



California Water Boards
DATA CENTER



Open Water

State Water Resources Control Board
DRAFT Strategic Data Action Plan
(SDAP)

version 11012019



FOREWORD

Today, more than ever, pressures over water quality and water quantity require efficient, effective and sustainable water management. Using water information in water governance is necessary in order to achieve those goals. The Water Board actions would only be effective if they are consistent; if stakeholders are properly engaged; if well-designed regulatory frameworks are in place; if there is adequate and accessible information; and if there is sufficient capacity, integrity, and transparency.

Making informed decisions requires data to be useful and to be used. For data to be used it needs to be identified, collected, processed, qualified, stored, analyzed and communicated. This is what is known as data life cycle management. A data governance plan is needed to make proper data management decisions. There is an urgent need to improve the way the Water Boards manage data and information.

This document describes a set of strategic actions that the Water Boards could implement in the next five years that we believe would significantly improve the way we manage and use data and information for both internal facing operations and external facing decisions about California's valuable water resources.

The Data Management Strategy Development Team

- *Greg Gearheart, Deputy Director, Office of Information Management and Analysis (OIMA)*
- *Melissa Morris, Assistant Deputy Director, OIMA*
- *Rafael Maestu, Chief Data Scientist and Research Data Specialist III, OIMA*

VISION

In 2023 the Water Boards collect, store and use data of known and useful quality to make decisions and treat all water data as a public asset. Our data are useful, accessible, and interoperable; they inform all water stakeholders and fundamentally support our mission.

MISSION

The mission of the Data Management Team and the Office of Information Management and Analysis is to enhance the value, quality, understanding, and accessibility of Water Boards data through strategic activities, coordination, and changing the culture of data management at the agency.

GUIDING PRINCIPLES

Make Data Accessible (“Open First”): our organization values transparency and strives to make all critical public data available in machine-readable datasets with metadata and data dictionaries.

Understand Data Quality and Integrity: our data are of known and acceptable quality and we deploy practices to protect its integrity with standards and protocols.

Improve Data Literacy: our whole organization understands its data needs and responsibilities, can speak the language of data science, and staff and managers have robust data science capacity.

Use Data to Govern: our organization uses data to govern and makes decisions that are in the best interest of our mission(s).

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.

Table of Contents

Introduction	5
Legislative Directive	5
Water Board Directive.....	5
California Department of Technology Directive.....	7
Strategic Goals, Priorities, and Measures	8
Goal 1: Unlock Water Board’s Data:	8
Priority 1. Increase the availability of our data to the public	9
Priority 2. Improve the data flow architecture of our agency.....	9
Priority 3. Transform our data into usable formats	10
Goal 2: Value our Data	11
Priority 1. Improve the quality and integrity of our data	11
Priority 2. Enhance the use of data in our work.....	12
Priority 3. Effectively govern our data.....	12
Goal 3: Enhance our Data Culture.....	14
Priority1. Develop and empower our workforce.....	14
Priority 2. Create a culture of data innovation	14
Priority 3. Establish a network of data leadership.....	15
Appendix 1: Evaluation our Data Ecosystem	16
Pre-Resolution Findings	16
Post-Resolution Findings.....	17

“To preserve, enhance, and restore the quality of California's water resources and drinking water for the protection of the environment, public health, and all beneficial uses, and to ensure proper water resource allocation and efficient use, for the benefit of present and future generations.” (State Water Board Mission Statement)

Introduction

The State Water Resources Control Board (State Water Board) directs and collects vast amounts of data and information on the state's water quality, quantity, and uses, as well as the agency business operations. These data are used to inform regulatory, operational, and water resource decisions and are therefore critical in achieving the agency mission and public trust through transparency and accountability.

As with all large and complex agencies, data at the State Water Board are siloed between offices, programs and even between staff. Additionally, the quality and usefulness of data vary between programs and the formats are often specific to the immediate data collector rather than supporting an agency-wide decision-making process. These disparities have hindered the ability of the agency to meet our commitments to transparency, well-informed decision making, efficient business processes, and innovation goals.

Legislative Directive

On September 23, 2016 Governor Edmund G. Brown Jr. signed into law Assembly Bill 1755, The Open and Transparent Water Data Act. The bill requires the Department of Water Resources, in consultation with the California Water Quality Monitoring Council, the State Water Resources Control Board, and the California Department of Fish and Wildlife, to create, operate, and maintain a statewide integrated water data platform; and to develop protocols for data sharing, documentation, quality control, public access, and promotion of open-source platforms and decision support tools related to water data.

In April 2018, the working group released the [Strategic Plan](#) for the Open and Transparent Water Data Act which outlined four main goals/guiding principles:

1. **Data are sufficient:** Data are sufficient to support water resources management and answer water resource-related questions.
2. **Data are accessible:** Data are available for use and discoverable.
3. **Data are useful:** Data are available in a form that facilitates use in various models, visualizations, and reports.
4. **Data are used:** Data are put to work in decision-making and innovation.

Water Board Directive

On July 10, 2018, the State Water Resources Control Board adopted [Resolution No. 2018-0032](#) adopting principles of open data as a core value and directing programs and

activities to implement strategic actions to improve data accessibility, usefulness, and governance (Open Data Resolution).

The resolution outlined five core principles for open data:

1. **Make Data Accessible** (“Open First”): our organization values transparency and strives to make all critical public data available in machine-readable datasets with metadata and data dictionaries.
2. **Understand Data Quality and Integrity**: our data are of known and acceptable 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, and staff and managers have robust data science capacity.
4. **Use Data to Govern**: our organization uses data to govern and 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.

Open and Transparent Data Act		Open Data Resolution	
1	Data Are Accessible	• Make Data Accessible	1
2	Data Are Sufficient	• Data Quality and Integrity	2
3	Data Are Used	• Improve Data Literacy	3
		• Use Data to Govern	4
4	Data Are Useful	• Govern Data	5

Figure 1. The four goals of AB1755 The Open and Transparent Data act mapped to the five guiding principles of the Water Boards Open Data Resolution.

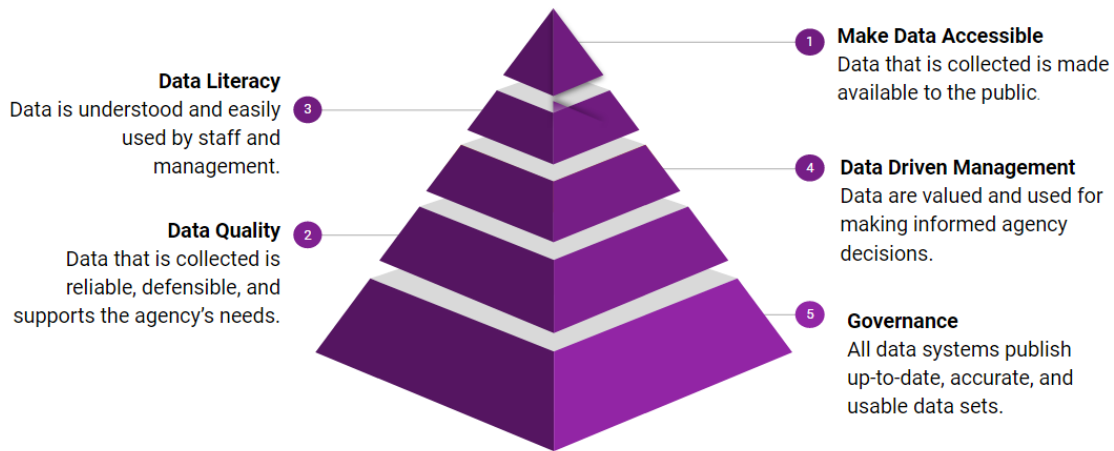


Figure 2. The five guiding principles of the Water Boards Open Data Resolution are shown as a pyramid to visualize the level of effort needed and their foundational and supportive role to the ultimate goal of publicly available datasets.

The Open Data Resolution calls for a Data Management Strategy that contains five minimum elements relating to these core principles:

- Open Data Strategy
- Data Literacy Strategy
- Data Driven Management Strategy
- Quality Management Strategy
- Data Governance and Administration Strategy

California Department of Technology Directive

In addition to legislative and Board drivers, the State of California's Department of Technology (CDT) has established goals and requirements for all state agencies that this plan must align with. CDT is responsible for establishing and enforcing statewide IT strategic plans, policies and standards. In 2017, CDT issued an updated strategic plan that highlights key requirements that this strategy addressed to achieve successful outcomes for both the agency and the state.

[Vision 2020 California Technology Strategic Plan](#)

October 2018 - [Open source technology letter](#)

On April 10th, 2019, CDT issued a statewide [Open Data Policy](#), detailing the requirements of each Agency/state entity to start or continue to enhance data-driven efforts to improve transparency, effective July 1st, 2019:

- Build or modernize Information Technology (IT) solutions in a way that maximizes interoperability and information accessibility when new systems are developed or significantly changed.
- Whenever feasible, make data broadly available to the public through the Agency/state entity’s open data site or portal.
- Describe information using standard metadata as the data is collected or created.

Strategic Goals, Priorities, and Measures

This strategic data management action plan establishes a framework for achieving our vision and goals through discovery, reflection and strategic actions that would help improve the way we use data and make decisions with information.

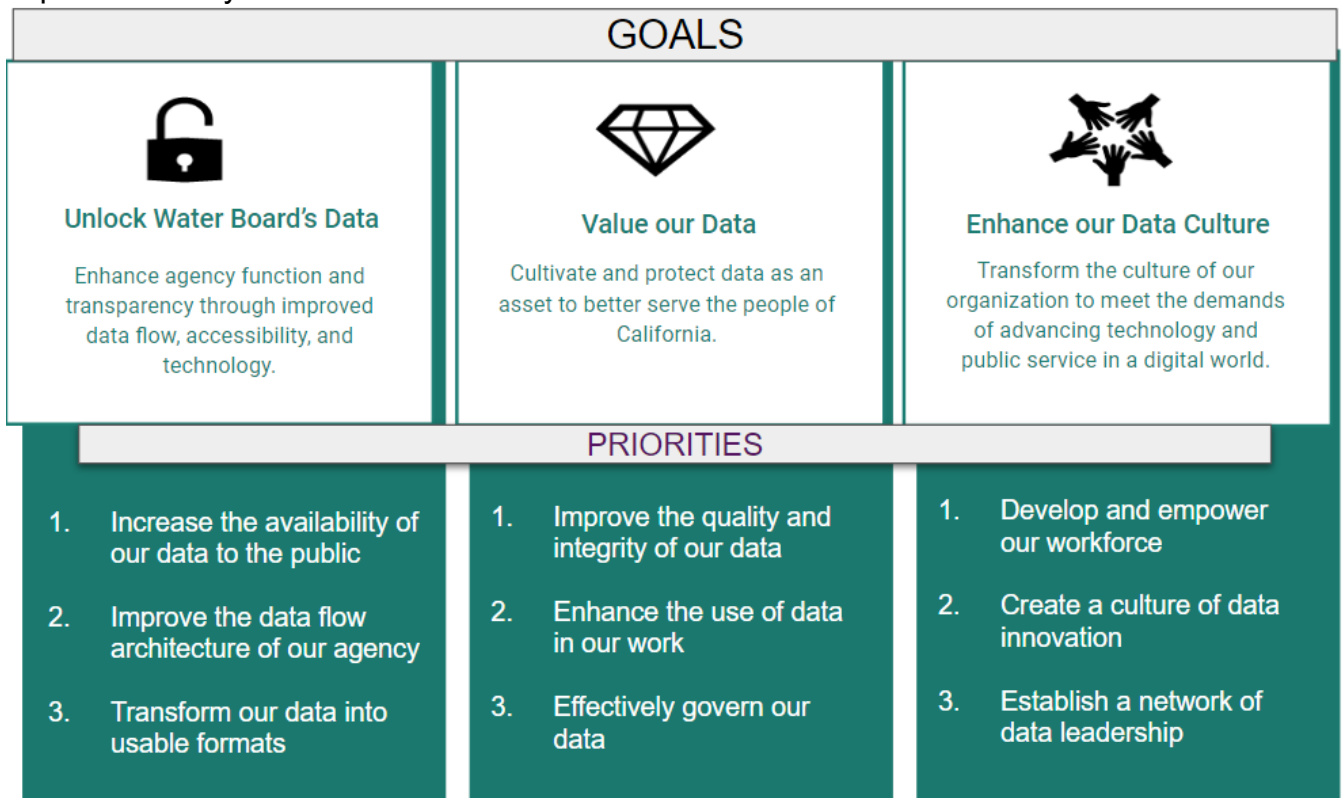


Figure 3. The three goals of this strategic plan are shown with their respective priorities.

[SDAP Project Portfolio Workbook](#): Overview and status of all projects.

[SDAP Project Portfolio Write Ups](#): Project write ups provide more detailed information about each project, timelines, and links to products and additional resources.

Goal 1: Unlock Water Board’s Data:

Enhance agency function and transparency through improved data flow, accessibility, and technology.

The State Water Board operates in organizational and database silos, limiting our ability to efficiently publish, combine and utilize datasets across our organization. This

limitation hinders our transparency goals and our ability to make coordinated and informed decisions to maximize our services to the public. It is crucial to our organization's success that we take steps to unlock those datasets from those silos, transform them into datasets that can be combined, used and published.

Outcomes: Water Boards data is accessible and usable to the Water Boards and we can easily publish up-to-date datasets to the public.

In preparation to achieve Goal 1, the Water Boards will complete the following, foundational projects as directed by the [Open Data Resolution](#):

[Project A. Data Set Inventory & Open Data Readiness](#): Complete an inventory of all State Water Board datasets. Evaluate and report on each Water Board dataset's open data readiness; to provide information to management on areas of needs and resource considerations to share data with the public.

[Project C. List of Critical Public High Priority Data Sets](#): In order to comply with the requirements set forth in the Open and Transparent Water Data Act (AB1755), OIMA will bring to the State Water Board a proposed list and description of critical public high priority datasets to be made public in machine- readable format according to the protocols for data sharing.

[Project D. Water Data Sharing on Open Data Portals](#): OIMA will coordinate the sharing of water data and information with other state and federal agency open data portals following the protocols set forth in the open data strategy.

[Project E. Agency Presentation Open Data](#): OIMA will work with EXEC to ensure that all presentations made by State Water Board staff in support of permits, policies, or other regulatory decisions using data or information derived from our data will be made available in an open data format.

Priority 1. Increase the availability of our data to the public

Measure: By the end of 2021, 90% of the Water Board's critical high priority public datasets will be available on an Open Data Portal.

[Project 1.1. Open Data Handbook](#): OIMA will establish protocols and guidance for sharing data on public platforms. ([Part of Project 2.6 Data Management Tool Kit](#)).

Priority 2. Improve the data flow architecture of our agency

Measure: By 2023, the agency will have 95% of the dataflow pipelines established for the Water Boards databases that allow this data to automatically flow into controlled, cloud places where specific communities can access the data.

[Project 1.2. Data Warehousing Around Key Topics \(Data Lakes and Data Services, e.g., Flat Files, Use Cases\)](#): Provide locations where all public Water Board data are copied

to, kept up-to-date, and stored with meta-data, data dictionaries, and other documentation with API functionality to make the data accessible, useful, and machine-readable. The Safe and Affordable Drinking Water Data Testbed is a pilot for this project.

Priority 3. Transform our data into usable formats

Measure: By the end of 2023, the Water Board will have identified the data models and standards for 95% of Water Board data types and have the ability to transform data from our current systems into these models for internal use and publication to the open data portal.

[Project 1.3. Protocols for Data Sharing \(Data Standards, Models, Metadata, and Workflows\)](#): Evaluate existing datasets to develop agency-wide data standards and metadata to ensure data interoperability. Establish and support workflows for data management and sharing data both internally and to the public, upholding principles of timeliness, data quality and integrity.

Goal 2: Value our Data

Cultivate and protect data as an asset to better serve the people of California.

Data are at the core of our business and functions and are crucial to our organization's success in making operational and water quality and management decisions. The State Water Board collects and uses data in the same organizational silos as our technological systems, limiting our ability to combine and utilize datasets to their maximum potential to make better-informed decisions. The State Water Board will aim to collect useful, comparable and defensible data to inform well-outlined management questions by establishing data quality objectives that include agency-wide data use needs. The State Water Board will also enhance the way we use data through coordination and alignment of our programs and performance metrics.

Outcomes: Water Boards data are of known and useful quality and we deploy practices to protect its integrity with standards and protocols. Our organization uses data to govern and makes decisions that are in the best interest of our mission(s).

In preparation to achieve Goal 2, the Water Boards will complete the following, foundational projects as directed by the [Open Data Resolution](#):

[Project F. Contracts & Grants - Data Quality and Integrity](#): Recipients of state funds managed by the State Water Board through grants or contracts for research or any other projects and studies shall, as a condition of the receipt of a grant or contract, adhere to the Data Management Strategy protocols for data sharing, transparency, documentation, and quality control. Every contract and grant that includes the gathering or analysis of data shall provide all the data and metadata used or collected in the format specified in the Data Management Strategy.6.0

[Project G. Data Quality and Integrity Infrastructure](#): New policies and permits adopted by the State Water Board that require collection of data or information or any other reporting requirement shall include a data management plan that documents the path that the data must take to become open. OIMA shall develop guidance and protocols on the content and structure of the data management plans.

Priority 1. Improve the quality and integrity of our data

Measure(s): By 2023, 90% of the agency programs will have established Quality Assurance Plans with well developed management questions, data quality objectives and data management plans.

By the end of 2021, the State Water Board will develop an active Quality Assurance Auditing Program for 100% of the consolidated contract laboratories.

[Project 4.1. Water Boards Quality Assurance Policy Implementation](#): Incorporate the Water Boards' Quality Management Plan (QMP) and 8.4.2. Quality Assurance Program Plan (QAPrP), and Quality Assurance Project Plan (QAPP) structure into all our core business workflows.

[Project 4.3 Universal Online Data Checker & Submission Tool](#): Develop a universal, online, interactive, format and quality assurance data checker to streamline data submission as well as educate data submitters and improve overall data quality for the Water Boards

[Project 4.4 Quality Assurance Coordination](#): The Quality Assurance Roundtable was established to ensure that an appropriate level of planning for data acquisition and analysis is applied consistently throughout the Water Boards through information sharing, and discussion of issues and solutions of mutual interest.

[Project 4.5 Implementation and Monitoring](#): Implement and monitor quality assurance and data management planning. Provide an infrastructure for tracking and auditing practices to protect data quality and integrity.

Priority 2. Enhance the use of data in our work

Measure(s):

By the end of 2022 the State Water Board Performance Metrics will be 100% updated to enhanced web-based, user-centered, interactive formats.

By the end of 2020, the State Water Board will establish outcome based performance measures for five major program areas of critical interest to the public.

[Project 3.1 Performance Measures Metrics Assessment & Workflows](#): Use data to establish and report on performance metrics for each program. Evaluate use of outcome and social science end-points to communicate agency work and value. Evaluate the data flow needed to inform performance measures.

[Project 3.2 Program Management Information Tools](#): Meet with managers to determine management questions, data needs, gaps, workflow bottlenecks and assess potential visualization tool solutions.

[Project 3.4 Data Storytelling Reporting Platforms](#): Evaluate and identify storytelling platforms for Water Boards use.

[Project 5.2 Data Planning and Use Alignment](#): Uses data management plans, quality assurance plans, and related documents to align our interests and practices to better achieve our mission, goals, etc;

Priority 3. Effectively govern our data

Measure: By the end of 2022 the State Water Board will develop Data Management Plans for 100% of the datasets listed within [Project C. List of Critical Public High Priority Data Sets](#).

[Project 2.6 Data Management Tool Kit](#): The Data Management Toolkit aims to provide resources to the Water Boards divisions and offices on all things data. Readers at any stage of their data journey will be able to utilize different aspects of this Toolkit, depending on their needs. This Toolkit is intended to be a living resource available to all Water Board staff that will adapt as our datasets, available tools, and related programmatic needs evolve

[Project 5.1 Data Lifecycle Governance](#): Addresses data governance and management for all data life cycle stages through the development of Data Management Plans

[Project 5.3 Data Sharing Governance](#): Framework for data sharing, approvals, and a mechanism to address emerging legal and/or governance issues with new data collection, storage and accessibility technologies.

Goal 3: Enhance our Data Culture

Transform the culture of our organization to meet the demands of advancing technology and public service in a digital world.

The State Water Board will invest in our workforce through training, engagement, and making available resources.<draft placeholder>

Outcomes: 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. 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.

Priority1. Develop and empower our workforce

Measure:

[Project 4.2 Water Board Quality Assurance Curriculum](#). Expand staff knowledge and understanding of quality systems, applicable laws, and integration of quality systems into planning, data use, and all Water Board business operations.

[Project 2.1. Data Science Curriculum](#). Develop staff and management with skills and knowledge to better derive information from data.

[Project 2.2 Business Intelligence Tools, Web Resources, and Staff Engagement:](#)

Develop staff and management in the areas of business intelligence, data analysis, data science, and data engineering. Provide information on business intelligence tools which could potentially be adopted by the State Water Board. Coordinate user groups that focus on different data tools, such as Tableau User Group, Power BI User Group, and R User Group.

Priority 2. Create a culture of data innovation

Measure:

[Project 2.4 Conduct Annual Water Data Science Symposiums:](#) Enhance how water quality monitoring generates meaningful data that informs water quality management decisions, engage the public, and foster collaborative partnerships with other governmental and non-governmental organizations, and the communities we serve.

[Project 2.5 Civic Data Engagement:](#) Enable collaboration among Water Board Divisions and engage the public and interested stakeholders to, in a focused environment, work towards answering challenging water management questions using available data.

Priority 3. Establish a network of data leadership

Measure:

[Project 5.4 Data Management Innovation Team \(DMIT\)](#): Develop and deploy a team of representatives throughout the agency to assist with the improvement of how the Water Board collects, stores and makes accessible its vital data and information.

[Project 5.5 Data Leadership and Organizational Culture Change](#): Establish an interim executive steering team convened to help drive, evaluate and adapt the implementation of the plan and to recommend and form a more permanent Data Governance system.

Appendix 1: Evaluation our Data Ecosystem

While the Open Data Resolution was adopted in 2018, work began as early as 2014 to evaluate the current Water Board data ecosystem and look for opportunities to improve data quality, flow and transparency. Exploratory findings drove the development of the proposed resolution as an urgent need for the agency to find solutions to our aging and siloes data systems.

Pre-Resolution Findings

Quality Assurance

A Water Boards Quality Management Strategy has been in place as early as 2010. In an update that was finalized in 2017, it was found that data quality support was lacking in many Water Board programs, Regions, and business areas. To date the roundtable is mostly an informational group and has only six of the nine regions represented and almost no representation from the district offices where the drinking water programs are implemented. The State Water Board has only developed two Quality Assurance Program Plans that have been approved by US EPA Region 9. The Water Boards currently has X water quality monitoring and/or data use programs.

Contracts

In 2016, OIMA performed an evaluation of existing laboratory contracts to review current scope of work language, services offered, and laboratory prices to inform internal contract development. This evaluation revealed that the Water Boards often held several contracts with the same vendor with notable differences in the requirements within Scope of Work and prices, and inconsistencies in quality control and data flow requirements.

Data Systems and Open Data Publication 2015-2018

The enterprise systems discussed in the 2008 Agency Information Management Strategy (AIMS) have grown substantially to now be well over 100 applications and systems. The core data flow strategies discussed in that document and the architecture diagram referred to as the “tuna can diagram” have shifted, as well. No longer is all water quality program regulatory data flowing explicitly to the system called, California Integrated Water Quality System (CIWQS), for example. And we have inherited many new systems and key datasets with the integration of the drinking water program to our organization. In 2016 we embarked on an effort to build “flat file” services from the new systems to use as data wholesale sources from which we could publish open water datasets at our open data platform, data.ca.gov. This effort has resulted in 13 new flat files services from the water rights database (called electronic Water Rights Information System, eWRIMS). Additionally we developed new and improved services from the stormwater system (Stormwater Multiple Application and Reporting System, SMARTS) and the multi-purpose off-premise system called GeoTracker. The flat file services from CIWQS, SMARTS and GeoTracker were used to publish our first 30 data resources on the state’s open data platform.

Post-Resolution Findings

Evaluation of Current State of Data Management “Ecosystem”

Item #6 of the State Water Board resolution directed the Office of Information Management and Analysis (OIMA) to review existing database governance and use (delivery) of data from the primary database applications and to develop an inventory of all existing datasets including an assessment of the open data readiness of those datasets.

Beginning in November 2018 the staff in OIMA began the evaluation of the Water Boards’ data readiness and creating an inventory of datasets used for core business. This process set a target of over 100 interviews with staff, managers and executives of the State Water Board. The evaluation was designed to:

1. Conduct the strategic review of data and database governance;
2. Review data delivery and use;
3. Conduct an inventory of open data readiness for business and mission critical datasets (and a process for all others);
- and 4. Conduct an assessment of opportunities to make quick progress on data management strategies.

To date, we have completed 87 interviews. We plan to complete the remaining interviews by the end of 2019, and over the early implementation period we may continue to interview other parts of the organization to ensure we understand the data management needs of all our programs and business functions. The results of these initial interviews, along with some analysis of our enterprise architecture and civic engagement around our data, serve as the foundation of our evaluation of the current, baseline data management ecosystem upon which the recommendations will build.

Summary of Evaluation Findings

The existing Water Boards data ecosystem is a close reflection of the organizational structure by programs which at the same time is defined in legislation and implemented through the budget through the identification of allowable expenditures and revenue sources. It is a complex system where solutions are developed and considered in isolation without a vision and understanding of how the data and the solutions developed could be applied to the rest of the organization. We have identified about 120 datasets that could be considered “open data” candidates, so far, and these are in various levels of open data readiness. We will prepare a list of them with readiness scores and include this in our documentation. Per the statute we are required to publish all our data to the open data platform by September 2019. We plan to do our best to get our data published by that deadline but we understand that open data publication is a continuous process.

Conclusions gained from evaluation:

- Data is at the core of our business and functions. But in many cases, the data we have available is not used, we don’t have the data we need, it is not in the right format, or it is difficult to understand.

- We are not taking full advantage of new technology. Most programs/staff rely on “system functionality” to view, analyze and report data.
- We must significantly increase the type and amount of professional and organizational development in the data science skills necessary to integrate our data, create better data structures, improve our quality management systems, perform critical data cleaning, and turn our data into useful information and knowledge.
- Most of our legacy IT systems (databases) have been developed for case management, not “longitudinal” or statewide analytics. Many of them now act as silos with different formats and business rules. Our data is not well integrated nor is it very interoperable.
- Almost all of our data is not open. Although data is in most cases “legally open” it is not “technically open” in a machine-readable format with a sufficient level of detail and disaggregation.

In addition to the evaluation we conducted we recruited a team of employee “data innovators” that has been exploring the results of the interviews, exploring innovative topics (like how to pilot machine learning, build data literacy capacity, and use of standard data models to make data more interoperable and accessible). We will provide an update in the next month on the Data Management Innovation Team, it’s expert advisory panel, and the products they are working on and how they relate to this effort.