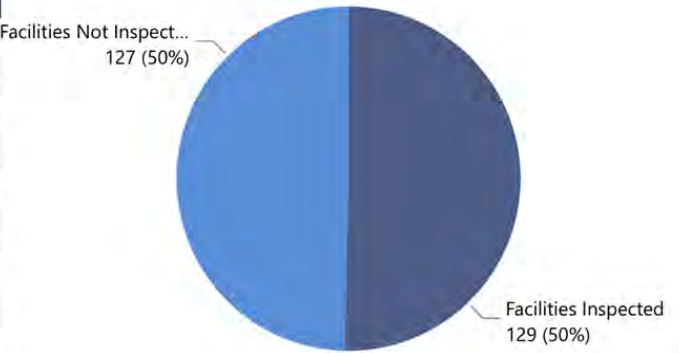
- <https://arcg.is/z5Km1>

# New Performance Report - Open Data Based

## Measurements

[View Data](#)

RB	Facilities	Inspections	Facilities Inspected	Percentage of Facilities Inspected
1	12	11	10	83%
2	48	21	21	44%
3	20	13	11	55%
4	49	31	29	59%
5	50	26	25	50%
6	2	2	2	100%
7	9	6	6	67%
8	24	9	8	33%
9	42	17	17	40%
<b>Total</b>	<b>256</b>	<b>136</b>	<b>129</b>	<b>50%</b>



# Water Quality Report Card - Algae in the Ventura River

<b>Regional Water Board:</b>	Los Angeles, Region 4	<b>STATUS</b>	<input type="checkbox"/> Conditions Improving		
<b>Beneficial Uses Affected:</b>	REC-1, REC-2, WARM, COLD, EST, WILD, RARE, MIGR, SPWN, WET, MUN		<input type="checkbox"/> Data Inconclusive	<input checked="" type="checkbox"/> Improvement Needed	
<b>Implemented Through:</b>	NDPES Permits, MS4 Permits, Conditional Waivers	<b>Pollutant Type:</b>	<input checked="" type="checkbox"/> Point Source	<input checked="" type="checkbox"/> Nonpoint Source	<input type="checkbox"/> Legacy
			<b>Pollutant Source:</b>	Urban Storm Water Runoff	Irrigated Crop Production
<b>Effective Date:</b>	June 28, 2013		Onsite Wastewater Treatment Systems	Wastewater Discharges	
<b>Attainment Date:</b>	2023		Horses and Livestock	Non-Point Source Runoff	

## Water Quality Improvement Strategy

The Ventura River watershed is in Ventura and Santa Barbara Counties in Southern California. The Ventura River, including its estuary and tributaries, is impaired due to algae, eutrophic conditions, low dissolved oxygen, and elevated nitrogen. The primary sources of these impairments are nutrients discharged from the municipal separate storm sewer system (MS4), agriculture operations, livestock facilities, onsite wastewater treatment systems (OWTS), and the Ojai Valley Waste Water Treatment Plant (WWTP). In 2013, USEPA approved the [TMDL for Algae, Eutrophic Conditions, and Nutrients in the Ventura River and its Tributaries](#) to restore water quality. The TMDL includes numeric targets for algal biomass, dissolved oxygen, and pH, and load allocations (LAs) and waste load allocations (WLAs) for total nitrogen and total phosphorus. The TMDL assigns more stringent nitrogen and phosphorus allocations for dry weather than wet weather because dry weather (May 1 to September 30) is the growing season. The TMDL allows the Ojai WWTP 12 years, MS4 permittees six years, agriculture operations six years, livestock facilities 10 years, and OWTS 10 years to attain allocations. The Ojai WWTP intends to attain WLAs by upgrading its nutrient removal processes. Agriculture operations will implement iterative management practices to control nutrients in their discharges. The MS4 permittees' compliance approach is to eliminate dry-weather discharges by implementing best management practices (BMPs). Horse facilities will implement manure management plans. Individual responsible parties are monitoring their discharges to demonstrate compliance with allocations and multiple responsible parties are jointly monitoring algal biomass, nutrients, and other constituents in receiving waters to assess watershed-wide conditions. The Board intends to adopt a Conditional Waiver for horse facilities in FY 18-19. Agriculture operations will implement nutrient management as required by the Conditional Waiver.

## Comparison of MS4 Effluent to Dry Weather WLA



Blank Total Nitrogen values, as seen for Happy Valley outfall after 2012, are due to zero flow and represent WLA attainment.



## Water Quality Outcomes

- Monitoring data show that algal biomass continues to exceed the numeric target. Total nitrogen in MS4 outfalls exceeds the WLA when there is sufficient flow to sample. However, no flow and no sample in the outfalls amounts to WLA attainment.
- WLAs have not been incorporated into the MS4 permits, but permittees are implementing BMPs, including a bioswale at the Happy Valley outfall in Reach 4, which has reduced dry-weather flow.
- The Ojai WWTP is on schedule to implement the nitrogen removal upgrades required by its permit to attain the WLAs. Ventura County is studying which OWTS will be upgraded to advanced treatment. The agriculture LAs are incorporated into a Conditional Waiver.
- The TMDL is still in the early stages of implementation. The multiple sources, complex interaction between groundwater and surface water, and variable flow make this a complicated TMDL.
- Responsible parties will continue implementation actions.



# Water Quality Report Card- Pesticides in the Palo Verde Outfall Drain and Lagoon

## Regional Water Board -Colorado River Basin, Region 7

**Beneficial Uses Affected** - Contact Water Recreation (REC-1); Non-Contact Water Recreation (REC- 2); Warm Freshwater Habitat (WARM); Wildlife Habitat (WILD); Rare, Threatened, Endangered Species (RARE)

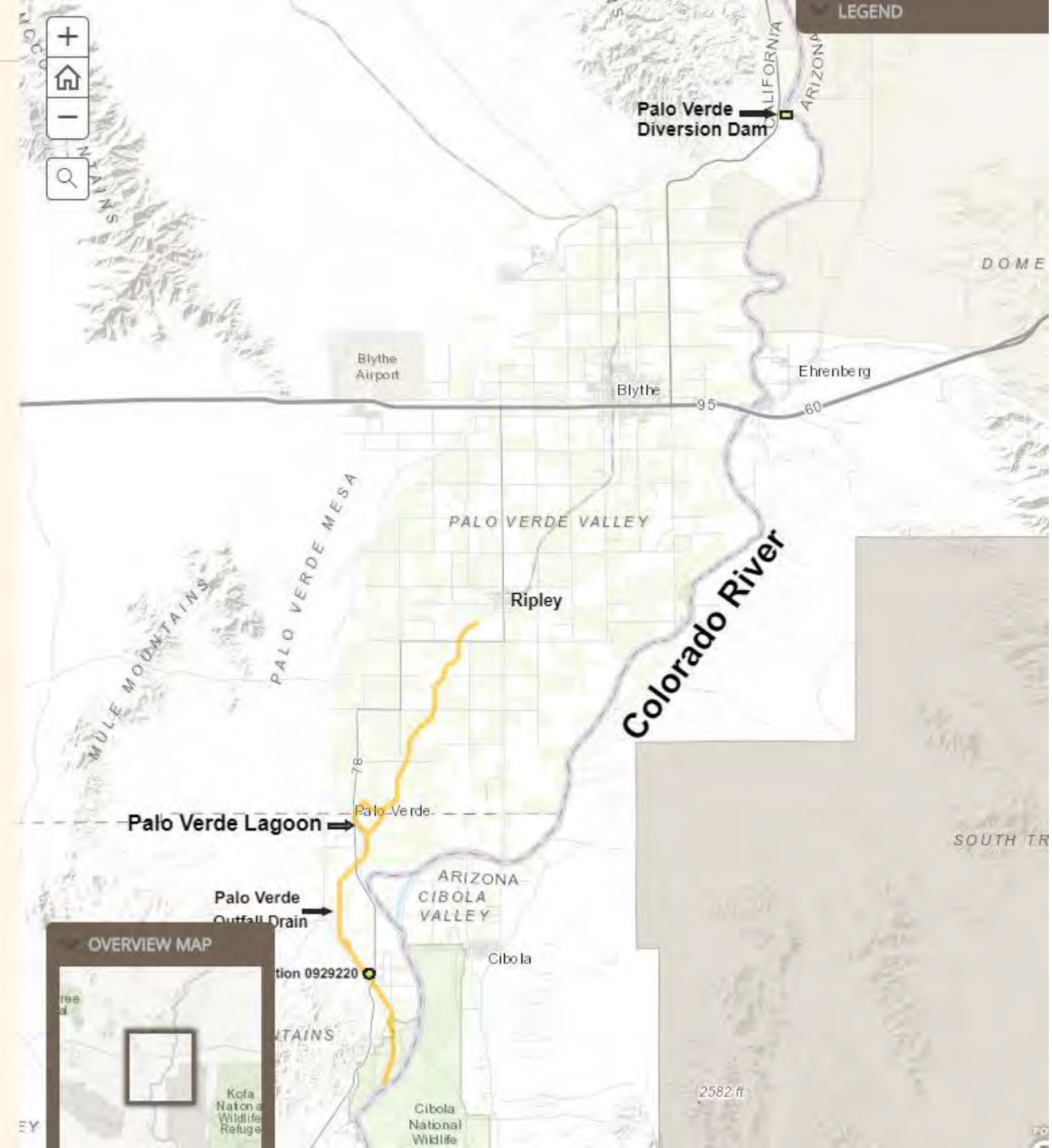
**Implemented Through** - Conditional Waiver of Waste Discharge Requirements (WDR)

**Pollutant Type:** Non-Point Source, Legacy **Pollutant Source:** Irrigated Crop Production

**Status:** Data Inconclusive

**Effective date:** September 20, 2012 **Attainment Date:** 2036

The Palo Verde Outfall Drain (PVOD) and Lagoon is located in Palo Verde Valley and Mesa (approximately 131,000 acres of agricultural land) in Imperial and Riverside counties. Palo Verde Outfall Drain and Lagoon are impaired by the legacy pesticides, Dichloro-Diphenyl-Trichloroethane (DDT) and Toxaphene and listed on the



# Water Quality Report Card - North Coastal Basin Rivers Cyanobacteria



## Regional Water Board - North Coast Region, Region 1

**Beneficial Uses Affected** - Cold Freshwater Habitat (COLD); Rare, Threatened, and Endangered Species (RARE); Migration of Aquatic Organisms (MIGR); Spawning, Reproduction, and/or Early Development (SPWN); Commercial and Sport Fishing (COMM); Tribal Tradition and Culture (CUL) Contact Water Recreation (REC-1); Non-Contact Water Recreation (REC-2)

**Implemented Through** - Restoration, Coordination efforts

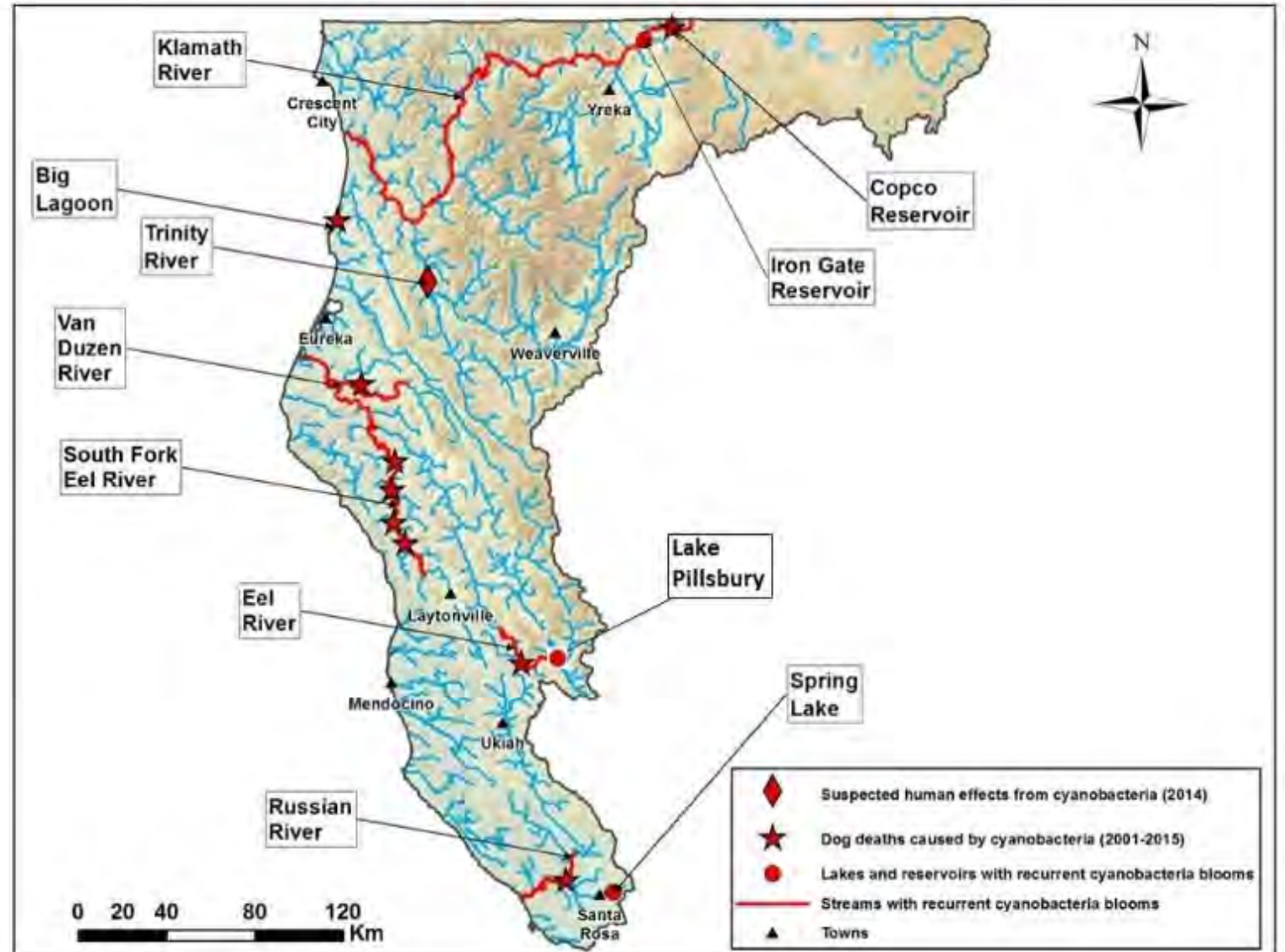
**Pollutant Type:** Non-Point Source, Legacy

**Pollutant Source:** Irrigated Crop Production, Hydromodification, Non-Point Source Run-off, Naturally Occurring, Logging, Grazing

**Status:** - **Improvement Needed**

**Effective date** - December 28, 2012

**Attainment Date** - 2050 or longer





# Build Capacity

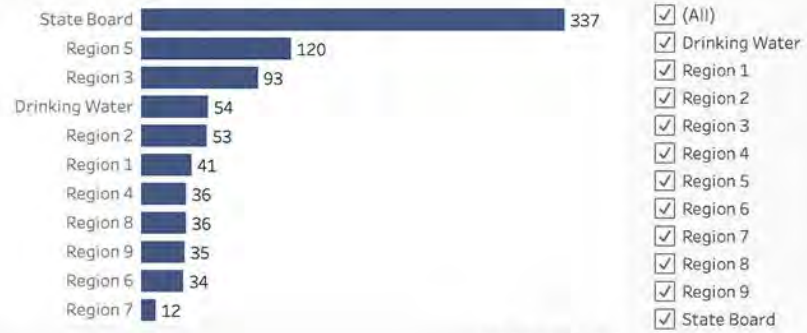
Improve Data Flows / Focus on Workforce and Workflows  
Enhance organization-wide 'data literacy'

# Workforce Survey on Data Literacy

## 2019 Water Boards Data Use and Skills Survey

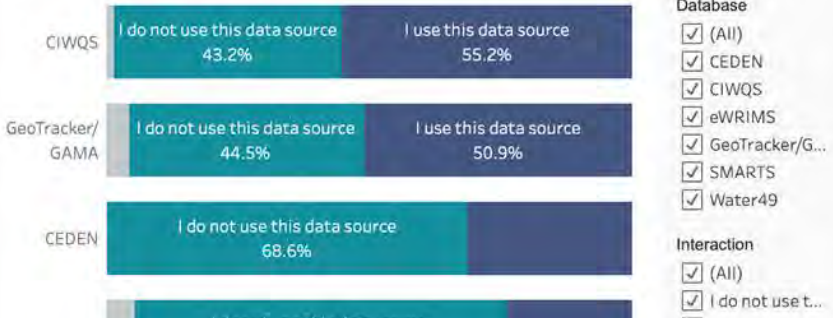


### Office/Location

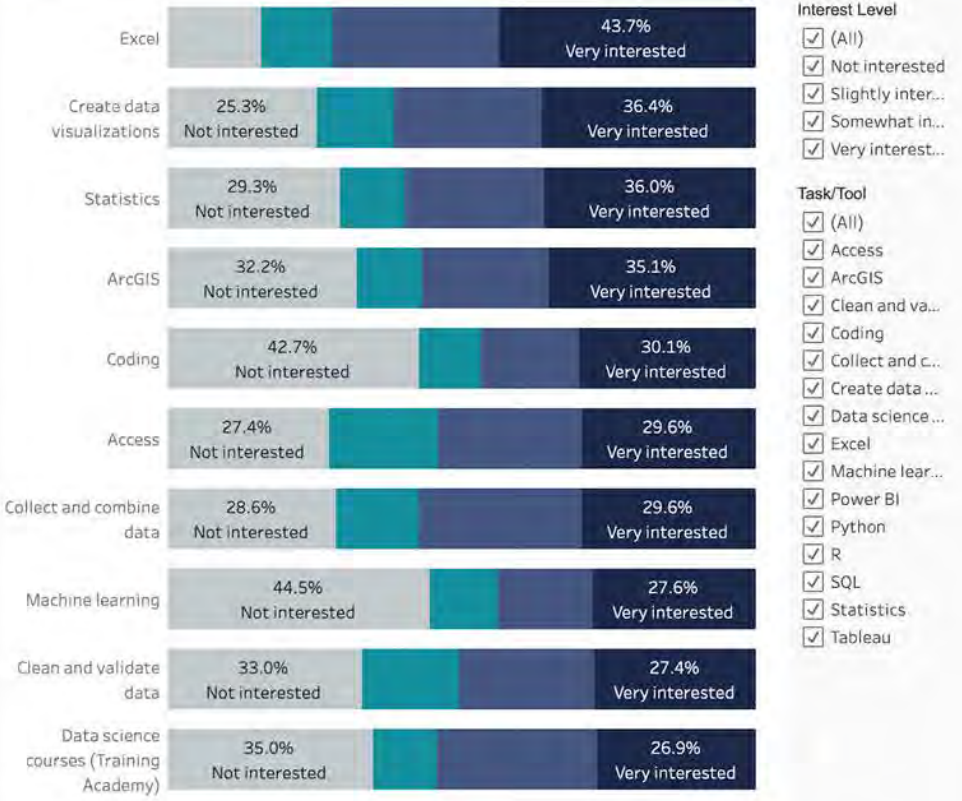


Total: 851

### Database/Data Source



### Interest



# Build Capacity

Pilot assessment and training program:

- Reinforce basic concepts (how to engage with data)
- Apply best practices throughout the data life cycle (collecting, storing, managing)

Tuva K-12 BUSINESS SUSTAINABLE DEVELOPMENT

## Tuva's Data Literacy Training Solutions



### Assessments to Measure Strengths & Weaknesses

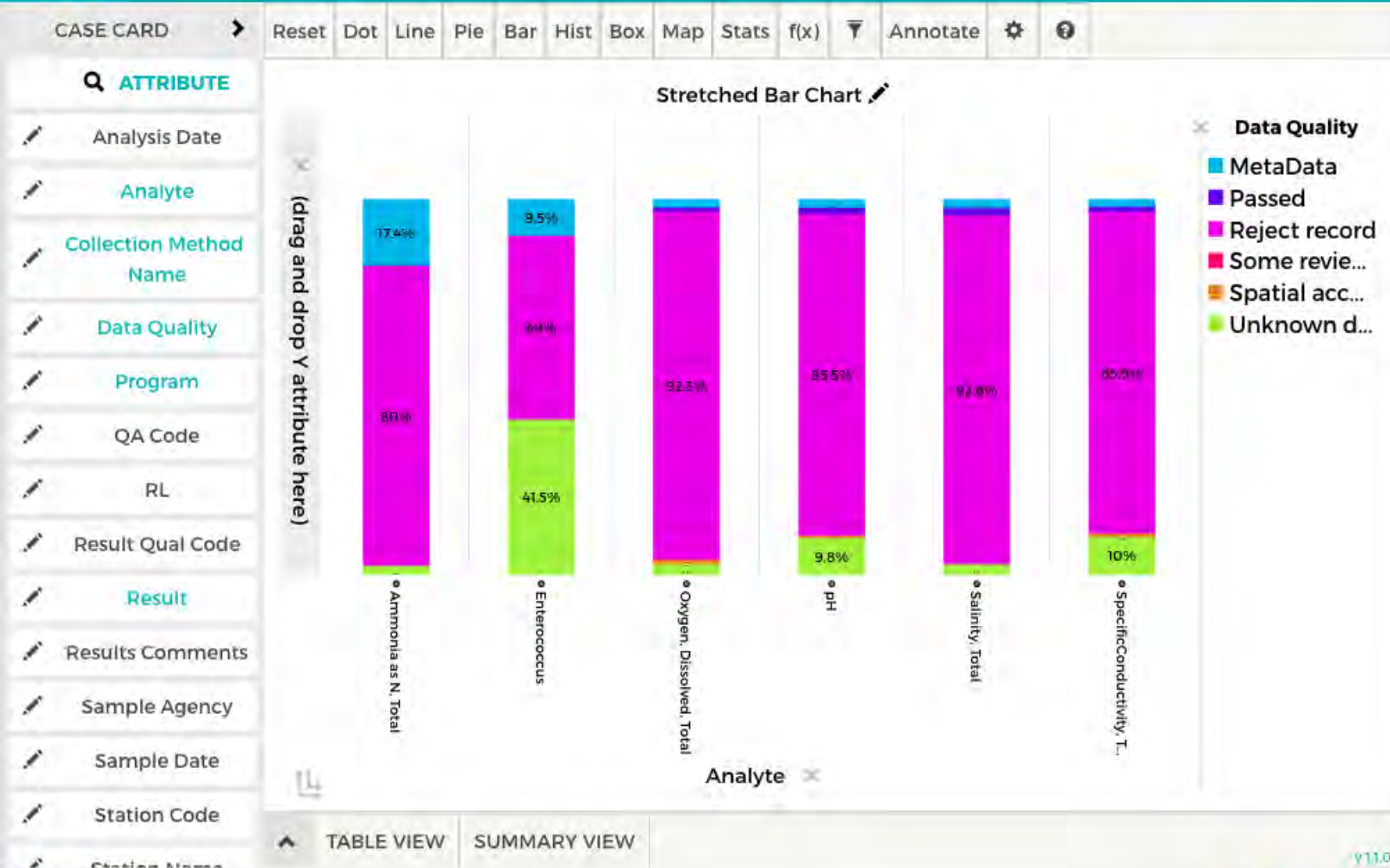
Assess Candidates for:

- Statistical Thinking & Reasoning
- Analytical Thinking & Evidence-Based Decision Making
- Data & Statistics Literacy



### Data & Statistics Literacy L&D Program

- A personalized learning pathway to build a strong data and statistics foundation
- Learn from wide range of authentic lessons and business case studies
- Comprehensive Learning Analytics & Reporting



## CA Water Boards - Data Literacy Assessment (10 of 23)

### Question 8:

Based on the plot on the left, which analyte did not have any samples that passed the Data Quality check?

- A.  Ammonia as N, Total
- B.  Enterococcus
- C.  Oxygen, Dissolved, Total
- D.  Salinity, Total
- E.  SpecificConductivity, Total

↑ ↓ CASE CARD 1 of 3403

Reset Dot Line Pie Bar Hist Box Map Stats f(x) Annotate

ATTRIBUTE	VALUE
Analysis Date	
Analyte	pH
Collection Method Name	Field Method
Data Quality	Reject record
Program	National Pollutant Discharge Elimination System (NPDES)
QA Code	None
RL	
Result Qual Code	=
Result	7.48
Results Comments	

**Histogram**

TABLE VIEW SUMMARY VIEW

## CA Water Boards - Data Literacy Assessment (9 of 23)

Let us focus our exploration on one specific analyte - Ammonia.

- Click on icon next to **Analyte** to access its metadata
- Unselect all the analytes except Ammonia as N. Total to keep only the ammonia samples in our dataset
- Now drag and drop **Result** on the x-axis
- Click on **Histogram** so that we can better view the distribution of the results
- From **Stats**, select **Mean** & **Median** so that it is displayed in our plot

### Question 7:

You will notice that the Mean is larger than the

↑ ↓ CASE CARD 1 of 3403

Reset Dot Line Pie Bar Hist Box Map Stats f(x) Annotate

ATTRIBUTE	VALUE
Analysis Date	
Analyte	pH
Collection Method Name	Field Method
Data Quality	Reject record
Program	National Pollutant Discharge Elimination System (NPDES)
QA Code	None
RL	
Result Qual Code	=
Result	7.48
Results Comments	

Dot plot

TABLE VIEW SUMMARY VIEW

## CA Water Boards - Data Literacy Assessment (7 of 23)

The dataset on the left contains 3403 samples from chemistry and field analyses of the Tijuana River.

- Drag  to the y-axis
- Drag  to the x-axis

### Question 5:

Based on the plot, which of the following analytes is measured in more than one unit.

- A.  Ammonia as N, Total
- B.  Oxygen, Dissolved, Total
- C.  pH
- D.  Salinity, Total



# Questions?

**Greg Gearheart, Deputy Director**  
[greg.gearheart@waterboards.ca.gov](mailto:greg.gearheart@waterboards.ca.gov)  
Office of Information Management and Analysis  
<https://www.waterboards.ca.gov/resources/oima/>