Open Data Resolution

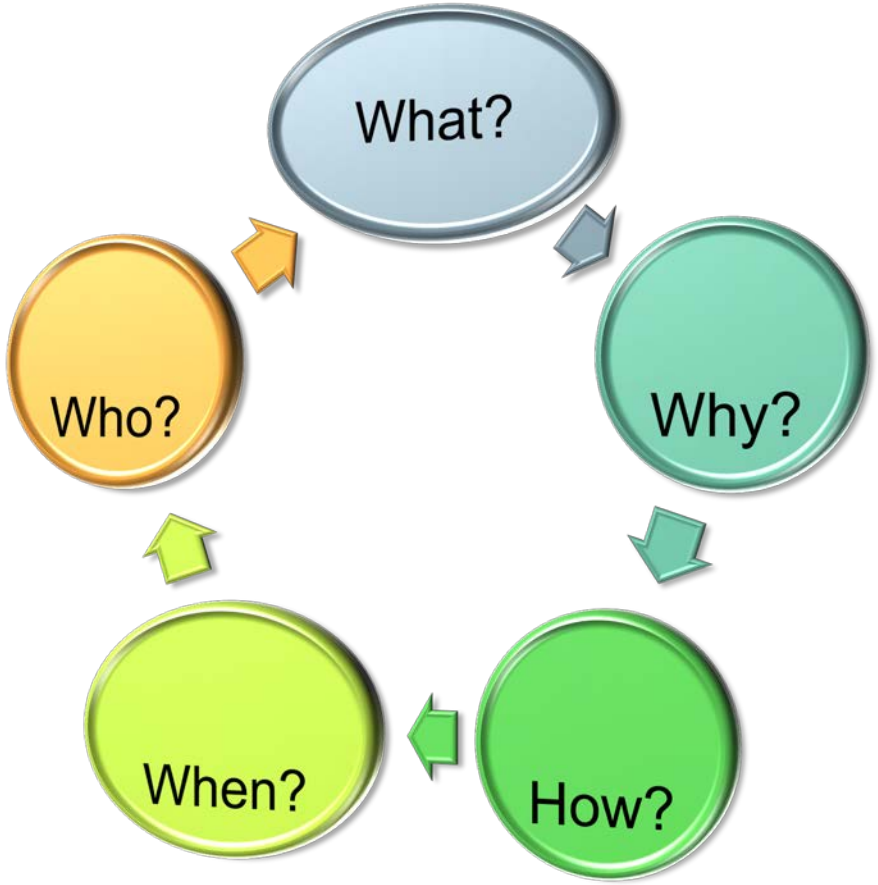
and Water Board Data Management

October 25, 2018. Brown-bag

Rafael Maestu

Office of Information Management and Analysis

State Water Resources Control Board / CalEPA


What?
Why?
How?
When?
Who?

SWAMP
Surface Water
Ambient Monitoring
Program

Why?

CALIFORNIA
Water Boards
STATE WATER RESOURCES CONTROL BOARD
REGIONAL WATER QUALITY CONTROL BOARDS

THE SACRAMENTO BEE

WATER & DROUGHT

360,000 Californians have unsafe drinking water. Are you one of them?

BY DALE KASLER, PHILLIP REESE AND RYAN SABALOW

June 01, 2018 03:55 AM
Updated June 01, 2018 04:36 PM

At the Shiloh elementary school near Modesto, drinking fountains sit abandoned, covered in clear plastic.

At Mom and Pop's Diner, a fixture in the Merced County town of Dos Palos, regulars ask for bottled water because they know better than to consume what comes out of the tap.

SUBSCRIBE ROUTE FIFTY
MANAGEMENT INFRASTRUCTURE TECH & DATA SMART CITIES PUBLIC SAFETY HEALTH & HU

A Moneyball Approach to Clean Water in California

Stanford | Water in the West
PROGRAMS PUBLICAT
News & Events / News & Insights

Assembly Bill 1755 and the Politics of California Water Data

July 31, 2018 | Water in the West | Insights
By Zachary Sugg

Each year millions of us collect receipts, financial statements and a host of other documents required to file our taxes. Some receipts are paper, some are digital, some files we have, while others are sent to us by employers, and for many, figuring out where things are is an arduous process. Accessing

Water Boards

Media Release

State Water Board Adopts Open Data Resolution

Commits to Improved Access to Water Data that Informs Regulatory and Operational Decisions on the Future of Water in California

For Release:
July 10, 2018

Contact: George Kostyrko
george.kostyrko@waterboards.ca.gov

SACRAMENTO – The State Water Resources Control Board adopted a resolution today on open data principles committing it and the Regional Water Boards to providing broader access to the data used to make local, regional and statewide water management and regulatory decisions in California.

Why?

Effective water resource management demands that California better integrate existing water and ecological data into an authoritative open-access platform to help water managers operate the State of California's (State) water system more effectively and help water users make informed decisions based on water availability and allocation.

Why?

Why do we collect, store and use data?

1. To inform our **data-driven management** and planning activities – performance report cards, workplans, resource assignment, *evaluating program effectiveness*, and many others examples;
2. To inform our **critical decisions** regarding our mission(s) and water management responsibilities – water allocation and use, water quality planning and "policies," permitting, program prioritization, etc.; and
3. To provide **transparency** to our many partners and stakeholders for their use, interests and purposes.

Why?

The Open and Transparent Water Data Act (AB1755)

September 23, 2016

Why?

The Open and Transparent Water Data Act

CWC § 12401 The Legislature finds and declares all of the following:

(a) The recent drought reveals that California needs to integrate existing water and ecological data into an authoritative open-access platform to help water managers operate California's water system more effectively and help water users make informed decisions based on water availability and allocation.

(b) State and federal leadership, increased awareness by business, governmental, and nongovernmental organizations through open and transparent access to data, and improved technology and availability of open-source platforms create a unique opportunity that California should seize upon to integrate and increase access to existing water data.

(c) California is working to increase access to water data collected by state agencies. The state board is piloting a project to make water quality datasets available online through an open data portal. The portal creates an opportunity to foster collaboration among state agencies, share and integrate existing datasets, improve state agency operations through data-driven decision making, and improve transparency and accountability.

(d) State agencies should promote openness and interoperability of water data. Making information accessible, discoverable, and usable by the public can foster entrepreneurship, innovation, and scientific discovery.

(e) Water data and information technology tools and applications developed and gathered using state funds should be made publicly accessible. State delegation of data management to contractors should not result in the public losing access to its own information.

(f) The availability of open-source tools makes it easier to access and explore water and ecological data and could facilitate the creation of an online integrated water data platform without the need to create an expensive new centralized database.

Why?

AB1755 Strategic Plan Goals

1. Data are Sufficient

   (enough data in the right format with the right quality)

2. Data are Accessible

   (everyone would have access to the data, and the desired data would be easily discoverable)

3. Data are Useful

   (the right data, measuring the right things, at the right time, interoperability, integration)

4. Data are Used (from data to information, visualization, analysis, communication..)

Why?

AB1755 requirements operative to WBs

- DWR is leading effort to build platform, develop protocols, etc.
- Water Board and WQ Monitoring Council is supporting DWR effort(s)
- Water Board must publish open data by September 2019. In addition, by August 2020 we must:
  - **12415.** The statewide integrated water data platform created pursuant to Section 12410 shall, at a minimum, do all of the following:
  - (a) Integrate existing water and ecological data information from multiple autonomous databases managed by federal, state, and local agencies and academia using consistent and standardized formats.
  - (b) Integrate the following datasets, as available:
  - (1) The department's information on State Water Project reservoir operations, groundwater use, groundwater levels, urban water use, and land use.
  - (2) The state board's data on water rights, water diversions, and water quality through California Environmental Data Exchange Network (CEDEN)

Why?

Goal 5 of the 2008 Water Board's Strategic Plan is to "Improve transparency and accountability by ensuring that Water Board goals and actions are clear and accessible, by demonstrating and explaining results achieved with respect to the goals and resources available, by enhancing and improving accessibility of data and information, and by encouraging the creation of organizations or cooperative agreements that advance this goal, such as establishment of a statewide water data institute."

Why?

- Water Boards are leading the State's efforts in transparency and accountability with the Performance Report .
- State Water Board has been publishing its core water datasets on the State's main open data portal.
- The releasing of open data can improve the effectiveness and efficiency of Water Board programs, improve public trust, facilitate conversations with stakeholders, and encourage participation.
- There is increasing demand for government data and information from the public and stakeholders.

How?

How do we currently manage our data?

- We **collect** some of it from our stakeholders, some we enter ourselves, some we acquire from other sources
- We **store** some of it on our own servers, some on the cloud, some is stored on desktop and other personal computers
- We **make it accessible** various ways, including public search forms and open data publishing
- We **use** it by transforming it into information via software tools like spreadsheet, GIS and business intelligence

Data life cycle management involves managing the flow of information from the initial identification of data needs through the stages of data collection, data storage, data accessibility, and the final process of turning data into information and knowledge. Data life cycle management often requires the development and/or enhancement of information systems to address various life cycle stages, including functions that ensure quality data are collected, managed, reformatted, and made accessible to users (including machine readability). The data and information systems (data ecosystem), that the State Water Board is a steward of, is part of the critical infrastructure needed by all Californians to best address the State's current and future water priorities.

Stage 1 - Collecting Data

Data

Data Collection

Standards

Data Checkers

S2 - Storing Data

Databases

Web Services

S3 - Publishing Open Data

CalData data.ca.gov Portal

Water Rights Datasets

Water Quality Datasets

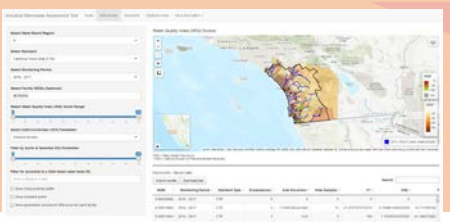
Other Open Data

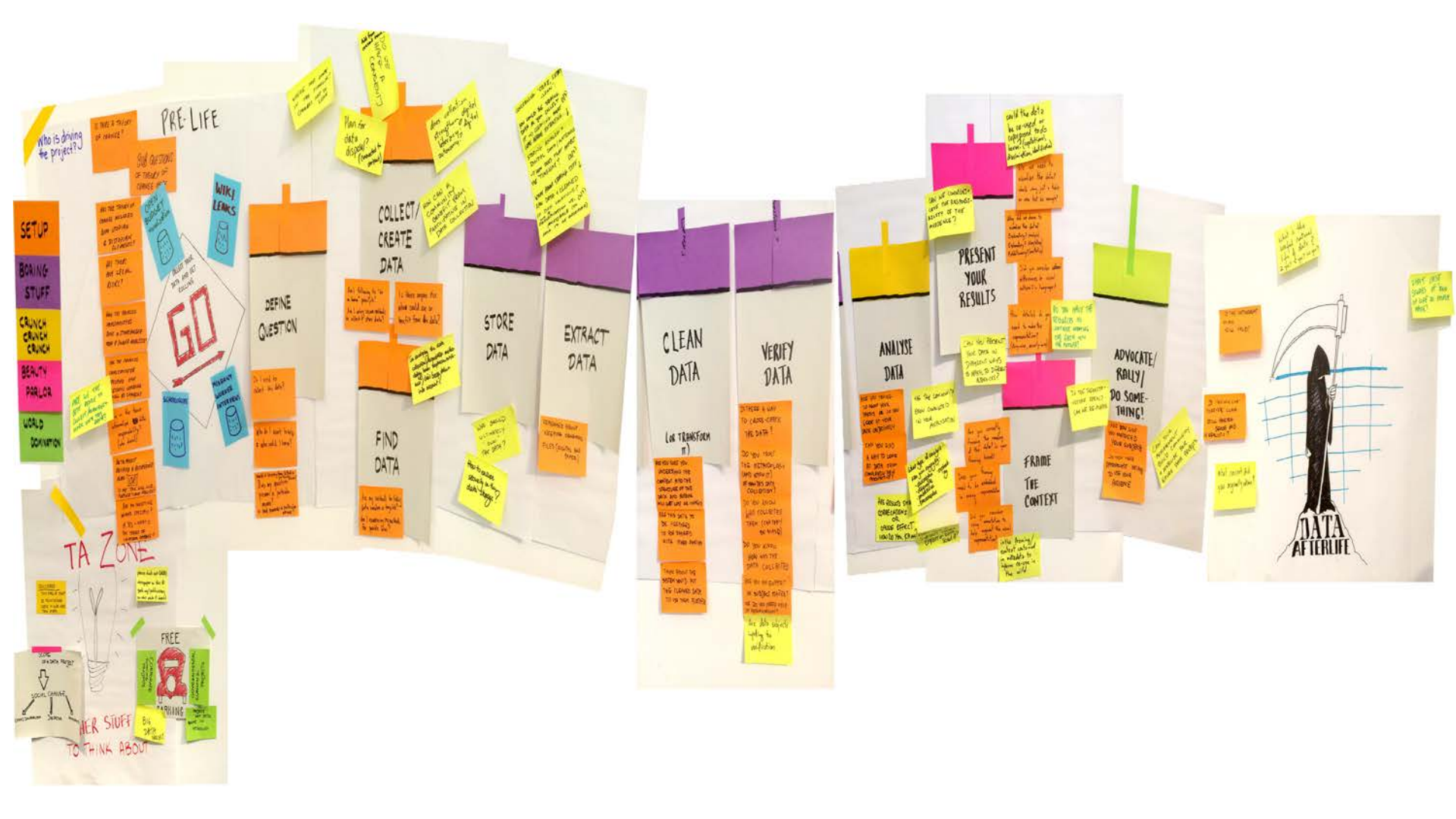
Federal

CA DWR

APIs / Web Services

S4 - Turning Data into Information




PRE-LIFE
Who is driving the project?
SETUP
BORING STUFF
CRUNCH CRUNCH CRUNCH
BEAUTY PARLOR
WORLD DOMINATION
GO
WIKI LEAKS
DEFINE QUESTION
COLLECT/ CREATE DATA
FIND DATA
STORE DATA
EXTRACT DATA
CLEAN DATA
VERIFY DATA
ANALYSE DATA
PRESENT YOUR RESULTS
FRAME THE CONTEXT
ADVOCATE/ RALLY/ DO SOME-THING!
DATA AFTERLIFE
TA ZONE
FREE
PARKING
HER STUFF TO THINK ABOUT

5 Guiding Principles for WB Data Management

1. **Make Data Accessible ("Open First")**: our organization values transparency and strives to make all critical data available in machine readable datasets with metadata and data dictionaries
2. **Understand Data Quality and Integrity**: our data are thoughtfully planned, of known and useful quality, and we deploy practices to protect its integrity with standards and protocols
3. **Improve Data Literacy**: our whole organization understands its data needs and responsibilities, can speak the language of data science the staff and managers have robust data science capacity

5 Guiding Principles for WB Data Management (cont.)

4. **Use Data to Govern:** our organization uses data to govern, or makes decisions that are in the best interest of our mission(s)
5. **Govern our Data:** our organization takes proactive steps to develop effective data and information technology management practices to ensure our data flows to where it is needed in a timely manner while complying with our data sharing policies

What?

What is “Open Data?”

- Open data is a term used around the world for past 6+ years to describe specific test of data accessibility and interoperability
- Publicly available
- Machine readable
    - Two dimensional flat files and/or available via APIs and/or web services
    - Curated and structured in a manner that makes it accessible and useful
- Well documented
    - Metadata – a robust and structured guide to the data to help users find context, limitations, issues including data dictionary

Data Management Strategic Elements

1. Open Data Strategy
2. Data Literacy Strategy
3. Data Driven Management Strategy
4. Quality Management Strategy
5. Data Governance and Administration Strategy

Open Data Strategy

1. Open Data Handbook
2. Make “flat file” pipelines of raw, high value data from all our priority databases applications
3. Identify data stewards in every part of the organization with charge and capacity to curate datasets around org
4. Datasets published on open data platforms, machine readable, refreshed as often as “needed” and well documented
5. Services and tools (and training) to support needs to analyze and visualize data (e.g., Tableau, etc.)
6. Predictive analytics, machine learning, fun stuff!
7. Civic engagement projects
    a. Water Data Challenges, etc.

CALIFORNIA
OPEN DATA PORTAL

search

Datasets | Topics | Groups | About | State Portals | Civic Engagement | Documentation | Log in

Welcome to California Open Data

search

California believes in the power of unlocking government data. We invite all to search and explore our open data portal and engage with our data to create innovative solutions. We believe the California open data portal will bring government closer to citizens and start a new shared conversation for growth and progress in our great state.

Natural Resources

Health & Human Services

Water

Government

Economy and Demographics

Transportation

Partners

Air Resources Board
California Protected Areas
Department of Aging
Department of Child Support Services
Department of Conservation
Department of Water Resources
California Environmental Protection Agency
Health and Human Services Agency
Public Utilities Commission
State Water Resources Control Board
Department of Health Care Services
Department of Managed Health Care
Department of Rehabilitation
Department of State Hospitals
Emergency Medical Services Authority

CA.GOV

Safe Drinking Water Data Challenge

#CAWaterDataChallenge

Search

Data & Resources | The Challenge | Get Involved | FAQ | Contact

Photo by The California Department of Water Resources

1 2 3 4 5

WEST BIG DATA INNOVATION HUB | WATER FOUNDATION | GOVERNOR'S OFFICE OF PLANNING AND RESEARCH · STATE OF CALIFORNIA · | IMAGINE H2O | BAYAREA COUNCIL

2018 Safe Drinking Water Data Challenge

2018 California Safe Drinking Water Data Challenge

Imagine H2O will award $1,500 and select accelerator services to a team that best demonstrates Impact, Scalability, Innovation in their solution. More information and terms

California Safe Drinking Water Data Challenge

Launch Event June 26, 2018 Collaborators

California Safe Drinking Water Data Challenge

Submit your open data project by Oct. 1, 2018!

WEST BIG DATA INNOVATION HUB

WATER FOUNDATION

GOVERNOR'S OFFICE OF PLANNING AND RESEARCH · STATE OF CALIFORNIA

IMAGINE H2O

BAYAREA COUNCIL

waterchallenge.data.ca.gov

#CAWaterDataChallenge

Data Literacy Strategy

1. Training and capacity development
    a. Organization (culture, systems and access to business intelligence tools)
    b. Individual (staff vs. managers vs. executives)
    c. Blend of customized online, in person and in-house providers
2. Engagement (internal and external)
    a. Data concepts and terms in regular work products and conversations
    b. Data science club (for projects, sharing, etc.)
    c. Targeted projects to practice what is learned

Data Driven Management Strategy

- Migrate performance report cards to open data-based visualizations and interactive tools
  a. Program tools
  b. Executive data storytelling
  c. Critical path stuff
     1. Business intelligence software distributed to staff, managers and executives
     2. A storytelling platform
     3. Open data web services that work reliably

Data Governance and Administration

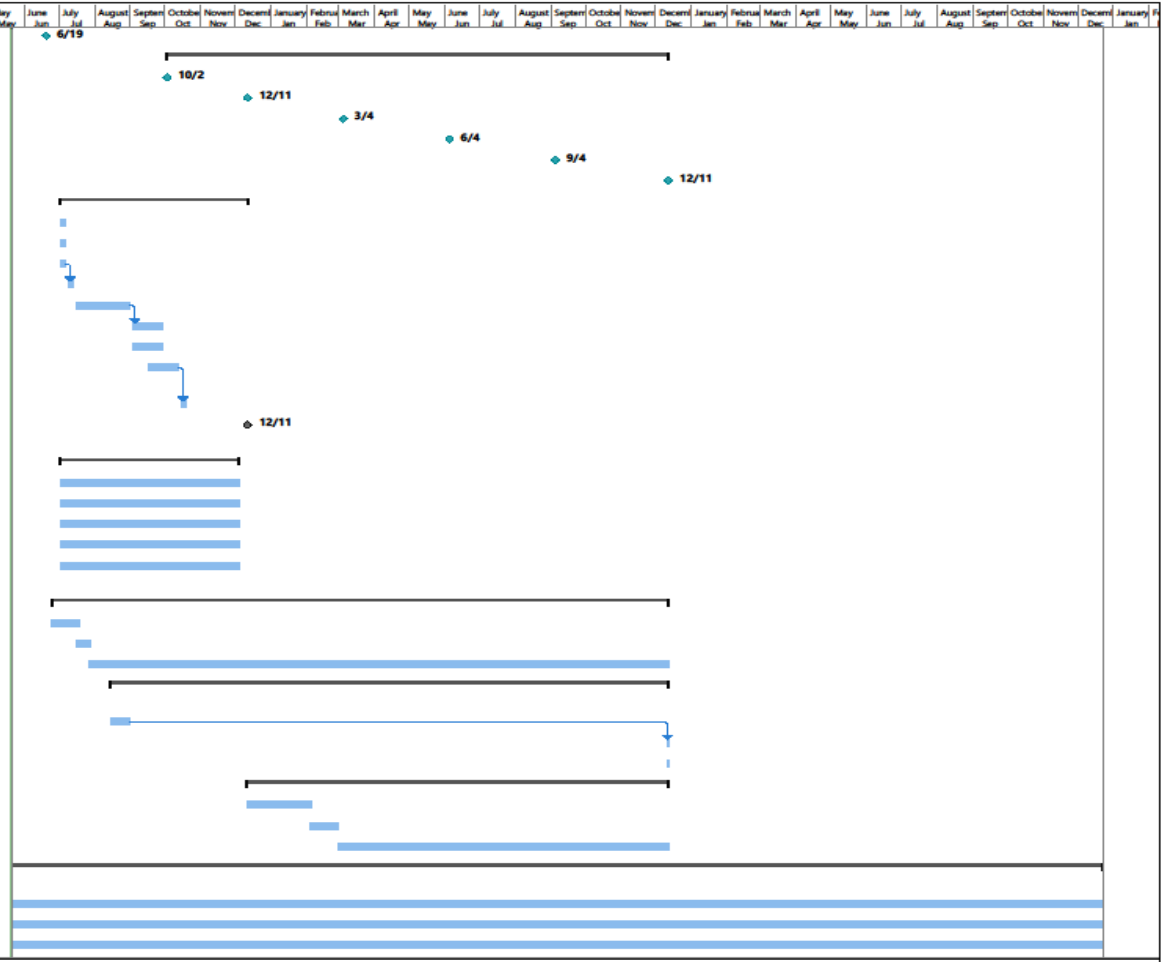
1. Governance - to address questions like:
    a. Who reviews and approves the open data datasets?
    b. What new data does the org need to do its job better?
2. Coordination with data collection and storage methods and approaches
    a. How we collect and store data to make it easier to flow towards flat files, open data, etc.
    b. When to use “spreadsheets” vs. building new DBs, etc.
3. Legal and privacy compliant practices in place

Actions and Next Steps

- Strategic review of database governance and delivery for our primary enterprise database applications.
- Development of a Data Management Strategy, including
  - An Open Data Strategy
  - A Data Literacy Strategy
  - A Data Driven Management Strategy
  - A Quality Management Strategy
  - A Data governance and Administration Strategy
- AB 1755 Compliance
- Procedures for Making Board Presentation Data Open
- Making Grant and Contract Data Open
- Developing Data Management Plan Resources
- Quarterly Progress Reports on Resolution Implementation

When?

ID	Task Mode	Task Name	Duration	Start	Finish
1		**Open Data Resolution Adopted**	1 day	Tue 6/19/18	Tue 6/19/18
2		**QUARTERLY REPORTS**	**312 days**	**Tue 10/2/18**	**Wed 12/11/19**
3		Q1 Report	1 day	Tue 10/2/18	Tue 10/2/18
4		Q2 Report	1 day	Tue 12/11/18	Tue 12/11/18
5		Q3 Report	1 day	Mon 3/4/19	Mon 3/4/19
6		Q4 Report	1 day	Tue 6/4/19	Tue 6/4/19
7		Q5 Report	1 day	Wed 9/4/19	Wed 9/4/19
8		Q6 Report	1 day	Wed 12/11/1	Wed 12/11/19
9		**Review Database Governance**	**117 days?**	**Mon 7/2/18**	**Tue 12/11/18**
10		Develop criteria for review	5 days?	Mon 7/2/18	Fri 7/6/18
11		Identify databases	5 days?	Mon 7/2/18	Fri 7/6/18
12		Identify data stewards	5 days?	Mon 7/2/18	Fri 7/6/18
13		Schedule interviews	5 days?	Mon 7/9/18	Fri 7/13/18
14		Conduct interviews/meetings	35 days?	Mon 7/16/18	Fri 8/31/18
15		Collect and analyze data	20 days?	Mon 9/3/18	Fri 9/28/18
16		Prepare inventory	20 days?	Mon 9/3/18	Fri 9/28/18
17		Write draft report and assessment of open data readiness	20 days?	Mon 9/17/18	Fri 10/12/18
18		Edit Final Report	5 days?	Mon 10/15/1	Fri 10/19/18
19		Presentation to Management and Board	1 day?	Tue 12/11/18	Tue 12/11/18
20		**Data Management Strategy**	**111 days?**	**Mon 7/2/18**	**Mon 12/3/18**
21		Open Data Strategy	111 days?	Mon 7/2/18	Mon 12/3/18
22		Data Quality Management Strategy	111 days	Mon 7/2/18	Mon 12/3/18
23		Training Program Data Literacy	111 days	Mon 7/2/18	Mon 12/3/18
24		Data Analysis and Communication	111 days	Mon 7/2/18	Mon 12/3/18
25		Database Coordination and Development Plan	111 days	Mon 7/2/18	Mon 12/3/18
26		**Data Management Plan**	**383 days?**	**Mon 6/25/18**	**Wed 12/11/19**
27		Develop Data Management Template	18 days?	Mon 6/25/18	Wed 7/18/18
28		Develop Training Materials	10 days?	Mon 7/16/18	Fri 7/27/18
29		Implementation and Coordination	359 days?	Fri 7/27/18	Wed 12/11/19
30		**Data Presentations to the Board**	**346 days?**	**Wed 8/15/18**	**Wed 12/11/19**
31		Prepare Memo and Instructions	13 days?	Wed 8/15/18	Fri 8/31/18
32		Coordinate with Boards Agenda	1 day?	Wed 12/11/1	Wed 12/11/19
33		OIMA Coordination and Follow Up	1 day	Wed 12/11/1	Wed 12/11/19
34		**Data from Contractors**	**262 days?**	**Tue 12/11/18**	**Wed 12/11/19**
35		Prepare Template Language	40 days?	Tue 12/11/18	Mon 2/4/19
36		Implement with Contracts Unit	19 days?	Mon 2/4/19	Thu 2/28/19
37		OIMA Follow up and coordination	205 days?	Thu 2/28/19	Wed 12/11/19
38		**OIMA Coordination With open Data Portals**	**677 days?**	**Mon 5/21/18**	**Tue 12/22/20**
39		Coordination with AB1755	677 days?	Mon 5/21/18	Tue 12/22/20
40		Coordination with Data.CA.Gov	677 days	Mon 5/21/18	Tue 12/22/20
41		Coordination with Monitoring Council	677 days	Mon 5/21/18	Tue 12/22/20



When?

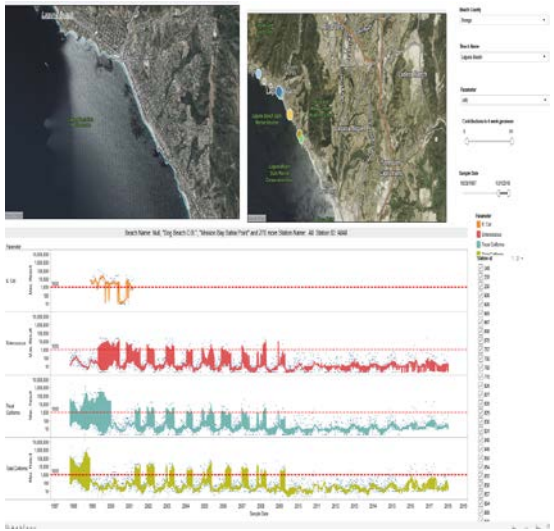
Key Dates

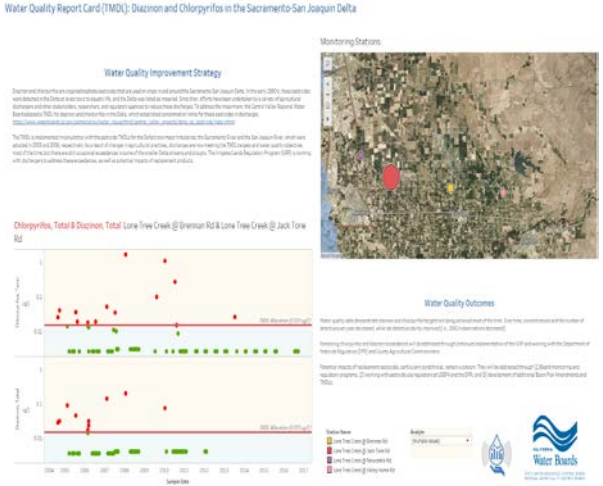
- Review and Assess Database Governance (by December, 2018)
- Data Management Strategy (March 1, 2019)
- Open Data for AB 1755 (September, 2019)
- Develop Data Management Plans (December, 2019)

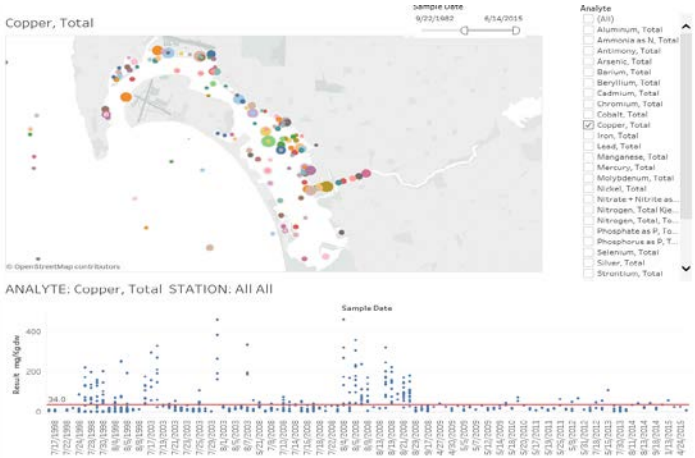
Who?

SWAMP
Surface Water
Ambient Monitoring
Program

Questions.. .. .. …? ? ? ?…







Rafael Maestu

Economist/Chief Data Scientist
State Water Resources Control Board
Office of Information Management and Analysis
California Environmental Protection Agency
1001 I Street, Sacramento, CA 95814
Ph: (916) 322-1838
Rafael.Maestu@waterboards.ca.gov