JANUARY 2018

DATA FOR WATER DECISION MAKING

Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement

BerkeleyLaw
UNIVERSITY OF CALIFORNIA
Center for Law, Energy & the Environment

UC WATER
Security and Sustainability Research Initiative

CCST
CALIFORNIA COUNCIL ON SCIENCE & TECHNOLOGY

DEPARTMENT OF WATER RESOURCES
STATE OF CALIFORNIA

Cantor, Alida, Michael Kiparsky, Rónán Kennedy, Susan Hubbard, Roger Bales, Lidia Cano Pecharroman, Kamyar Guivetchi, Christina McCready, and Gary Darling. 2018. Data for Water Decision Making: Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement. Center for Law, Energy & the Environment, UC Berkeley School of Law, Berkeley, CA. 56 pp. Available at: https://doi.org/10.15779/J28H01 or law.berkeley.edu/datafordecisions

Wheeler Water Institute
Center for Law, Energy & the Environment
University of California, Berkeley, School of Law
390 Simon Hall
Berkeley, CA 94720-7200

CLEE@law.berkeley.edu
clee.berkeley.edu
wheeler.berkeley.edu

UC Water Security and Sustainability Research Initiative
www.ucwater.org

JANUARY 2018

DATA FOR WATER DECISION MAKING

Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement

AB 1755 STAKEHOLDER WORKING GROUP SYNTHESIS REPORT

Alida Cantor, Michael Kiparsky, Rónán Kennedy, Susan Hubbard, Roger Bales, Lidia Cano Pecharroman, Kamyar Guivetchi, Christina McCready, Gary Darling

A multi-institution research collaboration led by
Wheeler Water Institute
Center for Law, Energy & the Environment
UC Berkeley School of Law

BerkeleyLaw
UNIVERSITY OF CALIFORNIA
Center for Law, Energy & the Environment

UC WATER
Security and Sustainability Research Initiative

CCST
CALIFORNIA COUNCIL ON SCIENCE & TECHNOLOGY

DEPARTMENT OF WATER RESOURCES
STATE OF CALIFORNIA

ABOUT THIS REPORT

This report is a collaborative effort of UC Water, the California Department of Water Resources (DWR), and the California Council on Science and Technology (CCST). In 2017, UC Water, DWR, and CCST convened a workshop series that serves as a foundation for this report. Workshops in February and May of 2017 engaged stakeholders and decision makers in the development of use cases—short examinations of how decision processes employ data—to inform a decision-driven water data system. The approach of defining objectives and outlining data needs was iteratively refined over the course of the workshops and subsequent engagements with stakeholders.

This document is one of several products related to AB 1755 implementation. The state agencies leading implementation of AB 1755 are currently working to develop a strategic plan,[1] while a technical working group is developing system requirements[2] and a set of standard operating procedures that will inform technical data system development.

The target audience for this report includes state agencies involved in data provision (including, for example, DWR, California State Water Resources Control Board (SWRCB), California Department of Fish and Wildlife (CDFW), California Water Quality Monitoring Council (CWQMC), and others); other data providers (including federal agencies as well as NGOs and universities); technical developers of data systems; and data users (including water managers, environmental managers in other sectors, regulators, researchers, policy makers, and other decision makers).

CONTENTS

About this report 4

Figures and tables 6

Executive Summary 7
- Engaging stakeholders in data system development 7
- Recommendations for an open and transparent water data system in California 8

1. Introduction 11

2. Water data in California 13
   - A. Information as a fundamental element of water security 13
   - B. The Open and Transparent Water Data Act (AB 1755) 15

3. Decision-driven data systems 17
   - A. Big data and information technologies in water management 17
   - B. Science for policy 18

4. A process for engaging stakeholders in assessing water data needs 20
   - A. Developing use cases through stakeholder workshops and collaborations 20
   - B. Summary of use cases 21
   - C. Methodological limitations 24

5. Data availability, sources, and gaps 27
   - A. Assessing data needs 27
   - B. Data providers 29
   - C. Data gaps and limitations 31
   - D. Additional stakeholder observations and perceptions 31

6. Lessons for open and transparent water data 34
   - A. Basic cyberinfrastructure considerations 34
   - B. Stakeholder engagement and data system governance considerations 37

7. Looking forward: Recommendations for implementation of AB 1755 40

Institutional Acronyms 43

Glossary 44

Acknowledgments 45

Review 46

About the Authors 47

About CLEE and UC Water 49

Endnotes 50

FIGURES AND TABLES

Figure 1: Data life cycle. 12

Figure 2: Key elements of water security 13

Figure 3: Elements involved in understanding a groundwater basin water balance. 14

Figure 4: AB 1755 implementation timeline 16

Figure 5: Common components of decision support systems. 18

Figure 6: Relationship between stakeholder use cases and functional requirements. 21

Figure 7: Classification of use cases by topic and decision objective 25

Figure 8: Alignment between Water Action Plan goals and use case topics 26

Figure 9: Data sources cited in use cases, categorized by topic 29

Figure 10: Number and topic of data sources cited within each use case 30

Figure 11: Percent of unique data sources used by one or multiple use cases 30

Figure 12: Percent of data sources cited in use cases provided by specific agencies 31

Table 1: Use case template 21

Table 2: Summary of use cases and objectives. 22

Table 3: Data types and parameters included within topic categorizations 28

Table 4: Stakeholder concerns and perceptions 33

EXECUTIVE SUMMARY

California's water management systems are struggling to meet environmental and human needs in the face of stresses from climate, population, and land use changes. Calls to improve the operation, management, and infrastructure of water systems are frequent. Making rational and equitable water management decisions depends upon timely knowledge of the current and projected state of the water system. This, in turn, requires robust data and information.

Currently, California's water data fall far short of the fundamental goal of being complete, accessible, and usable—that is, readily available in formats that suit users' needs, at relevant spatial and temporal resolutions that are useful for making the decisions at hand. Data sources are incomplete and inconsistent. Fragmented data systems often cannot exchange data with one another. Existing data are frequently difficult to access or use. Ironically, the birthplace of modern information technology lacks both basic information and the capacity to use it effectively to manage its most precious resource.

Strategic, coordinated investments in California's water information infrastructure are necessary. Without basic information on where, when, and how much water is available and being used, as well as physical, chemical, and biological measurements of water quality, we cannot improve how we manage our water resources. A modern water data system that enables accurate, timely, transparent accounting of water supply, quality and use could enhance water security and sustainability. With better and more usable data informing water management, California's existing water resources could better meet urban, agricultural, ecological and industrial needs.

California's 2016 Open and Transparent Water Data Act (AB 1755) provides an important opportunity for improving the state of water data in California. The bill charges state agencies with integrating currently fragmented water and environmental data systems. However, while AB 1755 provides an opening for improved data systems, the law itself does not ensure increased usability of data for decision making. Increased usability of data requires a broader rethinking of data systems and their interactions as applied to decision making contexts.

ENGAGING STAKEHOLDERS IN DATA SYSTEM DEVELOPMENT

If California is to enable better data-driven decision making in its water sector, it must begin by developing decision-driven data systems. In order to be useful for decision making, data must not only be open and transparent, but must also be presented in a way that is relevant to the needs of decision makers.

This report presents a case for basing the development of data systems upon end users' needs; describes a process for engaging stakeholders in the assessment of data needs and the design of a data system; and summarizes lessons emerging from the engagement process. The goal of the report is to support California's efforts to make the most of AB 1755. In concert with other efforts, we seek to develop an inclusive and actionable vision for the future of water data in California. This requires understanding, articulating, and communicating decision makers' processes, and using these requirements to inform the design of data systems.

Through a series of workshops in 2017, we engaged data users and data providers from multiple sectors, including academia, state agencies, and NGOs, to systematically analyze data needs for decision making. To do so, we developed a method to generate 'use cases'—short examinations of how decision makers use data. Use cases are a tool for assessing stakeholder data needs in specific decision contexts and communicating those needs to technical developers. The 20 use cases developed and analyzed in this report cover a range of topics in California water management, from water supply to water quality to long-term infrastructure investments.

Thousands of decisions are made daily regarding California water management. Sampling and analyzing a range of decision making contexts provides a glimpse into how these decisions use—or could use—data, which allowed us to draw some preliminary observations and recommendations. We hope the emergent conclusions from this preliminary exploration will help inform the State and relevant stakeholders as California solidifies its vision and process for modernizing its water data. We also hope that the analysis informs the development of data system functions and requirements, keeping the interests of end users in mind.

RECOMMENDATIONS FOR AN OPEN AND TRANSPARENT WATER DATA SYSTEM IN CALIFORNIA

The following recommendations are based on the results of our analysis of use cases, combined with insights drawn from workshop participants and lessons drawn from science integration and data system development in other contexts. If the state of California is to successfully navigate its current opportunity to develop an open and transparent water data system that effectively informs decision making, the following recommendations will aid its planning.

1. **To ensure relevance, an understanding of the way data are used in decision making should guide the development of data systems.** Achieving relevance requires first understanding the intended purpose of data, and then designing data collection, provisioning, and analysis systems to serve decision making. Beginning with the questions that are of concern to stakeholders, and incorporating a recognition that these questions will evolve over time, can ensure that relevant data are collected and useful analyses are developed. This can help to avoid the costly need to re-engineer the system after it has been developed. Involving end users in a two-way conversation can help to ensure that scientific information is truly useful for decision makers and policymakers.

2. **A wide variety of data must be easily accessible and highly interoperable to serve many different user needs.** Water decisions in California draw from a wide range of data types and data sources, including data about the flows and quality of water, as well as agricultural, land use, ecological data, and many other data classes. A system with high interoperability among different data sectors and topics could increase the ability of decision makers to integrate data from multiple data sources and sectors. Measurable metrics of data interoperability and accessibility should be used to gauge success.

3. **Data gaps take a variety of forms and will need to be filled in distinct ways.** Some data and information critical to decision making are not available at all and will need to be estimated or collected in order to move forward. Other data and information are available, but are not in a suitable format to be useful for decision making processes. For example, data are not interoperable or are hard to access or use, which can be equally significant barriers to informed decision making. Thus, the State's efforts to identify and address data limitations will need to take into account not only the existence of data, but also accessibility and interoperability.

4. **For California water, an integrated data system will need to connect data from multiple sources, while maintaining the autonomy of those independent systems.** Water data are produced and distributed by numerous state and federal agencies. Given the distributed nature of water data in California, the independence of disparate agencies, and the need for interoperability, a federated data system that enables exchange across distributed data sources is likely to have distinct advantages. However, the institutional, financial, and computational costs and benefits of any approach should be considered carefully. Any form of data system integration must provide clear standards for data quality, documentation, and archiving. To facilitate data integration, protocols and methods must be employed to ensure that data are properly collected, handled, processed, used, and maintained at all stages throughout the data life cycle.

5. **A water data system must address needs for data at multiple spatial and temporal resolutions, and in multiple, distinct forms and formats.** Analysis of use cases reveals that different decision makers have varying data and information needs, even on similar topics. Different formats and resolutions of data and information are necessary for different decisions, and a useful data system must take these needs into consideration. Enabling flexibility of data uses, while maintaining data quality and integrity, are non-trivial but crucial challenges.

6. **A water data system must enable the production of new information.** Ultimately, the goal is not only data provision, it is enabling the production of information (data that have been processed in such a way as to be useful). To this end, decision makers need data sources that can be readily integrated with one another, that are consistently updated with quality data, and that can support specific outputs such as analytical tools to guide informed decisions. Many models and decision support systems that process data into information incorporate a range of disparate data sources, which is particularly important since water resources management is often integrative by nature. At the same time, for many audiences such as academics and researchers, access to raw source data is key in order to produce or verify visualizations or other data products.

7. **Engagement between data system developers and end users is, ideally, an ongoing and iterative process.** Our use case efforts focused on identifying "*who* needs *what data* in *what form* to make *what decisions.*" An important next step could be to define "*how* are diverse datasets used to produce *what output* needed for what *specific objective*" for select representative use cases. In tandem with production of data system architecture, the State and end users of data could work to develop on ongoing process of assessing data user needs, including the specific analytical needs of decision makers, and developing a process for long-term follow up on use case implementation. Such iterative development could help to continually improve the design and functionality of data systems.

8. **Basing water data on principles of usability and stakeholder engagement requires robust cyberinfrastructure, good governance, and stable funding.** Stakeholders described needs for interoperability, data quality, and documentation that must be addressed in the development of a data or information system or platform. Beyond these basic cyberinfrastructure properties, a data system that is sustainable over the long term requires good governance and stable funding. Specific resources should be dedicated to information management and operability. Commitment to ongoing communication with end users is fundamental to ensuring that a data system will be able to contribute to improved decision making by meeting users' needs. Water data infrastructure should be conceptualized as a complex adaptive system: it must meet the needs of users, but be modular and flexible enough to allow alterations and reconfigurations. A functional institutional framework will require

clear, long-term financial commitments by policy makers, lead agencies, and data providers. Both an initial investment and consistent and ongoing resources will be required.

AB 1755 is a major step towards developing an open and transparent water data and information system for California. Working towards open and transparent water data for California will require, first and foremost, good governance. Developing quality data and information systems in a useful and usable form requires not only resources, but also substantial commitment to the processes of building relationships and working with stakeholders. These processes start with AB 1755, but the effort of engaging decision makers in working towards greater availability of data and information to inform modern water management must continue as a broader intitiative. The current momentum and collaborative efforts between agencies and stakeholders are encouraging, and these efforts represent meaningful progress towards actualizing a vision of data-driven decision making for the sustainability of California water.

1. INTRODUCTION

California faces many water management challenges, from balancing urban, environmental, and agricultural water needs to managing the impacts of drought and climate change.[3] Addressing these challenges involves making decisions, and making sound, evidence-based decisions in turn requires reliable, usable data.[4]

While data are a crucial ingredient for enabling effective decisions, providing data does not in and of itself ensure that data can or will be effectively utilized for better water management. The form, accessibility, and usability of data provided can make the difference between an invaluable resource and a stranded asset. To be useful for decision making, data must be open, transparent, and relevant to the needs of decision makers. Generating data that meet these criteria requires understanding and articulating the processes and contexts of decision making and using these articulations to inform the development of water data systems.

In order to support state agencies and stakeholders in their efforts to modernize California's water data systems, this report outlines a case for basing water data systems on decision makers' needs. We describe a process for engaging stakeholders to assess data needs, and lessons emerging from the process.

Currently, California's water data system does not meet the fundamental goal of being usable for decision making. Data are diverse and fragmented. They are produced, housed, and maintained by multiple entities. Interoperability—the ability of information technology systems to exchange meaningful information with each other in standard ways that allow for common comparison, aggregation, and analysis—in many cases is low. Some fundamental data do not exist or are not gathered using standardized approaches—for example, data about individual groundwater extraction is not yet collected in any systematic way. Other data are hard to access or hard to use— for example, most legal information on California water rights and water use exists only in paper form and awaits digitization.[5] The lack of uniform, annotated, high quality, accessible, and ultimately usable data, as well as the limited interoperability of data systems, hampers evidence-based water management in California.[6] The net result is less-informed decisions on how to best manage a foundational resource for California's environment and economy.

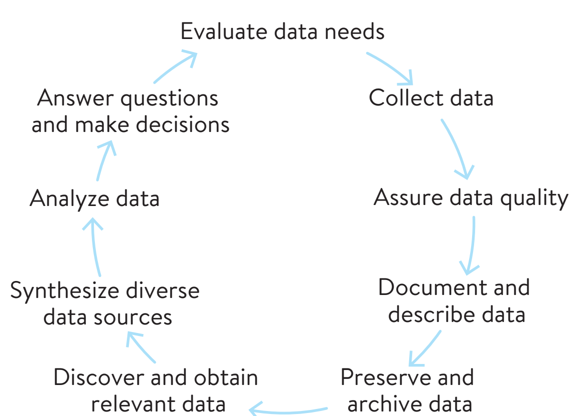
WHAT ARE WATER DATA SYSTEMS?

In this report, we use "water data" to refer to the broad suite of data and information that inform decision making and research on water-related topics, including data to characterize systems and to monitor systems. "Water data" goes beyond streamflow and precipitation measurements to include many areas relevant to water management, including data on ecology, land use, and agriculture. We use the term "data" to refer to measurements of basic properties in the world, while "information" refers to data that has been processed or synthesized in order to answer questions. "Data and information systems" refers to software and hardware systems that collect, organize, archive, distribute, integrate, process, analyze, or synthesize data.

California needs to develop a comprehensive, integrated system for water data.[7] By data, we mean more than just numbers: the production and use of data involves a broader life cycle that begins and ends with evaluating data needs (Figure 1).[8] Improving data-driven decision making requires a rethinking of data systems along this entire life cycle, in conjunction with relevant decision makers. Starting at the top of the data life cycle, carefully evaluating needs of decision makers can facilitate the development of more useful data and information. Conversely, a data system that does *not* put users and decisions at its center raises the risk of spending time and money while failing to improve decision making capacity. Ideally, water data can contribute to both efficiency and equity in operational, economic, and regulatory decision making.

AB 1755, the Open and Transparent Water Data Act, requires coordination among multiple agencies to integrate existing data that are currently fragmented. The bill provides an opening for improving California's water data system, but careful implementation and follow-through will be crucial. AB 1755 enables crucial movement towards a fully realized water data system, raising an opportunity to generate momentum on improving water data usability. Implementation of AB 1755, with careful forethought, will be crucial for enabling informed water management for California's future.

Figure 1: Data life cycle. Adapted from NSF DataONE project.



The main goals of this report are to:

1. Illustrate the value of starting with decision makers' needs in data system development.

2. Develop and describe a method for assessing data needs from a stakeholder perspective.

3. Begin to prioritize water data efforts by assessing data availability and limitations based on a preliminary analysis of decision making processes in the California water sector.

4. Present key lessons on decision-driven data systems that emerged from this work.

Starting from the premise that engaging decision makers in data system development increases potential usefulness of data systems,[9] we worked with stakeholders to think systematically about data needs for decision making. In collaboration with decision makers we developed and analyzed a set of 'use cases'—short examinations of how water management decision processes employ data. The use cases serve as a tool for assessing stakeholder data needs and communicating those needs to technical developers.[10] The approach of starting with decision makers' perspectives may help to develop a stronger understanding of what a data system needs to do in order to be useful.

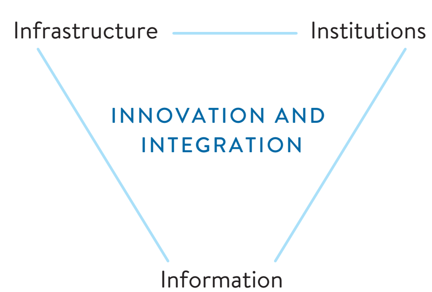
2. WATER DATA IN CALIFORNIA

Information is, or should be, a foundation for water management and decision making. Sound information can improve decision making for water security and sustainability. However, the reality of California's data systems do not always match this ideal.

A. INFORMATION AS A FUNDAMENTAL ELEMENT OF WATER SECURITY

Water security is defined as the reliable availability of an acceptable quantity and quality of water for ecosystem and human health, livelihoods and production, coupled with an acceptable level of water-related risks.[11] Sustaining California's water security in the face of unprecedented changes requires investments in three tightly linked areas: infrastructure, institutions, and information (Figure 2).

Figure 2: Key elements of water security. Adapted from UC Water.



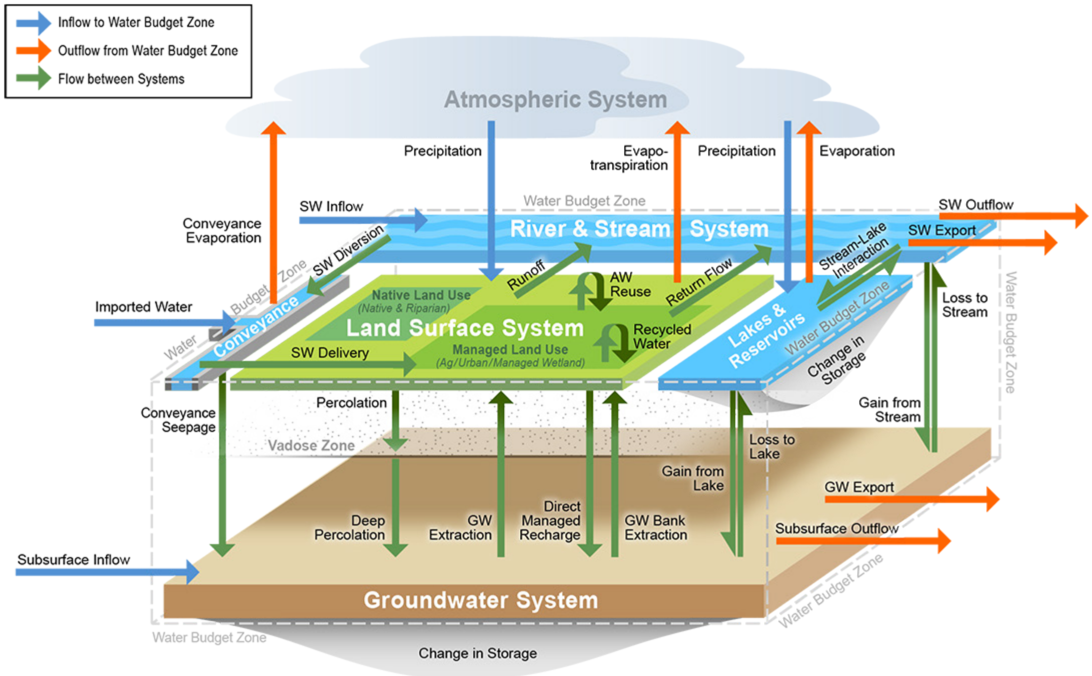
A sustainable water future involves planning for, investing in, and maintaining natural and built infrastructure, as well as digital information infrastructure. Strong institutions, including laws, regulations, agencies and organizations, but also the norms and conventions that govern decisions, are also essential for sound water management.

Information, the focus of this report, is a key element of water security. Information can guide better decision making, inform the design and maintenance of infrastructure, guide the development and operation of institutions, and ideally ultimately improve water security,[12] both under current conditions and in the face of an uncertain future. Data are quantitative or qualitative representations of basic properties of the world. Processing and synthesizing those data so they can be used to answer questions produces information.[13] Information is generated from data through analytics, integration, and other processes. In the water sector, infrastructure, institutions, and information are by no means separate from one another, since institutions are responsible for managing both infrastructure and information.

Water is an inherently integrative and interdisciplinary field of inquiry. Water management requires incorporation of multiple streams of information into cross-cutting analyses.[14] Thus, the provision of data and development of data and information systems that can handle diverse data are crucially important for decision making in the water sector.

Accessibility of information at appropriate spatial and temporal resolution can be a bottleneck for making sound decisions. Data and information can and should inform decision making around water management, such

Figure 3: Elements involved in understanding a groundwater basin water balance. Figure from DWR.



as drought management, flood response, and ecosystem protection. However, the lack of basic numbers on water supply, demand, and quality limits the accuracy and completeness of information derived from those data. This, in turn, limits our use of data to manage water and natural resources. Unmet data needs hinder efforts to characterize groundwater conditions, operate water supply infrastructure, achieve water conservation goals, reduce energy use, protect important ecosystems, and promote water transfers.[15, 16]

A simple example illustrates these points. Many water decisions require an estimation of a water balance over a watershed, groundwater basin, agricultural field or other area. For example, a decision maker might want to know how much water is likely to be available for consumptive use or export in the current water year; whether enough water is available to satisfy a given water right; or whether groundwater recharge is sufficient to make up for groundwater pumping. As Figure 3 illustrates, calculating a water balance for a given groundwater basin or other sub-basin requires a range of data, including inflow to the system (e.g., precipitation, water imports, subsurface flow) and outflow from the system (e.g., evapotranspiration, surface water outflow, groundwater and subsurface outflow), as well as flows between connected systems.

Generating a water balance is, in concept, attainable (Figure 3).[17] However, in practice, each component of the water balance is typically measured separately and at different spatial and temporal resolutions. Water balance measurements may be stored in different formats, in different repositories, and with different metadata. Each step—including assembling and integrating all of the diverse datasets, ensuring that the data are reliable, and performing analytics or synthesis—presents non-trivial challenges to converting diverse water data into water balance information.

Exacerbating these difficulties, the data required to compute a water balance are not readily available for all basins. Key elements, such as details of local aquifer properties, groundwater flows, and pumping, may be missing or difficult to obtain. These and other data challenges hamper real-time, transparent decision making in the California water management sector.

B. THE OPEN AND TRANSPARENT WATER DATA ACT (AB 1755)

California's Open and Transparent Water Data Act (AB 1755), passed in 2016, focuses on integration and interoperability of existing data sources. The bill requires the Department of Water Resources (DWR), in consultation with the State Water Resources Control Board (SWRCB), the Department of Fish and Wildlife (CDFW), and the California Water Quality Monitoring Council (CWQMC) to create and maintain a statewide integrated water data platform.[18] AB 1755 requires that these state agencies coordinate and integrate existing water and ecological data (Sidebar 1) from local, state, and federal agencies for several purposes, including implementation of the Sustainable Groundwater Management Act, increasing transparency of water transfers and markets, and water management more generally.

AB 1755 leaves undefined the format and purposes of the system, and does not specify with any granularity what decisions will be informed by it. For example, it is conceivable that the requirements of the bill could be met by simply posting static spreadsheets online. Although even this action would represent progress compared to the current state of water data, AB 1755 presents an opportunity to go much further toward the ultimate goal of improving water policy, management, and decision making through the improved provision of data.

SIDEBAR 1: EXAMPLES OF OTHER WATER DATA EFFORTS IN CALIFORNIA

Some examples of existing California systems[19] for collecting and sharing data include:

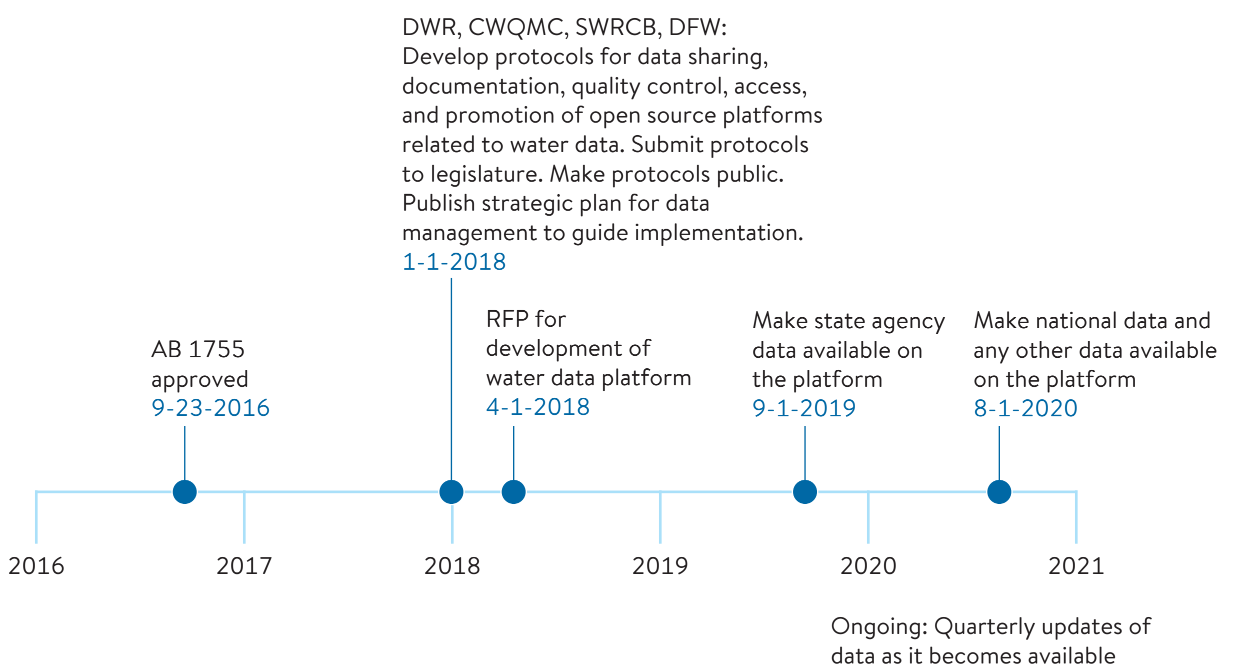
- **California Data Exchange Center (CDEC):** An extensive hydrologic data collection network.[20](DWR)
- **California Irrigation Management Information System (CIMIS):** A network of automated weather stations that produce evapotranspiration data.[21] (DWR)
- **Surface Water Ambient Monitoring Program (SWAMP):** Provides information on surface water quality and links data from water monitoring programs.[22] (SWRCB)
- **California Environmental Data Exchange Network (CEDEN):** A data network that aggregates data on water quality, aquatic habitat, and wildlife health.[23] (SWRCB)
- **Electronic Water Rights Information Management System (eWRIMS):** Tracks information on water rights in California.[24] (SWRCB)

Each of these systems focuses on a particular aspect of water management—for example, hydrologic conditions; water quality; or water rights. AB 1755 strives to add to the value of these existing data systems by connecting different areas of water management.

In spite of these ambiguities, AB 1755 specifies an aggressive timeline for implementation (Figure 4). The timeline is consistent with forthcoming data collection efforts—for example, the 2014 Sustainable Groundwater Management Act (SGMA),[25] which raises new data needs and calls for new groundwater data collection. However, the actions mandated by AB 1755 are extraordinarily ambitious, and without careful and concerted efforts for thoughtful and forward thinking implementation there is a real risk of wasting the opportunity afforded by AB 1755.

In this report we contend that AB 1755 implementation needs to begin with the end in mind—a clear understanding of the decisions that the data system is meant to support. We envision a data system that serves the needs of specific decision making entities, while enabling production and integration of new types of data and support for new uses of data. The messages about developing data systems that reflect end users' needs are not limited to the scope of this particular bill. The vision for a water data system that is rooted in decision makers' needs, and the lessons about how to achieve such a system, are applicable beyond the AB 1755 context and timeline.

Figure 4: AB 1755 implementation timeline



3. DECISION-DRIVEN DATA SYSTEMS

Data-driven decision making describes the practice of making decisions based on analysis of data rather than experience or intuition.[26] This report suggests a subtly different point: enabling data-driven decision making requires first developing *decision-driven data systems*. The premise is similar—that better decisions can be made using data—but the concept also recognizes the development of data systems themselves must be informed by decision makers' needs. Decision-driven data systems thus suggests a process that first examines stakeholder needs, then bases functional and technical requirements for data systems on these needs.[27]

A. BIG DATA AND INFORMATION TECHNOLOGIES IN WATER MANAGEMENT

Contemporary approaches to water management tend to move past simple supply/demand calculations to recognize the complexity of the resource.[28] This gives rise to a growing need for more diverse types of information such as socioeconomic, environmental, and sustainability indicators.[29] While the use of diverse data and information provides opportunities for informed decision making, it also presents challenges.

Several distinct types of data are important to understanding water quality and quantity, including data to characterize the system, and data to monitor the system over time. The former is collected over a shorter period of time, often at finer spatial or temporal resolution, with the distinct purpose of parameterizing the conceptual or numerical model of the physical system. The latter is

SIDEBAR 2. MANY Vs OF BIG DATA[30]

Volume	The amount of digital data is growing rapidly.
Variety	Different forms and formats of data challenge effective curation.
Velocity	The speed at which data are generated or streaming continues to increase.
Veracity	Biases, noise and abnormalities in the data grow with increasing automation.
Validity	Accuracy of data for intended uses requires evaluation and validation.
Volatility	Length of time that data are valid and should be stored is increasing.
Vulnerability	Susceptibility of data to loss or intrusion is high.
Versioning	Management of multiple iterations of a data source or product is challenging.
Value	It is challenging to value the importance of data for intended and secondary uses.

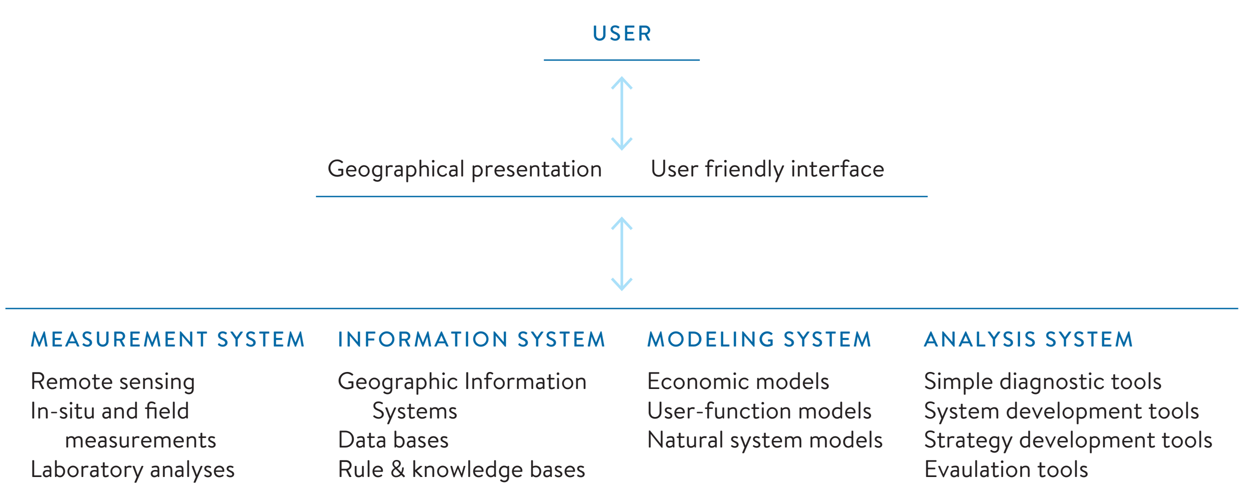
operational (i.e., continuous and ongoing) data acquisition to evaluate the current state of the system and to inform decision making. The results of operational data collection can also be used to validate or improve the initial parameterization.

Water data faces challenges related to the "Big Data V's" (Sidebar 2). Variety, for example, presents particular challenges for water data. Water data are often obtained using a variety of sensors that sample different properties. Some measurements are obtained directly and reflect local conditions (such as water levels near a wellbore), while other attributes are calculated using a suite of measurements. However, this variety of data may be collected at different scales and resolutions, raising challenges for interoperability.

Information technologies (IT), including new and existing technologies, can help incorporate scientific data into decision making.[31] Geographic information technologies—including Geographic Information Systems (GIS), Global Positioning Systems (GPS), remote sensing, and cloud-based tools—can help collect, aggregate, organize and share data. Wireless sensor networks can enable greater precision in hydrologic measurement.[32] These approaches require collective and collaborative sharing of information and re-thinking of processes to take account of new technologies.[33] Some of these technologies are already in use, but an 'innovation deficit' remains: there is still huge potential to unlock further innovation through the integration of information into decision making.[34]

Decision support systems (DSS)—modelling tools used by scientists, managers, and stakeholders—can help guide management decisions by formulating problems and evaluating solutions.[35] Decision support systems are a way of processing data into information: they incorporate diverse data sources in order to inform decisions,[36] and bring analytic functions together to help a user make sense of complex data and systems (Figure 5).[37]

Figure 5: Common components of decision support systems.



B. SCIENCE FOR POLICY

Data-driven decision making implies a tight relationship between science and policy. Scientific observations help to define the scope of environmental problems, the agenda for debate, and the range of alternative solutions to be discussed. While management decisions and policy questions are not fully solvable through purely scientific processes,[38] science can aid in ensuring that decisions are based on accurate information, clarifying the application of general goals to specific contexts, and ensuring that agencies act in the public interest.[39]

Many types of decision makers need data and information. Regulators need sound information in order to quantify and manage risk, and more accurately apply existing laws.[40] Managers of utilities, infrastructure, and water agencies need information to inform their daily operations and long-term investments. Nongovernmental organizations and the general public also need information in order to engage with environmental protection and other aspects of water resources.[41] In water resource management, many parties have interests in outcomes, and all stakeholders require accurate information to participate meaningfully in decision making processes.[42]

"**Credibility** involves the scientific adequacy of the technical evidence and arguments.

Salience deals with the relevance of the assessment to the needs of decision makers.

Legitimacy reflects the perception that the production of information and technology has been respectful of stakeholders' divergent values and beliefs, unbiased in its conduct, and fair in its treatment of opposing views and interests." [43]

Practical application of science requires effective use of information.[44] Scholarship on the use of science and IT in policy suggests that data must be salient, credible, and legitimate[45] in order to be useful and usable.[46] While "usefulness" refers to functionality and desirability, "usability" refers to application and the ability to fit decision making processes in practice.[47] That is, to be not only useful but also usable, data must be readily available in formats that suit users' needs for application to decision making contexts.

Achieving credibility requires data to be produced and stewarded according to best practices. Achieving salience requires first understanding and articulating the intended purpose of data and information, and then designing data systems to serve relevant decision making. Beginning from the questions that are of concern to stakeholders ensures that relevant data are collected and synthesized into information. Starting with user needs in this way can help to avoid the costly need to re-engineer the system after it has been developed. Achieving legitimacy is fostered by involving end users in a two-way conversation, providing mechanisms for bridging boundaries between the scientific and decision making communities, and developing end-to-end networks to ensure that scientific information is truly useful for decision makers and policymakers.

4. A PROCESS FOR ENGAGING STAKEHOLDERS IN ASSESSING WATER DATA NEEDS

In this section we describe a process for developing use cases through stakeholder engagement. For the purposes of this report, use cases describe water decision making processes and the data needs associated with those processes. The premise of use cases is that starting with the end user's goals can enable a more efficient and effective development of a data system. In the field of computer science, use cases are employed in software and hardware development describe system requirements from a user's perspective.[48] The approach puts the user front and center in the development process.

The main goals of developing the use cases were a) to systematically assess end user data needs; b) to evaluate the fit between data needs and existing data availability; and c) to communicate data needs to technical developers tasked with design of data systems and user-oriented applications. Taken collectively, the use cases can assist in identifying commonly used data sets where interoperability is particularly important, as well as gaps in usability or accessibility. The process also formalizes stakeholder engagement.

We define a use case as a set of answers to the questions of *who* needs *what data* in *what form* to make *what decisions*, framed around a single decision.

A. DEVELOPING USE CASES THROUGH STAKEHOLDER WORKSHOPS AND COLLABORATIONS

We developed the use case concept and format in an iterative process over the course of several day-long workshops and through additional meetings and interviews.[49] We defined stakeholders broadly as those with an interest in the outcomes of California's progress on water data. Invited attendees included data users and data producers from a variety of sectors, including academics and scientists, state agency representatives, NGO representatives, community members, the private sector, and water management practitioners, all of which were represented in workshops, subsequent discussions, and review of this document.

Use cases were framed around the questions of *who* needs *what data* in *what form* to make *what decisions*[50] A use case is therefore an internally consistent set of answers to these questions, anchored around a single decision. We developed a template (Table 1) to guide stakeholders in articulating use case elements. The template was developed in collaboration with technical data system developers to ensure that necessary information would be collected and conveyed. Using this template, a total of 20 use cases were developed. Eight use cases were produced through the workshops described above, which also served the purpose of developing and testing the use case concept and methodology. The Center for Law, Energy and the Environment (CLEE) also facilitated additional meetings and supported other organizations in contributing use cases using this template (available online at law.berkeley.edu/datafordecisions).

The use cases are being used in the development of functional requirements that will in turn inform the protocols called for in the legislation.[51] Functional requirements represent the translation of objectives—the stakeholder-generated goals defined through the use cases—into engineering terms and technical language describing how the objectives will be met (Figure 6). The goal of translating stakeholder objectives into functional requirements is ultimately to inform development of data system protocols and design in ways that increase usability.

B. SUMMARY OF USE CASES

By covering a diversity of topics and ways of using data to meet objectives, the use cases together illustrate a range of ways in which data are used for decision making in the water management sector. Table 2 provides a summary of the use cases developed and analyzed for this report. The full text of all of the use cases, including a description of decision makers and decision contexts for each, is available online at law.berkeley.edu/datafordecisions.

Table 1: Use case template

USE CASE ELEMENT	DESCRIPTION
Objective	Decision, goal or desired action
Description	Important context and background information
Participants	The main decision maker; other parties involved or affected
Regulatory context	Legal, regulatory, and reporting requirements
Workflow	Progression of steps and specific actions taken to accomplish objective
Data sources	Existing data sources; also data gaps
Data characteristics	Notes about type and format of data

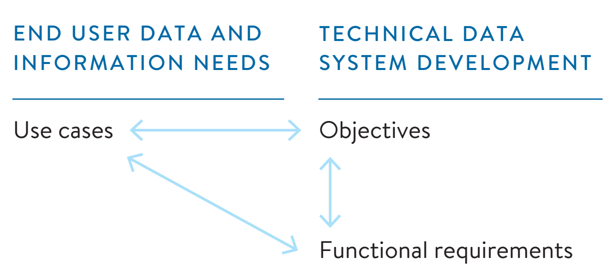


Figure 6: Relationship between stakeholder use cases and functional requirements.

Table 2: Summary of use cases and objectives. Use case numbers are based on the order of initiation during our process.

USE CASE TITLE	OBJECTIVE
1. Planning a groundwater recharge project	Determine when, where, and how to recharge groundwater, with what water, in order to avoid undesirable results including declining groundwater levels.
2. Financing a groundwater recharge project using Proposition 1 funding	Maximize expected return on investment for a groundwater recharge project financed under Proposition 1, based on an evaluation of costs, benefits, risks, and expected return on investment.
3. Management of environmental flows to protect salmon habitat	Manage environmental flows for winter run Chinook Salmon in the Upper Sacramento River through reservoir management.
4. Groundwater basin water budgets (DWR)	Quantify inputs, outputs, and changes in storage (i.e., water budget) within the basin, at appropriate spatial and temporal scale and resolution, with accuracy sufficient to inform groundwater management and pumping.
5. Delta hydrographs	Establish thresholds (e.g., mean sea level; levee conditions; water quality) as a proactive alternative to Temporary Urgency Change Petitions for management in the Delta.
6. Water transfers for environmental purposes	Assess a plan for a 1-year instream water transfer on a small stream in upper Sacramento River watershed. The goal of the water transfer is to maintain or enhance instream flows for environmental purposes.
7. Capital investment in headwaters restoration	Restore and maintain headwaters in a sustainable condition by building public support and knowledge, and quantifying the suite of benefits from headwater management and restoration.
8. Wetland and riparian mitigation and monitoring	Produce "State of the State Wetlands report" to provide regional estimates of the ecological integrity and biological conditions of wetlands, to ensure no overall net loss and achieve a long-term gain in the quantity, quality, and permanence of wetlands acreage in California.
9. Central Coast Ambient Monitoring Data Navigator	Acquire data to populate the CCAMP Data Navigator, an online data visualization and analysis tool used by Water Board staff, decision makers and the public to inform them about water quality status and trends.
10. Urban Water Efficiency Explorer tool	Provide a data visualization and scenario planning tool to help California water retailers estimate residential water efficiency targets in order to visualize the changing water conditions to enable them to effectively make decisions about adaptations.

11. Sacramento River real-time water and fishery coordination decision platform	Integrate flow, water operations, fishery, and water quality data into a single, open data platform that improves Sacramento River operation of the CVP and delivery of flows for fishery temperature management and downstream diversions.
12. Water availability analysis for curtailments to protect senior water rights	Determine, based on a drought water availability analysis, at what time and to whom notices of water unavailability (also called curtailment notices or water shortage notices) would be issued to protect senior water rights.
13. Water rights licensing process	Determine the circumstances under which the State Water Resources Control Board (SWRCB) can make the necessary findings to issue a license for a permitted consumptive use of water for a minor agricultural project.
14. Water shortage contingency planning vulnerability assessment	Conduct a vulnerability assessment of a rural community in the San Joaquin Valley to determine the extent to which water shortage puts it at risk of not having sufficient clean water supplies for household use, including consumption.
15. Decision support system for harmful algal bloom response, communication and mitigation	Effectively manage and utilize data regarding harmful algal bloom (HAB) incident verification, communication and mitigation to support and inform decision making.
16. Decision support system to track and evaluate mercury control actions	Implement mercury control actions to maximize effectiveness of reducing exposure to humans and wildlife. Evaluate the potential of wetland restoration, salmonid population restoration, and other on-the-ground projects to increase mercury exposure to humans and wildlife.
17. Groundwater basin water budgets (SWRCB)	Quantify groundwater basin water budgets so that SWRCB can determine: 1) whether probationary designation is required; 2) whether an interim plan is required, and if so; 3) how to manage extractions within the basin, consistent with established Water Rights law, so that the basin progresses toward sustainability.
18. Agricultural water management plan	Improve water management and water use efficiency, as well as plan and prepare for periods of limited water supply and severe drought by developing an Agricultural Water Management Plan (AWMP) to serve as a water management planning tool within an agricultural water supplier's service area.
19. Urban water management plan	Document current and future water supply reliability through the preparation of Urban Water Management Plans (UWMPs), to ensure that California communities have adequate water supplies especially in times of drought.
20. Source-water basin water budgets	Quantify inputs, outputs and changes in storage (i.e. water budget) within a basin, at appropriate spatial and temporal scale and resolution, with accuracy sufficient to inform reservoir operations, hydropower generation, downstream water deliveries, allocation decisions, infrastructure investments, flood protection, groundwater recharge and other economic and regulatory decisions.

The sample of 20 use cases is by no means intended to comprehensively cover the vast landscape of decisions in California water management. The use cases do, however, exemplify the complexity and range of water management contexts in California. An initial set of topics selected for development in the workshops were augmented by additional topics, which were prioritized to cover a wide range of subjects across California's water management landscape (Figure 7). The categorization of use cases by topic in the Venn diagram in Figure 7 represents only one possible way of doing so; cases could be categorized according to other topics as well.

Additionally, Figure 7 illustrates the variety of decision objectives represented by the use cases. Some of the cases address high-level investment and policy decisions; others focus on mid-level programmatic implementation; and still others address day-to-day operational decisions. Likewise, some of the cases draw directly upon data to answer a question or inform a decision, while other cases involve integrating disparate data sources for a decision support tool that, in turn, can be used to inform decisions or answer questions.

The sample of use cases also align with broader goals for California water management. The Governor's California Water Action Plan[52] is intended as a roadmap towards sustainable water management. Alignment between the use case topics and the Water Action Plan goals is illustrated in Figure 8.

C. METHODOLOGICAL LIMITATIONS

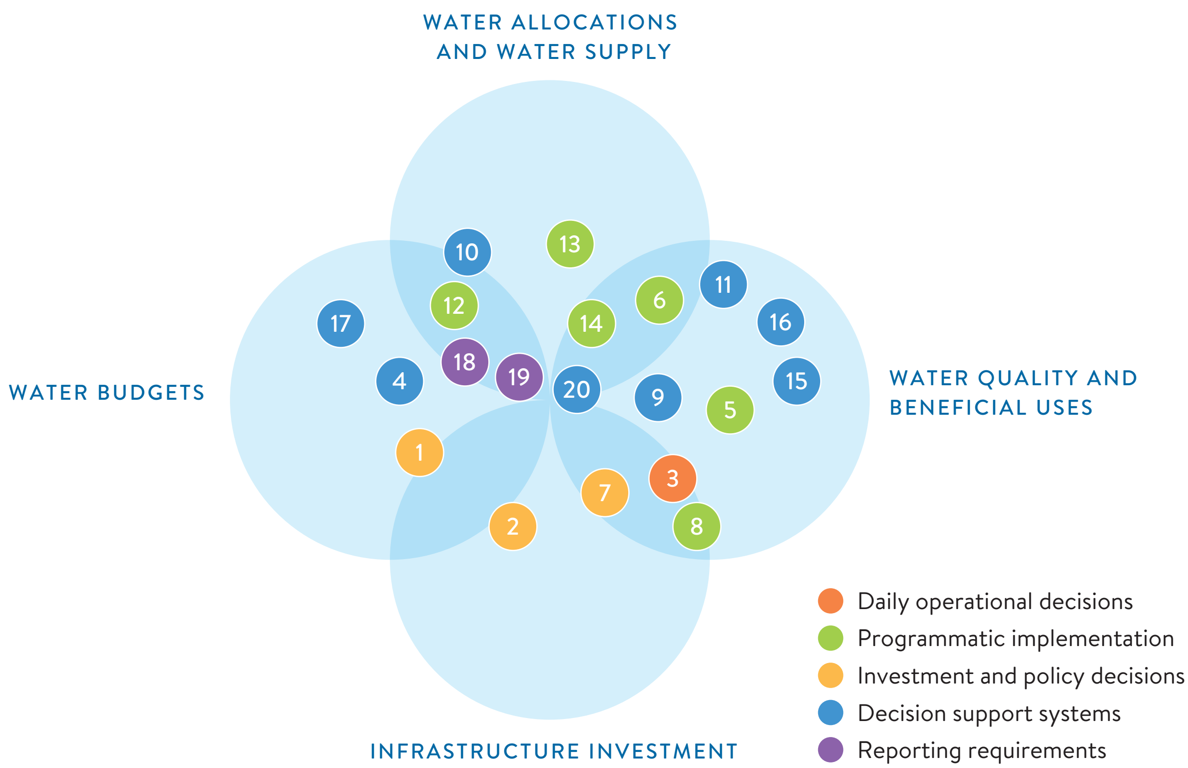
The process of identifying stakeholder objectives and translating these objectives into functional requirements comes with some limitations and caveats.

First, the 20 use cases represent a small sample among the many water-related decisions made in California each year. The aggressive timeline of AB 1755, while helpful in motivating stakeholder engagement, limited the feasibility of conducting multiple iterations and the total number of use cases. As such, the following analysis may not be representative of broader water management data needs. Rather, the following analysis illustrates some general trends and provides broad and potentially informative observations, not statistically significant conclusions.

Second, in our analysis, we did not fully characterize data sources by usability. For example, we did not examine whether the data sources provided for a use case were interoperable, whether temporal or spatial resolution matched decision maker needs, or whether they were accessible in a format suitable for analysis. As a result, a closer analysis of use cases may find that some stakeholder-listed data sources are actually low in usability, requiring additional processing to render them interoperable and useful as part of a data system. Further analysis could address the extent to which less-than-adequate data sources are currently being utilized, and what this means for the information that is produced using this data.

Third, the method of developing use cases was intended as the start of an ongoing and iterative process. We refined the method throughout the project, and as such, several challenges and limitations were raised during the course of the project. The process revealed that decisions are not always obvious or easy to articulate, and in a group format, agreeing upon the most important decisions was time consuming. Moreover, while our process focused on identifying 'decisions,' not all data uses necessarily take this form; for example, research may draw upon data sources to produce useful information that can inform decisions, but in and of itself does not represent a 'decision' per se. We learned that engaging stakeholders with a detailed understanding of workflow and data needs is crucial, as the use case format depends upon clear working knowledge of data sources (both existing and desired/gaps) required to make decisions. Sufficient buy-in to the process and outcome matters: clearly communicating purpose to stakeholders is important, and quality of facilitation can influence the quality of use case results. Finally, use cases were each generated by a different group of stakeholders, creating some inconsistencies in format and style.

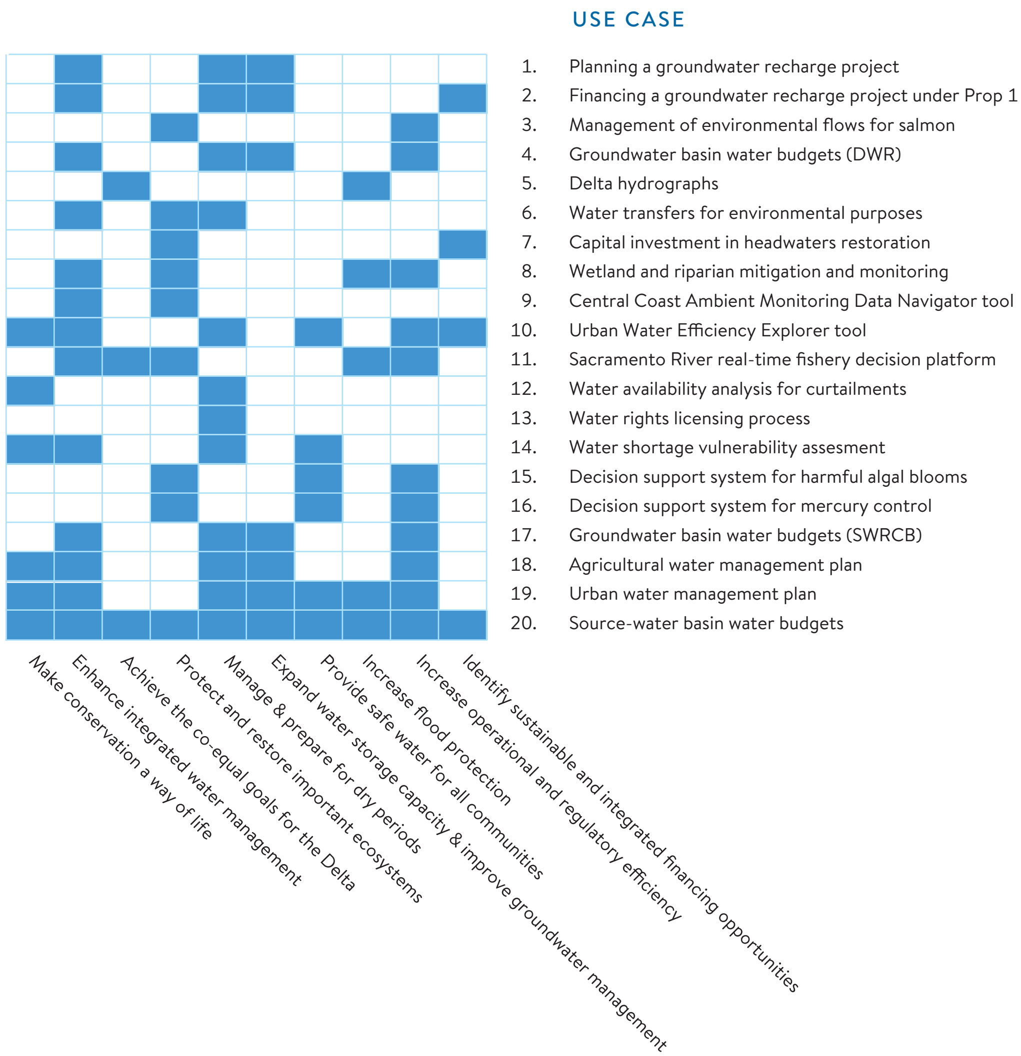
Figure 7: Classification of use cases by topic and decision objective



USE CASE TOPICS

1. Planning a groundwater recharge project
2. Financing groundwater recharge under Proposition 1
3. Management of environmental flows for salmon
4. Groundwater basin water budgets
5. Delta hydrographs
6. Water transfers for environmental purposes
7. Capital investment in headwaters restoration
8. Wetland and riparian mitigation and monitoring
9. Central Coast Ambient Monitoring Data Navigator tool
10. Urban Water Efficiency Explorer tool
11. Sacramento River real-time fishery decision platform
12. Water availability analysis for curtailments
13. Water rights licensing process
14. Water shortage vulnerability assessment
15. Decision support system for harmful algal blooms
16. Decision support system for mercury control
17. Groundwater basin water budgets (SWRCB)
18. Agricultural water management plan
19. Urban water management plan
20. Source-water basin water budgets

Figure 8: Alignment between Water Action Plan goals and use case topics



5. DATA AVAILABILITY, SOURCES, AND GAPS

Examination of use cases revealed preliminary insights into data availability and data gaps. In this report, data source refers to an individual unique access point (for example, a URL) for data. A complete list of data sources cited is available online at law.berkeley.edu/datafordecisions. Data sources were compiled and categorized by topic and by organization providing the data source. Data gaps and limitations were also identified. Results from these analyses are presented below. Throughout this section, it is crucial to recognize that the results describe general trends based on an examination of a limited number of use cases, rather than statistically meaningful analysis.

A. ASSESSING DATA NEEDS

Coding data sources cited in use cases by topic (Table 3) allowed for several observations.[53] First, data sources were diverse, spanning topics far beyond those directly related to the hydrologic cycle, such as precipitation and streamflow (Figure 9). Second, each use case drew from a variety of data spanning a range of topics to meet the stated objective (Figure 10). Third, while a few core data sources were common to many different use cases with markedly different objectives, the majority of data sources were only used in a few specific contexts (Figure 11).

Over half of the data sources cited in the 20 use cases were related directly to water, including water supply, demand, groundwater, water quality, and water storage (Figure 9). However, data on a wide variety of other topics were also often needed to support decisions. This suggests that a water data system will need to incorporate not only water data, but also other relevant data in order to fully serve the purpose of supporting water decisions.

The majority of use cases required data sources spanning diverse topics, regardless of the decision objective (Figure 10). This supports the argument that there is a need for data systems to not only focus on interoperability, a crucial point recognized by the authors of AB 1755, but on a broadly defined notion of water data. A system with high interoperability between different data sectors and topics could increase the ability of decision makers to integrate data from multiple data sources and sectors.

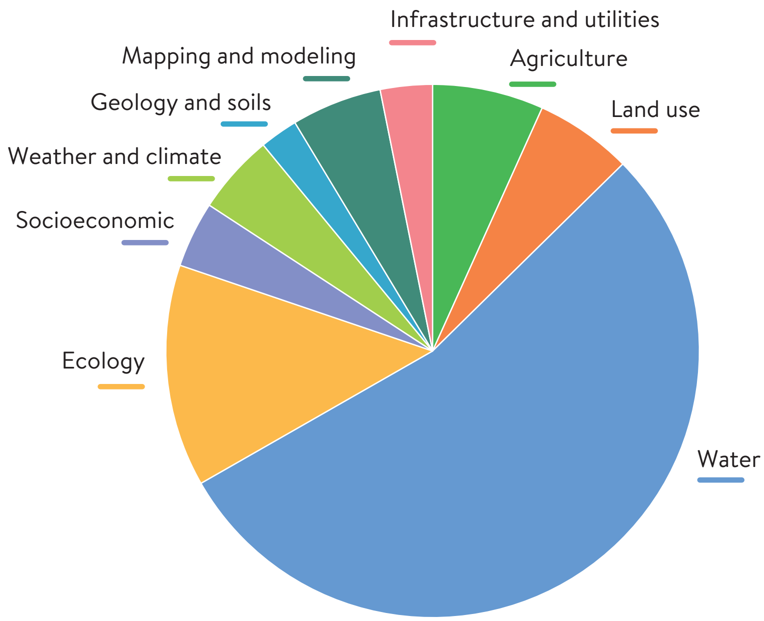
Only a very few data sources (individual URL access points) were commonly used across many use cases (Figure 11). For example, USGS streamflow was cited as a data source in 16 of the 20 use cases.[54] On the other hand, about three quarters of the unique data sources were each only cited in one of the 20 use cases.

If commonness of use reflects importance for decision making, then it is particularly important to ensure accessibility, interoperability and completeness for the very few data (e.g., streamflow and precipitation) that were common across many use cases, within a data system. However, usage across many use cases is not necessarily the only indicator of importance. Many of the less-frequently cited data sources may be crucial in certain decision making contexts. Moreover, importance of the decisions themselves varies. The wide diversity of topics and difficulty in assessing importance of data sources and of use cases themselves may present a challenge to data system designers seeking to ensure accessibility and interoperability.

Table 3: Data types and parameters included within topic categorizations

TOPIC	SUMMARY OF DATA PARAMETERS INCLUDED WITHIN TOPIC
Agriculture	Agricultural land use, crop types and acreage, evapotranspiration, maps of farmland, pesticide use
Ecology	Habitat parameters, biodiversity, fish counts, bird counts, invasive species, estimates of wildlife population and extent, forest conditions, fire vulnerability, native plants, timber harvest plans, vegetation classification and maps, aquatic resources, wetland boundaries, wetland mitigation sites
Geology & soils	Geology, hydrogeology, subsidence, soil types
Infrastructure & utilities	Service area boundaries, hydropower plans, water utility boundaries, pumping records, roads, water sources for particular communities, utility data on water and energy use
Land use	Aerial imagery, city and county land use, land cover, land use surveys, Landsat data, remote sensing data
Mapping & modeling	Watershed boundaries, surface waterways, terrain models, topographic surveys, elevation, county boundaries
Socioeconomic	Cost-benefit analyses, water pricing data, willingness to pay, economic impact assessments, policy analyses, population and demographics, population growth projections
Water	Includes subcategories of demand, supply, storage, quality, and groundwater—see below.
• Water demand & use	Water demand for different interests, water rights, water transfers, water usage, conservation, conjunctive use, urban water demand, water deliveries, imports and diversions, pump locations, per capita water use, consumptive use, beneficial use vs storage, domestic well data
• Water supply	Precipitation, hydrologic conditions, streamflow, stream gage data, hydrographs, full natural flow, flow projections, snowpack, return flows, river stages, volume, availability, water year type
• Water storage	Reservoir capacity, reservoir levels, reservoir surveys, snowpack storage, flood storage capacity
• Water quality	Water quality, temperature, Total Maximum Daily Loads (TMDLs), water chemistry, sediments, contaminants, bacteria, algal blooms, biological indicators
• Groundwater	Groundwater basin maps, elevation, models, pumping, quality, recharge suitability, storage, groundwater-dependent ecosystems, groundwater-surface water connectivity, GSA boundaries, well locations, well logs, aquifer storage capacity
Weather & climate	Temperature, weather forecasts, climate change forecasts and scenarios, climate patterns

Figure 9: Data sources cited in use cases, categorized by topic



B. DATA PROVIDERS

A relatively small number of organizations provided a large fraction of the data sources cited in the sample of use cases. Categorizing data by the organization providing each data source[55] showed that just six organizations provided approximately two thirds of the data sources cited across the sample of use cases (Figure 12).

State agencies provided approximately half of the data sources cited. Federal agencies provided over forty percent of the data sources. These two types of agencies combined make up for over 90% of the data sources (Figure 12). University and other research institutions, private, and nonprofit sources together made up the remainder of data sources cited in use cases.

While many of the state and federal data sources were cited across numerous use cases, the majority of private, nonprofit, and university data sources were more specialized, and for the most part were only used in single use cases.[56]

If a water data system seeks to support the full range of data providers contributing to California water management, then it should consider how data provisioning and integration beyond state agencies can be supported. This analysis shows that data providers span beyond AB 1755 governance partners (including DWR, SWRCB, CDFW, and CWQMC). This points to a need for developers of a data system to consider not only how the system will support these important partners, but also to take into account how other organizations (including federal organizations such as USGS, NOAA, and USDA) will participate.

Figure 10: Number and topic of data sources cited within each use case

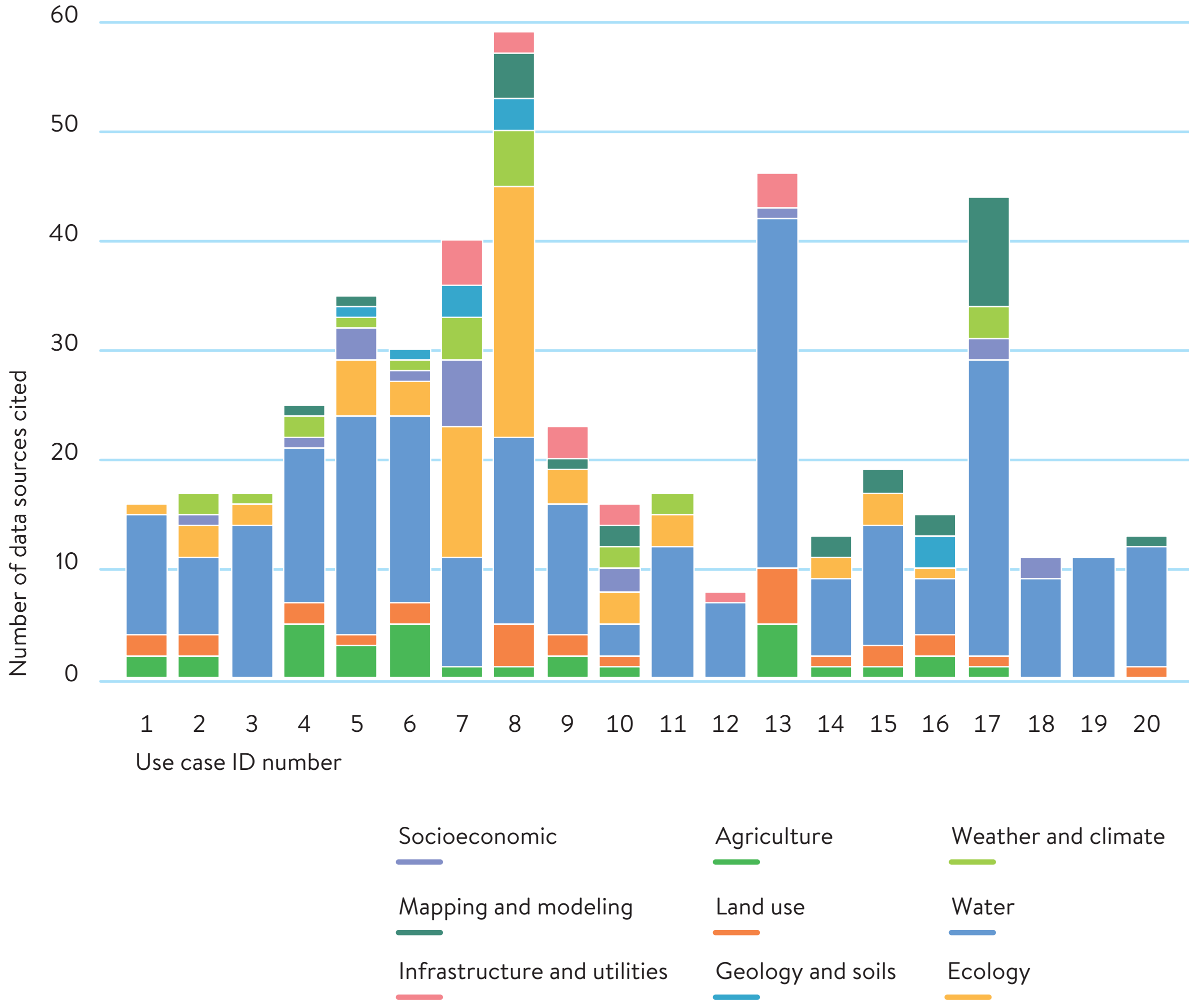


Figure 11: Percent of unique data sources used by one or multiple use cases

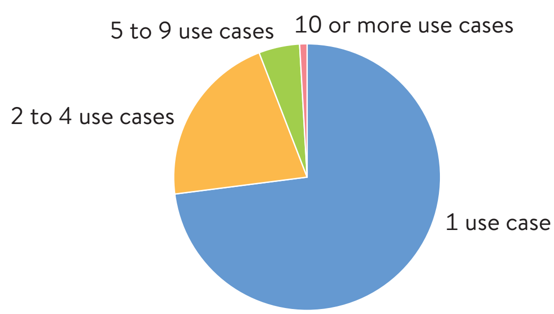
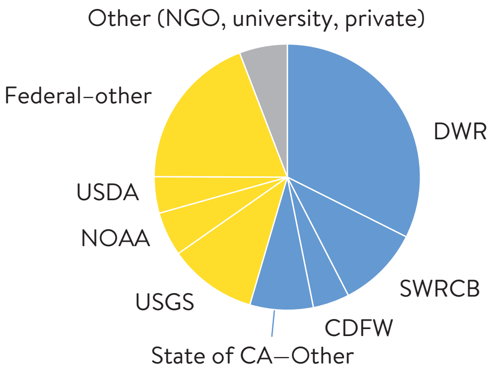


Figure 12: Percent of data sources cited in use cases provided by specific agencies (state agencies shown in blue, federal in yellow, other in grey)



C. DATA GAPS AND LIMITATIONS

Some data and information critical to decision making were not available at all, while other data and information were available, but were not interoperable or were hard to access or use.

Limitations in data and information availability were identified primarily based on stakeholder input during use case development. In our selection of use cases, commonly cited limitations included limited availability of economic information (water pricing data, economic impacts, and willingness to pay); the need for more accurate characterization of water rights; and the need for better data related to groundwater pumping and withdrawals.

Limitations in connectivity and accessibility mean that some data sources can constitute effective data gaps even if, strictly speaking, data are available (Sidebar 3).

Further analysis beyond the scope of this study could evaluate not only limitations in data availability and accessibility, but also the significance of these limitations. That is, some data limitations may represent a critical bottleneck to decision making, while other missing or limited data availability may not actively constrain effective decision making. Further analysis could also examine the costs and benefits of decision makers' reliance upon stand-in data sources that are less than ideal when the desired data are not available. Decision makers often "make do" with existing data because they need to proceed in making a recommendation or decision, even though additional or higher-resolution data would improve accuracy and analytical capacity. The use of less-than-adequate stand-in data sources to fill gaps in data availability, accessibility or usability may mask the existence of data gaps.

D. ADDITIONAL STAKEHOLDER OBSERVATIONS AND PERCEPTIONS

During the workshops, and in additional meetings and interviews, we tracked comments and responses from stakeholders. We coded and qualitatively synthesized this stakeholder input. Table 4 summarizes a number of stakeholder concerns and perceptions, sorted by themes including interoperability; data format and resolution; uncertainty and metadata; distinctions between data and information; and potential barriers to an open and transparent water data system.

SIDEBAR 3: ASSESSING DATA USABILITY

A selection of data sources were examined in depth to better understand the process of developing technical parameters out of the stakeholder-generated use cases. This exercise was meant as a starting point for the development of functional requirements for a data system. The analysis found that, when examined in closer detail, a number of data sources listed in the use cases did not, in their current form, contain the requisite parameters for integration into a data system. We identified several themes that presented challenges for integration of data:

- Some data sources were logistically difficult to access:
  - Data were accessible via maps or visualization tools for public use, but downloading data to create those visualization tools was not possible or unclear.
  - Data were only available in PDF formats.
  - Data were accessible only using an institutional login name and password to access.

- Some data sources lacked precision in definition:
  - The need to include "physical habitat parameters" was mentioned, many of which are available via US Fish and Wildlife Service, but it was unclear which habitat parameters were needed.

- Some data sources represented protocols and methodologies to gather data; the data themselves were not readily accessible:
  - An academic article described a method for data collection, but did not provide the data themselves.

- Some data sources lacked consistency in terms of availability and spatial and temporal resolution:
  - Reports were made available infrequently or inconsistently.
  - Spatial or temporal resolution of data was limited.
  - Parts of data were available digitally while older versions were only available as PDF.

These practical issues of data accessibility and interoperability present challenges for an integrated water data system because they hamper automated integration of data from multiple sources.

Table 4: Stakeholder concerns and perceptions

Interoperability	• Data standards vary across different agencies and organizations. • Multiple entities have their own processes for data collection, storage, and metadata documentation. • Data sources are not always connected to one another, nor to regulatory activities or decisions. • Data must be comparable and connectable across different agencies and different geographic locations.
Data format & resolution	• Not all data are digitized. • Data should be in a format that can accommodate a range of post processing and analysis. • Data may be collected at different resolutions than decisions call for (e.g., some data are collected seasonally, monthly, or daily; however, data collection must ideally be real-time in order to support decision making). • Different decisions require different resolutions of data.
Uncertainty & metadata	• Data are laden with uncertainties; it is important to reduce uncertainties where feasible, and then to characterize remaining uncertainty. • Describing uncertainties makes it possible for a decision maker to decide whether the uncertainties hamper decision making capacity. • Documenting uncertainties can help communicate not only what data can be used for, but what data should not be used for. • Quality assurance and quality control procedures must be documented. • Ensuring consistent and complete metadata is an important part of addressing concerns around uncertainty.
Transforming data into information	• Information gaps exist where relevant data may be available, but processing and synthesis to make data more useful have not yet taken place (for example, information on long-term historical impacts of groundwater pumping; or a system for tracking net loss of wetlands over time). • Transforming data into information requires knowledge of system workflows and tools that are used to inform decision making, such as graphs, figures, or reports. • Models are examples of useful decision support tools that connect multiple types of data. Hydrologic and groundwater models, while not strictly necessary for all types of decisions, can be a useful way to integrate data on precipitation, streamflow, evapotranspiration, soils, land use, and climate.
Potential barriers to an open and transparent water data system	• Time lags in data review and availability. • Issues of outdated technology. • Issues around sharing of data, including legal, privacy, and trust. • Cross-sector communication issues. • Lack of clarity regarding allocation of resources and responsibility between agencies for various aspects of data collection and provisioning.

6. LESSONS FOR OPEN AND TRANSPARENT WATER DATA

This section discusses lessons and considerations for open and transparent data system development. The material in this section emerged from the process of developing and analyzing use cases, as well as from review of literature and experience in other contexts. Lessons include (a) basic cyberinfrastructure considerations, and (b) stakeholder engagement and data system governance considerations.

A. BASIC CYBERINFRASTRUCTURE CONSIDERATIONS

Interoperable water data systems

Water management involves a wide variety of decisions drawing from many different forms of data and information. If the California water management sector is to reach a point where decisions are informed by the synthesis of disparate sources of data, then information infrastructure needs to ensure interoperability between existing data sets and systems.[57] This means that data themselves must be interoperable. Currently, this is not always the case. Data are often collected for a particular purpose without consideration of potential other uses or how they may intersect with other data sets. For example, DWR and USGS collect groundwater data from the same wells but use different well station identifiers for the same wells, reducing the ability to compare the two data sets.

Because decisions draw on a wide variety of data sources, interoperability is important among disparate data topics (i.e., not only data on flows and quality of water, but also ecological data, agricultural data, etc.). While many decisions use a few core data sets, they combine the data with other, more specific data sources in different ways depending upon context. Moreover, a few agencies are sources for a large fraction of those data; robust interoperability within and between these organizations is particularly important. Integrating data and data providers into an interoperable system will require a flexible or possibly phased approach.

Considering federated data systems as a strategy for increasing interoperability

Given the distributed nature of data, the independence of disparate agencies, and the need for interoperability, a federated data system is likely to have distinct advantages over a more centralized system. A federated data system connects multiple independent data systems through common standards and conventions, while keeping those independent systems as autonomous entities (for example, see Sidebar 4). Such systems can allow individual data providers to continue to control their own data management while increasing interoperability.[58] Maintaining control can be important for agencies and programs with mandates to collect, host, or present data in a specific way. Moreover, some agencies must steward their own data in order to maintain credibility and legitimacy.

However, federated data systems can also present challenges. Technical issues exist, given the difficulty in ensuring that data can be used beyond their original purpose.[59] Designing and enforcing versioning and metadata standards is also non-trivial. Institutional challenges

SIDEBAR 4: EARTH SYSTEM GRID FEDERATION (ESGF)

The Earth System Grid Federation (ESGF) is an example of an open source federated data collaboration sponsored by NSF, NASA, NOAA, DOE, and others. For over 10 years, ESGF has supported climate science research by providing a large-scale data management platform for observational data, model simulations, and analysis products related to Earth System Science. ESGF is a foundational data tool for the Intergovernmental Panel on Climate Change (IPCC) process among other research activities, and facilitates global data access, management, and use. The system includes decentralized "peer nodes" that are independently administered, yet are linked through common protocols and interfaces. The platform is being adapted for use in energy infrastructure and pharmaceutical development.

include potential disagreement around who sets standards; difficulty in securing buy-in from data producers; potential concerns around data ownership and control; and ensuring stable funding for data providers. Having some degree of centralized data storage can enable redundancy and backup, as well as facilitate value-added data processing, although the dispersed liability of federated data systems can also have value.

Despite these challenges, a federated data system could help ensure interoperability of data across multiple organizations. Clear standards for data quality, metadata, and technical requirements are key to the technical success of any interoperable data system. Governance mechanisms such as broad mandates for incorporating standard procedures could help ensure that agencies participate in a federated data system.

Spatiotemporal data

To be useful for water decision making, data must be spatially and temporally referenced, which presents several challenges. Differing scales and resolutions[60] of data collection and availability can hinder interoperability. For example, California's hydrologic regions are divided into Detailed Analysis Units (DAUs) by the Department of Water Resources, but these boundaries do not exactly match USGS hydrologic boundaries. This issue extends to field data collection: for example, flux tower measurements used to calculate evapotranspiration have a footprint about the size of a football field, while estimates of groundwater storage from the Gravity Recovery and Climate Experiment Satellite (GRACE) have a footprint larger than the Sacramento and San Joaquin basins combined. The variety and lack of consistency in spatial boundaries and resolutions hinders integration of different types of data.

A useful, responsive data system must consider how stakeholder data needs vary across decisions and decision makers, even when considering similar topics in a given location. For example, in the development of a groundwater basin water budget use case (Use Case 4), it became clear that the resolution of data needed to meet GSA reporting requirements differs from data resolution needs of individual groundwater pumpers. The Sustainable Groundwater Management Act (SGMA) requires groundwater basins to report annual aggregate basin-scale measurements. However, water pumping decisions by individual land managers require data at a finer spatial and temporal resolution. To address this issue, DWR is encouraging basin managers to collect data at more refined temporal and spatial resolutions to better define and meet sustainability objectives.

Another consideration regarding spatiotemporal data is that full datasets can be prohibitively large. A usable data system would ideally enable data users to retrieve data tailored to an area and time period of interest. This would enhance data usability, but presents a challenge for a federated data system because it requires data selection capacity that might be unavailable. For certain types of data (for example, very commonly used data sets, as well as very large data sets), a centralized server hosting a copy of the original data with additional search functionality might be a good solution. For other types of data, the opportunity to find data via a federation of data suppliers may better serve the needs of water managers and scientists.

Complete and consistent documentation and data quality

Clear standards for documentation can facilitate the incorporation of new data into a water data system, but developing them can be difficult and warrants care. Data must be traceable and sources clearly identified. Protocols and methods must be employed to ensure that data are properly collected, handled, processed, used, and maintained at all stages of the data life cycle. Documenting data lineage and quality is central for credibility. Documentation is also an essential component of integration of data from multiple sources. Data must be structured according to standards in order to facilitate integration and automated search capacity.[61] In the case of water data, it is also often important to include coordinates for georeferencing. Effective AB 1755 implementation may include developing new standards for documentation or implementing existing ones.[62] Archiving practices are important to protecting data against losses. One simple partial solution is the use of digital object identifiers (DOIs), unique handles used to identify digital objects (such as data sets).[63] DOIs do not change even if, for example, agency websites are reorganized, and can also assist with versioning, quality assistance/quality control, and referencing sources.

While error and uncertainties in data may persist, complete documentation that describes and characterizes uncertainties enables a user to decide whether the uncertainties are acceptable. Documenting uncertainty can be useful in communicating not only what data *can* be used for, but what data *should not* be used for. Data users have different needs regarding data quality and precision: for example, data used for regulatory compliance generally must meet agency-specified collection methods, whereas many other decisions can be based on best available data. The wide difference in user needs regarding data quality and precision highlights potential tradeoffs between cost of data collection and provision and data accuracy. Understanding user needs could clarify the costs and benefits of greater accuracy.

Bridging the gap between data and information

Ultimately information, not just data, is necessary for informed decision making. However, turning data into information can be an afterthought. For example, reporting requirements specified in regulations are often intended to ensure compliance, rather than being formulated to provide decision makers with the information they need to manage water resources. This can lead to large volumes of data but little useful information.[64]

Numerous methods exist for converting data sources into useful information products, including models, decision support systems, or other analytic tools. Hydrologic models of surface water and groundwater, while not necessary for all types of water-related decisions, can be a useful way to integrate data. While models can be developed by public or private entities, the state of California encourages open development of models,[65] and the University of California is working to create an open and organized venue for model development.[66]

Stakeholders suggested several principles for organizing and connecting data. Water budgets are needed at multiple scales and for multiple purposes, and regional approaches may also be beneficial (for example, connecting data around the Sacramento Bay-Delta, or connecting data across a watershed).

The variety of decision objectives covered in the use cases shows that different data users require data in different formats. In some use cases, stakeholders seek to develop decision support systems that will inform other decision makers. In other use cases, stakeholders seek to access data directly in order to process data into information themselves. For still other audiences such as academics and researchers, access to raw source data is key in order to produce or verify visualizations or other data products. If a data system is to facilitate the production of information, these varying needs must be taken into account. Additionally, data user needs and technological capacities change over time; the overall system must therefore closely meet the needs of users[67] but be modular enough that it can be altered as required.[68]

Considerations for open data

Open data can be of considerable benefit in efforts to make government more transparent, accountable, and efficient.[69] Openness and transparency are important elements in water governance systems that seek to involve multiple stakeholders,[70] and AB 1755 explicitly makes these ideas part of the approach to manage California's water resources.

Open data can ideally provide significant assistance in citizen empowerment.[71] However, best practices in open data involve more than just making existing data more readily available. Without adequate processing capacity or decision support systems, many individuals and groups may not have the capacity to use large volumes of data effectively. There is therefore a need to consider access and usability issues, including the capacities of community and civil society groups to make use of large volumes of data. There may be a need to engage in outreach with the public and provide ongoing technical assistance.

Implementation issues may exist for open data. Data may carry legal restrictions or concerns around intellectual property, privacy or potential liability. There are costs to providing data, and to meeting metadata standards for effective data sharing. Data can also be misinterpreted by secondary users. Data providers may see the opening of their data as the loss of a potential source of revenue.[72] These concerns and other potential barriers to successful implementation of an open data system should be considered.

B. STAKEHOLDER ENGAGEMENT AND DATA SYSTEM GOVERNANCE CONSIDERATIONS

Stakeholder engagement as an ongoing process

Proactively engaging data users helps to ensure that resource investments in data systems will ultimately be worthwhile.[73] If AB 1755 governance partners want to ensure that a data system will be as usable as possible, then stakeholder engagement should continue formally as part of AB 1755 implementation and future data system development.

As new stakeholder needs and new technologies arise, a data system must take these changes into account to remain useful. Developing a responsive data system includes ongoing analysis of stakeholder objectives. The process of generating stakeholder objectives, translating objectives into functional and technical requirements, and informing development of data systems, should be repeated in numerous iterations over time. Data development can be conceptualized as an adaptive management cycle involving multiple iterations of planning, implementation, and evaluation.[74]

Stakeholder engagement is important, but not easy. Several lessons emerged from our process. First, stakeholder engagement requires time, resources, and commitment, but ultimately can help inform usability of a data system. Second, engaging knowledgeable stakeholders with sufficiently detailed understanding of data needs and workflows involved in decision making is crucial to identifying key aspects of data system usability. Third, it is important to clearly communicate with stakeholders in order to generate buy-in to the process and outcome.

Our use case efforts focused on identifying "*who* needs *what data* in *what form* to make *what decisions.*" An important next step could be to define '*how* are diverse datasets used to produce *what output* needed for what *specific objective*' for select, representative use cases. In these future iterations, we envision state agencies and data users working together to refine use cases that reflect the specific analytical needs of decision makers. Such refined definitions could then in turn help inform data architecture design.

Long-term investment and funding

Developing a water data system that serves stakeholders well is no small undertaking. Data systems are complex, and development of a useful system requires adequate funding for the institutions responsible for developing, processing, housing, maintaining, and analyzing data. Data system development and maintenance is a long-term project that will require sustained investment of resources. Water data and information activities are often funded as a part of other projects targeting a different outcome. This funding model can be inefficient and ineffective. Water data and information systems are important enough to warrant independent funding. Data and information cost money both up front and over time, even though the real value of some water data may not become apparent until an infrequent event such as heavy precipitation, drought, or an infrastructure emergency.

Investment in data collection, monitoring, processing, and sharing is important for improving water management in California—as long as that investment is targeted toward meeting the needs of decision makers.[75] Each new technological advance that is proposed should be carefully assessed in a long-term decision making context. Technological advances will require stakeholder capacity and buy-in to justify their cost. Stakeholder engagement is itself also a resource-intensive process that requires funding.

Funding is important along the entire information pipeline, from data collection to quality control to analysis to archiving. Ensuring adequate funding for entities producing core data is crucial to the stability and utility of a data system. At the same time, investment in the less frequently used, more specialized data providers must be considered, since these data may be important in particular decision making contexts. Finally, investment in addressing gaps and limitations in data availability, accessibility, and usability is important.

Data producers may well need incentives to participate in a system if adhering to protocols carries costs that outweigh perceived benefits. These incentives could come in the form of financial support or regulatory mandates. "Intervener funding" (financial support that helps stakeholders to effectively participate in agency proceedings) may be a useful mechanism to support engagement of data users and producers.[76] Some mechanisms for participation already exist: for example, AB 1755 requires recipients of state funds for research or water-related projects to adhere to protocols to be eligible for state funding.[77] Moving forward, California may consider a requirement similar to the National Science Foundation's requirement of data management plans for documenting and archiving data collected during a study. The state could likewise require all data collected during state-funded projects to be made available to the public in an interoperable format.

More broadly, generating a sustainable funding source for a water data system must be given further consideration. A number of funding models exist, but any model must be carefully thought out with special consideration given to issues of equity and openness of data systems.

Importance of good governance

In addition to the technical development of information technology tools, an effective data system requires robust governance. An insufficiently strong institutional framework runs the risk of expending resources on a system that may ultimately fail due to lack of coordination, lack of investment, or lack of trust and buy-in. Institutions are important in setting priorities in a cognizant fashion, encouraging creative research, building cross-disciplinary collaborations, and ensuring that once information is gathered, it is archived and accessible.[78]

However, developing and managing a cross-organizational network for the sharing, processing, and archiving of data is not straightforward.[79] Currently, institutional and cultural barriers may stand in the way of an improved data system. These barriers include potential trust issues around sharing of data, as well as communication issues—for example, terminologies may differ across different sectors and must be clarified. An institutional framework for an effective water data system for the state of California will need to take into account issues such as trust, confidence, and credibility, which are key to the ongoing work of water resources management but are not easily resolved by technical fixes.[80]

A number of factors must be considered for institutional arrangements. Agencies must articulate a clear purpose, commit to building long-term partnerships, adopt methods to manage complexity, build teams with relevant expertise, organize a variety of resources, and engage relevant stakeholders.[81] Institutional frameworks will need to take into account information needs, the environment within which each organization operates, the diversity of operating processes and expectations, legislative contexts, and the importance of individual managers.[82] Cross-institutional agreements should be structured in a way that allows for flexibility and takes into account political, legal, financial, and technological constraints.[83]

Best practices for building robust institutional frameworks for data systems include stable funding, long-term commitment by lead agencies to the processes of shared data management and preservation, as well as clear standards for data.[84] Champions are needed across a variety of levels to successfully create organizational change and to push the project forward across the board. As we have maintained throughout this report, a strong institutional framework should include a plan for ongoing stakeholder communication and engagement.

7. LOOKING FORWARD: RECOMMENDATIONS FOR IMPLEMENTATION OF AB 1755

AB 1755 is an important step towards developing an open and transparent water data and information system for California. However, by itself the bill does not ensure that such a system will improve water decision making. AB 1755 is a start but not an end point. Working beyond the letter of AB 1755 towards an open and transparent water data system for California will require not only careful consideration of technical cyberinfrastructure, but also good governance.

By emphasizing water data system governance, we emphasize that data systems are about more than database design. Data provisioning involves multiple steps and stages, including production, curating, quality assessment/quality control, security, and access. Transforming data into information requires the development of workflows and tools that can integrate or perform analytics on diverse datasets to create specific output that can be used to guide decision making. Throughout these multiple steps, transparency must be preserved and goals of interoperability and usability must be maintained. Along each step, end user needs must be carefully considered. Over the long term, successful data systems require ongoing, stable funding, along with institutional commitment.

The 'decision first' model that we describe in this report contributes to a stronger understanding of what a data system needs to do to be useful. A data system must be responsive to stakeholders' needs in order to support the decisions made by water managers and other decision makers. A data system that cannot serve decision makers' needs can be at best of limited utility, and at worst a waste of resources in a funding-constrained environment.

Based on the analysis of use cases, we reiterate our considerations and recommendations for the development of an open and transparent water data system. If the state of California is to successfully navigate its opportunity to develop a robust water data framework, it should incorporate the following considerations into its planning.

1. **To ensure relevance, an understanding of the way data is used in decision making should guide the development of data systems.** Achieving relevance requires first understanding the intended purpose of data, and then designing data collection, provisioning, and analysis systems to serve decision making. Beginning with the questions that are of concern to stakeholders, and incorporating a recognition that these questions will evolve over time, can ensure that relevant data are collected and useful analyses are developed. This can help to avoid the costly need to re-engineer the system after it has been developed. Involving end users in a two-way conversation can help to ensure that scientific information is truly useful for decision makers and policymakers.

2. **A wide variety of data must be highly accessible and interoperable to serve many different contexts.** Water decisions in California draw from a wide range of data types and data sources, including not only data about the flows and quality of water, but also agricultural, land use, and ecological data, among many other classes. A system with high interoperability among different data sectors and topics could increase the ability of decision makers to integrate data from multiple data sources and sectors. Measurable metrics of data interoperability and accessibility should be used to gauge success.

3. **Data gaps take a variety of forms, and will need to be filled in distinct ways.** Some data and information critical to decision making are not available at all and will need to be estimated or collected in order to move forward. Other data and information are available, but are not in a suitable format to be useful for decision making processes—for example, data are not interoperable or are hard to access or use. The latter can form an equally significant barrier to informed decision making. Thus, the state's efforts to

identify and address data limitations will need to take into account not just presence of data, but also accessibility and interoperability.

4. **For California water, an integrated data system will need to connect data from multiple independent sources, while keeping those independent systems as autonomous entities.** Water data are produced and distributed by numerous state and federal agencies. Given the distributed nature of water data in California, the independence of disparate agencies, and the need for interoperability, a federated data system that enables exchange across distributed data sources is likely to have distinct advantages. However, the institutional, financial, and computational costs and benefits of any approach should be considered carefully. Any form of data system integration must provide clear standards for data quality, documentation, and archiving. To facilitate data integration, protocols and methods must be employed to ensure that data are properly collected, handled, processed, used, and maintained at all stages throughout the data life cycle.

5. **A water data system must address needs for data at multiple resolutions, and in multiple distinct forms and formats.** Analysis of use cases reveals that different decision makers have varying data and information needs, even on similar topics. Different formats and resolutions of data and information are necessary for different decisions, and a useful data system must take these needs into consideration. Enabling flexibility of data uses, while maintaining data quality and integrity, are non-trivial but crucial challenges.

6. **A water data system must enable the production of information.** Ultimately, the goal is not only data provision, it is enabling the production of information (data that have been processed in such a way as to be useful). To this end, decision makers need data sources that can be readily integrated with one another, that are consistently updated with quality data, and that can support specific outputs such as analytical tools to guide informed decisions. Many models and decision support systems that process data into information incorporate a range of disparate data sources, which is particularly important since water resources management is often integrative by nature. At the same time, for many audiences such as academics and researchers, access to raw source data is key in order to produce or verify visualizations or other data products.

7. **Engagement between data system developers and end users is, ideally, an ongoing and iterative process.** Our use case efforts focused on identifying "*who* needs *what data* in *what form* to make *what decisions.*" An important next step could be to define "*how* are diverse datasets used to produce *what output* needed for what *specific objective*" for select representative use cases. In tandem with production of data system architecture, the state and end users of data could work to develop on ongoing process of assessing data user needs, including the specific analytical needs of decision makers, and developing a process for long-term follow up on use case implementation. Such ongoing iterative development could help ensure continuing usability of data system design.

8. **Basing water data on principles of usability and stakeholder engagement requires robust cyberinfrastructure, good governance, and stable funding.** Stakeholders described needs for interoperability, data quality, and documentation that must be addressed in the development of a data or information system or platform. Beyond these basic cyberinfrastructure properties, a data system that is sustainable over the long term requires good governance and stable funding. Specific resources should be dedicated to information management and operability. Commitment to ongoing communication with end users is fundamental to ensuring that a data system will be able to contribute to improved decision making by meeting users' needs. Water data infrastructure should be conceptualized as a complex adaptive system: it must meet the needs of users, but be modular enough to be altered as required. A functional institutional framework will require clear, long-term financial commitments by lead agencies and data providers. Both an initial investment and consistent and ongoing resources will be required.

The principle of open data is increasingly becoming a standard for state governments, and it is promising that this is beginning to apply to water data as well. Developing data systems that are not only open and transparent but also useful and usable is an ambitious but worthwhile project. Attaining these goals will require funding as well as commitment to the sometimes lengthy and time-consuming processes of building relationships and working with stakeholders. We recognize that this project is likely not possible in a single step; instead, achieving the goals outlined in this report will require adaptive management and commitment to an iterative process, with progress and improvement over time.

Evaluating success of such a project should reflect not only technical considerations, but also the importance of good governance and usability for decision making. Here, we summarize our key considerations in the form of a checklist to evaluate the success of data system implementation (Sidebar 5).

California's efforts thus far to implement AB 1755 have been significant and admirable. State agencies have taken on the substantial task of engaging decision makers in working towards a data and information system that informs modern water management—a task that we hope will continue not only as part of AB 1755 implementation, but also as a broader project. Given the extensive groundwork laid since the passage of the Act, there is reason for optimism that the law can be used as a springboard to achieve a larger and more potent vision of useful and usable water data, for the benefit of the Californians who will continue to rely on efficient and effective management of water resources.

SIDEBAR 5: CONSIDERATIONS FOR EVALUATING IMPLEMENTATION SUCCESS

There are several key questions to ask when evaluating the success of implementation of a data system or platform.[85] These questions touch upon institutional as well as technical elements.

Usability considerations:

- Is the system based on an understanding of decision making contexts and user needs?
- Do users believe the system is credible, salient, and legitimate?
- Is the system actually used in practice to inform decision making?

Cyberinfrastructure considerations:

- Are appropriate data available and accessible in usable formats?
- Are data from multiple sources interoperable?
- Are data available at appropriate spatial and temporal resolution?
- Is documentation complete and consistent?
- Does the data system support synthesis and analysis to transform data into information?
- Are the system, data, and models included regularly updated?
- Is data open and transparent?

Governance considerations:

- Are data users engaged meaningfully in an ongoing way?
- Are sufficient resources allocated to data system development and maintenance?
- Is there sufficient institutional commitment by key organizations to use and maintain the system?
- Do incentives exist to ensure participation by data providers and users?

INSTITUTIONAL ACRONYMS

CCST California Council on Science and Technology

CDEC California Data Exchange Center

CDFW California Department of Fish and Wildlife

CEDEN California Environmental Data Exchange Network

CIMIS California Irrigation Management Information System

CLEE Center for Law, Energy & the Environment

CWQMC California Water Quality Monitoring Council

DWR California Department of Water Resources

eWRIMS Electronic Water Rights Information Management System

NASA National Aeronautics and Space Administration

NOAA National Oceanic and Atmospheric Administration

SGMA Sustainable Groundwater Management Act

SWAMP Surface Water Ambient Monitoring Program

SWRCB State Water Resources Control Board UCOP University of California Office of the President

UC Water University of California Water Security and Sustainability Research Initiative

USDA United States Department of Agriculture

USGS United States Geological Survey

GLOSSARY

AB 1755: The Open and Transparent Water Data Act, legislation passed in 2016 that requires the creation, operation, and maintenance of a statewide integrated water data platform.

Data system: A software or hardware system that collects, organizes, archives, distributes, or integrates data.

Data: Quantitative or qualitative representations or measurements of basic properties of the world.

Data-driven decision making: The practice of making decisions based on analysis of data rather than experience or intuition.

Decision support system: A modelling or analytic tool used to help guide decisions by processing and synthesizing data into information.

Decision-driven data provision: An approach that recognizes that decision makers' needs must inform the development of data provisioning systems themselves.

Federation: A federated data system connects multiple independent data systems through common standards and conventions, while keeping those independent systems as autonomous entities.

Functional requirements: The translation of objectives into engineering terms and technical language describing how the objectives will be met.

Georeferencing: Associating data with locations in physical space.

Information system: A software or hardware system that supports the processing, analysis, or synthesis of data so they can be used to answer questions.

Information: Data that have been processed, analyzed, or synthesized so they can be used to answer questions.

Interoperability: The ability of computing systems to operate on the same data and obtain the same analytical analysis.

Metadata: Data that describes and gives information about other data.

Objectives: The stakeholder-generated goals defined through use cases. The goals for the data system's intended uses and outputs.

Open Water Information Architecture (OWIA): An organizing structure for an open and transparent water data system created in response to the mandate of AB 1755.

Open: The provision of access to data using open-source and open-architecture protocols and methods.

Salience: The relevance of data and information to the needs of decision makers.

Stakeholder: For this report, defined as those with an interest in the outcomes of California's progress on water data, including data users and data producers from a variety of sectors.

Usability: Data that meets the needs of decision making processes in practice. Data that are readily available in formats that suit users' needs for making decisions.

Use case: For this report, defined as an example of a water decision making process and the data needs associated with that process. An answer to the set of questions of *who* needs *what data* in *what form* to make *what decision.*

ACKNOWLEDGMENTS

We are grateful to all workshop participants for sharing their time and expertise, and for their willingness to engage in open discussion around water data systems.

Workshops were organized by Alida Cantor, Michael Kiparsky, Roger Bales, Leigh Bernacchi, Luke Sherman (UC Water), Susan Hubbard (LBNL, CCST), Amber Mace (CCST), Gary Darling, Christina McCready, and Kamyar Guivetchi (DWR).

We thank workshop facilitators, including Ariel Ambruster and Alex Cole-Weiss from the Center for Collaborative Policy, and those assisting in workshop breakout session planning and facilitation, including Michael Anderson, Martha Conklin, Paul Shipman, Abdul Khan, Jerry Johns, and Shakoora Azimi-Gaylon. We also thank additional use case contributors, including Karen Worcester, Patrick Atwater, Mark Tompkins, Nicole Kuenzi and members of the State Water Resources Control Board Division of Water Rights staff, Greg Gearheart, Erik Ekdahl, Samuel Boland-Brien, Sonia Saini, and others who have contributed to the development of use cases.

We thank John Helly for ongoing discussions on use cases and functional requirements. We thank Richard Roos-Collins, Holly Doremus, and Nell Green Nylen for early conversations on the concepts presented in this report.

This work is supported by the University of California Office of the President (UCOP) through the UC Water Security and Sustainability Research Initiative (UCOP Grant No. 13941) and by the Water Foundation.

REVIEW

We wish to thank the following individuals for their review of this report and the associated use cases:

Ariel Ambruster
Center for Collaborative Policy

Shakoora Azimi-Gaylon
Delta Conservancy

David Blodgett
United States Geological Survey

Sam Boland-Brien
State Water Resources Control Board

Matthew Correa
California Department of Water Resources

Ali Douraghy
Lawrence Berkeley National Lab

Tony Hale
San Francisco Estuary Institute

Sam Hermitte
Texas Water Development Board

Nicole Kuenzi
State Water Resources Control Board

Sara Larsen
Western States Water Council

Lauren Patterson
Duke University Nicholas Institute

Doug Rotman
Lawrence Livermore National Laboratory

Paul Shipman
California Department of Water Resources

Ate Visser
Lawrence Livermore National Laboratory

Lisa Williams
Arizona Department of Water Resources

Greg Young
Tully and Young LLC

We are grateful for the efforts of reviewers, whose constructive comments and suggestions have been carefully considered and have helped us to improve the report. Reviewers were not asked to endorse the report's conclusions or recommendations, nor did they see the final draft of the report before its release. Responsibility for its final content rests entirely with the authors, and any errors are our own.

ABOUT THE AUTHORS

Alida Cantor is an Assistant Professor of Geography at Portland State University; she was previously a Research Fellow with the Wheeler Water Institute at the Center for Law, Energy and the Environment at Berkeley Law. Her teaching and research focus on natural resource management and policy with an emphasis on water resources. Dr. Cantor holds a Ph.D. in Geography from Clark University.

Michael Kiparsky is the Director of the Wheeler Water Institute at the Center for Law, Energy and the Environment at Berkeley Law, and a co-Director of UC Water. He has worked on problems as the intersection between technical and policy aspects of water resources for over 15 years. Dr. Kiparsky earned an A.B. in Biology from Brown University and a Ph.D. from U.C. Berkeley's Energy and Resources Group.

Rónán Kennedy researches and teaches environmental law, information technology law, and the intersections between these at the School of Law at the National University of Ireland Galway. He has worked in the IT industry as an analyst/programmer, and was Executive Legal Officer to the Chief Justice of Ireland, Mr. Justice Ronan Keane. He was a Visiting Scholar at the School of Law at UC Berkeley in 2017. He holds a Ph.D. from University College London.

Susan Hubbard is the Associate Lab Director for Earth & Environmental Sciences at Berkeley Laboratory. She earned her Ph.D. in Civil and Environmental Engineering at UC Berkeley, where she has been recognized as a Distinguished Alumni. Susan's research focuses on quantifying how terrestrial environments function, with a particular emphasis on using geophysical methods to explore how hydrological, geochemical and biological processes interact to govern larger scale system behavior. Dr. Hubbard has served on many scientific advisory and editorial boards, including the California Council of Science and Technology (CCST).

Roger Bales is Distinguished Professor of Engineering and founding faculty member at UC Merced, where he also directs the Sierra Nevada Research Institute. Previously, he was Professor of Hydrology at the University of Arizona. His water research spans over 30 years, and he is a fellow in multiple professional societies. Currently, his work focuses on building the knowledge base to adapt California's water supplies, ecosystems and economy to the impacts of climate warming.

Lidia Cano Pecharroman is a Graduate Research Fellow at the Earth Institute's AC4 in Columbia University. Her work focuses on the intersection between conflict prevention, natural resources exploitation, and climate change. She is a consultant and researcher for governments and international organizations in these topics. Lidia earned her MIA with a concentration on Energy and Natural Resources Management at the Columbia University School of International and Public Affairs.

Kamyar Guivetchi is Manager of the Division of Statewide Integrated Water Management with the California Department of Water Resources. During his 39 years with DWR, Kamyar has worked on many technical and planning projects with the North-Central Region Office, Division of Planning, Bay-Delta Office, and Division of Environmental Services managing the Suisun Marsh Program. He chairs the State Agency Steering Committee for preparing Water Plan Update 2018, and is the State Co-chair of the California Biodiversity Council's Executive Committee.

Christina McCready is a Principal Engineer with the California Department of Water Resources, where she manages the Integrated Data and Analysis Branch of the Division of Statewide Integrated Water Management. Chris' work with DWR has included civil engineering design, construction inspection, Oroville facilities relicensing, project communications, and, most recently, integrated regional water management. Chris is a licensed professional civil engineer and holds a bachelor of science degree in civil engineering from California Polytechnic State University, San Luis Obispo.

Gary Darling is an Operations Research Specialist at the California Department of Water Resources. He has previously served as the Geographic Information Officer for California, the Chief Information Officer of the California Resources Agency and served as the Internet domain authority for the State of California. He has an A.B. in Economics and a B.S. in Political Economy of Natural Resources from the University of California at Berkeley.

ABOUT CLEE AND UC WATER

The Wheeler Water Institute develops interdisciplinary solutions to ensure clean water for California. Established in 2012 at the Center for Law, Energy & the Environment (CLEE) at Berkeley Law, the Institute conducts projects at the intersection of law, policy and science.

The Center for Law, Energy & the Environment (CLEE) at Berkeley Law educates the next generation of environmental leaders and develops policy solutions to pressing environmental and energy issues. The Center's current initiatives focus on reducing greenhouse gas emissions, advancing the transition to renewable energy, and ensuring clean water for California's future.

The UC Water Security and Sustainability Research Initiative is focused on strategic research to build the knowledge base for better water resources management. UC Water applies innovative science, technology, and implementation strategies to surface water and groundwater management.

ENDNOTES

1 DWR strategic plan is forthcoming, to be released in January 2018.

2 Helly, J. (2017). Open Water Information Architecture System Requirements Document. Forthcoming. Available at: ftp://mae2.sdsc.edu/Download/Project-OWIA-Documentation/

3 Hanak, E. (2011). *Managing California's water: from conflict to reconciliation.* Public Policy Institute of California.

4 The Aspen Institute. (2017). *Internet of Water: Sharing and integrating water data for sustainability*. Report from the Aspen Institute Dialogue Series on Water Data; California Council on Science and Technology. (2014). *Achieving a Sustainable California Water Future through Innovations in Science and Technology.*

5 See, e.g., the 2017 winner of the Imagine H20 2017 California Water Policy Challenge, Center for Law, Energy and the Environment at Berkeley Law (CLEE) and Water and Power Law Group LLC's project to create a modern information system of water rights and water use that will enable deliberate, real-time decision making on the allocation of water in California. Available at: https://www.imagineh2o.org/policy

6 Escriva-Bou, A., McCann, H., Hanak, E., Lund, J. & Gray, B. (2016). *Accounting for California's Water.* Public Policy Institute of California.

7 California Council on Science and Technology (2014), note 4; Bales, R., Conklin, M., Viers, J., Fisher, A., Fogg, G., & Kiparsky, M. 2016. *Foundations for California's Water Security in a Changing Climate*. Institute on Science for Global Policy Working Paper; California Natural Resources Agency (2016). California Water Action Plan 2016 Update.

Bales, R., Conklin, M., Viers, J., Fisher, A., Fogg, G., & Kiparsky, M. 2016. *Foundations for California's Water Security in a Changing Climate.* Institute on Science for Global Policy Working Paper.

California Natural Resources Agency (2016). California Water Action Plan 2016 Update.

8 NSF DataONE (2017). Data Life Cycle. Available at https://www.dataone.org/data-life-cycle

9 Lemos, M. C., & Rood, R. B. (2010). Climate projections and their impact on policy and practice. *Wiley interdisciplinary reviews: climate change*, *1*(5), 670-682.

10 Alexander, I.F. and Maiden, N. eds. (2005). *Scenarios, Stories, Use Cases: Through the Systems Development Life-Cycle.* John Wiley & Sons.

11 United Nations Water. (2013). *Water Security and the Global Water Agenda: A UN Water Analytical Brief.*

12 California Council on Science and Technology (2014), note 4.

13 Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3-9.

14 Kallis, G., Kiparsky, M., Milman, A. & Ray, I. (2006). Glossing over the complexity of water. *Science*, *314*(5804), pp.1387-1388.

15 California Natural Resources Agency, Department of Food and Agriculture, and Environmental Protection Agency. (2016). *California Water Action Plan, 2016 Update.*

16 Escriva-Bou et al. (2016), note 6.

17 California Department of Water Resources. (2016). *Water Budget Best Management Practices: Best Management Practices for the Sustainable Management of Groundwater.* Sacramento, CA: California Department of Water Resources. California Natural Resources Agency. Available at: http://www.water.ca.gov/groundwater/sgm/pdfs/BMP_Water_Budget_Final_2016-12-23.pdf Last Updated: Dec. 2016

California Department of Water Resources. Unpublished. Water Budget Framework Project Tulare Lake Hydrologic Region Pilot Study Phase I Results and Recommendations. Sacramento, CA: California Natural Resources Agency.

18 In addition to the lead agencies named in the bill (DWR, SWRCB, CDFW, and CWQMC), the Act identifies five additional agencies that hold data which are to be aggregated, including the U.S. Bureau of Reclamation (USBR), the U.S. Fish and Wildlife Service (USFWS), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Geological Survey (USGS), and the U.S. Forest Service (USFS), but does not specify precisely what those data are. See Cal. Water Code §12400 et seq.

19 Beyond these California-based data systems, many other national and international efforts have advanced hydrological, earth and environmental cyberinfrastructure constructs and systems. A few examples of broader U.S.-based efforts: Hydroshare, a collaborative environment for sharing hydrologic data and models geared toward academic water researchers; EarthCube, supported by the National Science Foundation (NSF), which aims to connect Earth science researchers; and Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), a cyberinfrastructure system supported by the Department of Energy (DOE) to integrate diverse environmental datasets.

20 California Department of Water Resources, California Data Exchange Center. Available at https://cdec.water.ca.gov/

21 California Department of Water Resources, California Irrigation Management Information System. Available at http://www.cimis.water.ca.gov/

22 California State Water Resources Control Board, Surface Water Ambient Monitoring Program. Available at https://www.waterboards.ca.gov/water_issues/programs/swamp/

23 California State Water Resources Control Board, California Environmental Data Exchange Network. Available at http://www.ceden.org/

24 California State Water Resources Control Board, Electronic Water Rights Information Management System. Available at https://www.waterboards.ca.gov/waterrights/water_issues/programs/ewrims/

25 Cal. Water Code §10720 et seq.; *See also* Green Nylen, N. (2014). California's new groundwater law: An interactive timeline. *Legal Planet*. Available at: http://legal-planet.org/2014/10/08/californias-new-groundwater-law-an-interactive-timeline/

26 Provost, F. and Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), pp.51-59.

27 Delta Stewardship Council. (2015). *Enhancing the Vision for Managing California's Environmental Information.* White paper, Jul 2015.

28 Gleick, P. H. (2000). A look at twenty-first century water resources development. *Water International*, *25*(1), 127-138.

29 Del Moral, L., Pita, M. F., Pedregal, B., Hernández-Mora, N., & Limones, N. (2014). Current paradigms in the management of water: Resulting information needs. *Progress in water geography-Pan-European discourses, methods and practices of spatial water research*, 21-31.

30 Adapted from McAfee, A., Brynjolfsson, E., & Davenport, T. H. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.

31 Salewicz, K.A. & Nakayama, M. (2004). Development of a Web-Based Decision Support System (DSS) for Managing Large International Rivers. *Global Environmental Change, 14*(25), 28–30.

32 Kerkez, B., Glaser, S. D., Bales, R. C., & Meadows, M. W. (2012). Design and performance of a wireless sensor network for catchment-scale snow and soil moisture measurements. *Water Resources Research*, *48*(9).

33 Pita López, M. F., Pedregal Mateos, B., Hernández Mora, N., Limones Rodríguez, N., & Moral Ituarte, L. D. (2014). Key Data and Information Requirements in the Context of Current Debates on Water Management. Technical Report, May 2014.

34 Eggimann, S., Mutzner, L., Wani, O., Schneider, M. Y., Spuhler, D., Moy de Vitry, M., ... & Maurer, M. (2017). The Potential of Knowing More: A Review of Data-Driven Urban Water Management. *Environmental Science & Technology*, *51*(5), 2538-2553; Kiparsky, M., Sedlak, D. L., Thompson Jr, B. H., & Truffer, B. (2013). The innovation deficit in urban water: the need for an integrated perspective on institutions, organizations, and technology. *Environmental engineering science*, *30*(8), 395-408.

35 Rizzoli, A. E., & Young, W. J. (1997). Delivering environmental decision support systems: software tools and techniques. *Environmental Modelling & Software*, *12*(2-3), 237-249.

36 Zhang, K., Zargar, A., Achari, G., Islam, M. S., & Sadiq, R. (2013). Application of decision support systems in water management. *Environmental Reviews*, *22*(3), 189-205.

37 Figure adapted from Loucks, D.P. (2005). Water Resource Systems Modelling: Its Role in Planning and Management. Chapter 2 from Loucks, D.P., van Beek, E., Stedinger, J.R., Dijkman, J., & Villars, M.T. (2005). *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*. Paris: UNESCO

38 Wagner, W. E. (2003). The" bad science" fiction: reclaiming the debate over the role of science in public health and environmental regulation. *Law and Contemporary Problems*, *66*(4), 63-133.

39 Doremus, H., & Tarlock, A. D. (2005). Science, judgment, and controversy in natural resource regulation. *Pub. Land & Resources L. Rev.*, *26*, 1.

40 Sparrow, M. K. (2011). *The regulatory craft: controlling risks, solving problems, and managing compliance*. Brookings Institution Press.

41 Kysar, D. A., & Salzman, J. (2007). Making Sense of Information for environmental protection. *Tex. L. Rev.*, *86*, 1347.

42 Pedregal Mateos, B., Cabello Villarejo, V., Hernández Mora, N., Limones Rodríguez, N., & Moral Ituarte, L. D. (2015). Information and Knowledge for Water Governance in the Networked Society. *Water Alternatives, 8 (2), 1-19*.

43 Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., ... & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences*, *100*(14), 8086-8091.

44 Acreman, M. (2005). Linking science and decision-making: features and experience from environmental river flow setting. *Environmental Modelling & Software*, *20*(2), 99-109.

45 Cash et al. (2003), note 43.

46 McNie, E. C. (2007). Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environmental science & policy*, 10(1), 17-38; Lemos et al. (2010), note 9. .

47 Lemos et al. (2010), note 9.

48 Kulak, D., & Guiney, E. (2012). *Use cases: requirements in context*. Addison-Wesley.; Alexander et al. (2005), note 10.

Alexander, I.F. and Maiden, N. eds. (2005). *Scenarios, Stories, Use Cases: Through the Systems Development Life-Cycle.* John Wiley & Sons.

49 Two major workshops were convened by UC Water, DWR, and CCST. The first workshop was held on February 9, 2017 at UC Berkeley, with approximately 60 attendees. The second workshop was held on May 9, 2017 at UC Davis, with approximately 80 attendees. In addition, a capstone workshop was hosted by LBNL in collaboration with UC Water, DWR, and CCST on November 29, 2019, in Berkeley, with about 100 attendees observing presentation of preliminary results from the Stakeholder Working Group, Technical Working Group, and partner agency efforts and a look forward to next steps.

50 Kiparsky, M. & Bales, R. (2017). Advanced data would improve how California manages water. *The Sacramento Bee,* February 8, 2017. Available at http://www.sacbee.com/opinion/op-ed/soapbox/article131528914.html

51 A technical workshop was held in San Diego in July 2017 with members of both the Stakeholder Working Group and the Technical Working Group for the purpose of translating the use cases into functional requirements, see Helly et al. (2017), note 2.

52 California Natural Resources Agency (2016). California Water Action Plan 2016 Update.

53 Coding of data topics, as shown in Table 3, was first generated by the Technical Working Group and refined by the Stakeholder Working Group. Coding was an iterative process conducted inductively based on the content of the data sources in the use cases. The authors re-coded the data by topic several times throughout the process of analysis to ensure consistency and accuracy. Throughout the coding process, the authors attempted to maintain consistency throughout the data sources in order to give a broad sense of what topics were most frequently cited by data users.

54 USGS Daily Streamflow Conditions. Available at https://waterdata.usgs.gov/ca/nwis/rt

55 Data were first categorized by program and then by agency hosting and providing the resources for that source of data. For instance, the Groundwater Ambient Monitoring and Assessment Program (GAMA) is an assessment program hosted by the State Water Resources Control Board.

56 There were a few exceptions of university data sources that were used more frequently, including the UC Davis Soil Agricultural Groundwater Banking Index, available at https://casoilresource.lawr.ucdavis.edu/sagbi/ and the Oregon State University PRISM Climate Group climate data sets, available at http://www.prism.oregonstate.edu/.

57 The Aspen Institute (2017), note 4.

58 *Ibid*.

59 *Ibid*.

60 Gibson, C. C., Ostrom, E., & Ahn, T. K. (2000). The concept of scale and the human dimensions of global change: a survey. *Ecological economics*, *32*(2), 217-239.

61 Blodgett, D., Read, E., Lucido, J., Slawecki, T., & Young, D. (2016). An analysis of water data systems to inform the open water data initiative. *JAWRA Journal of the American Water Resources Association*, *52*(4), 845-858.

62 Helly et al. (2017), note 2.

63 Paskin, N. (2010). Digital object identifier (DOI®) system. *Encyclopedia of library and information sciences*, *3*, 1586-1592.

64 Karkkainen, B. C. (2000). Information as environmental regulation: TRI and performance benchmarking, precursor to a new paradigm. *Georgetown Law Journal 89*(257), 283–85.

65 For example, SGMA requires groundwater models supporting groundwater sustainability plans to be developed using public domain, open source software code. See Cal Code of Regulations, §352.4(f)(3)

66 For more information, see the Hobbes Project. This bottom up approach to improve and organize data for water modeling efforts in California is a project of the UC Davis Center for Watershed Sciences. For more information see https://hobbes.ucdavis.edu/.

67 McNie (2007), note 46.

68 Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: the case of building internet. *Journal of information technology*, *25*(1), 1-19.

69 Story, D. & Mayton, H. (2017). *Towards a Sustainable Water Future: Synthesis of Key Recommendations for California's Water Data Management.* California Council on Science and Technology.

70 De Stefano, L., Hernández-Mora, N., López Gunn, E., Willaarts, B., & Zorrilla-Miras, P. (2012). Public participation and transparency in water management. In *Water, agriculture and the environment in Spain: can we square the circle*, eds. De Stefano, L. & Llamas, R.M. CRC Press/Balkema, Taylor & Francis Group. 217-225.

71 Sandoval-Almazán, R., Gil-Garcia, J. R., Luna-Reyes, L. F., Luna, D. E., & Rojas-Romero, Y. (2012). Open government 2.0: citizen empowerment through open data, web and mobile apps. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (pp. 30-33). Association for Computing Machinery.

72 Fioretti, M. (2010). *Open data, open society: a research project about openness of public data in EU local administration.* Laboratory of Economics and Management, Scuola Superiore Sant'Anna, Pisa.

73 Story and Mayton (2017), note 69.

74 California Natural Resource Agency. *EcoRestore Adaptive Management Program.* Available at http://resources.ca.gov/ecorestore/ecorestore-adaptive-management-program/.

75 Story and Mayton (2017), note 69.

76 Kiparsky, M., Owen, D., Green Nylen, N., Christian-Smith, J., Cosens, B., Doremus, H., Fisher, A. & Milman, A. (2016). *Designing Effective Groundwater Sustainability Agencies: Criteria for Evaluation of Local Governance Options.* Center for Law, Energy & the Environment, U.C. Berkeley School of Law.

77 Cal. Water Code §12406 (a)

78 Doremus, H. (2008). Data Gaps in Natural Resource Management: Sniffing for Leaks Along the Information Pipeline. *Indiana Law Journal 83,* 407.

79 Agranoff, R. (2006). Inside collaborative networks: Ten lessons for public managers. *Public administration review*, *66*(1), 56-65.

80 Jackson, S. (2006). Water models and water politics: design, deliberation, and virtual accountability. In *Proceedings of the 2006 international conference on digital government research* (95-104). Digital Government Society of North America.

81 Dawes, S. S., & Pardo, T. A. (2002). Building collaborative digital government systems. *Advances in digital government*, 259-273.

82 Pardo, T. A., Gil-Garcia, J. R., & Burke, G. B. (2008, January). Governance structures in cross-boundary information sharing: Lessons from state and local criminal justice initiatives. In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual* (pp. 211-211). IEEE.

83 Bekkers, V. (2009). Flexible information infrastructures in Dutch E-Government collaboration arrangements: Experiences and policy implications. *Government Information Quarterly*, *26*(1), 60-68.

84 Lynch, C. (2008). Big data: How do your data grow? *Nature*, *455*(7209), 28-29.

85 Van Delden, H. (2009). Lessons learnt in the development, implementation and use of Integrated Spatial Decision Support Systems. In *Proceedings of 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia (pp. 13-17).*