Progress Report

Implementing the Open and Transparent Water Data Act

with Initial Draft Strategic Plan and Preliminary Protocols

January 2018

Prepared by the California Department of Water Resources

In Consultation with the Partner Agency Team



STATE OF CALIFORNIA Edmund G. Brown Jr., Governor

THE NATURAL RESOURCES AGENCY John Laird, Secretary for Resources

DEPARTMENT OF WATER RESOURCES Karla Nemeth, Director

> Cindy Messer Chief Deputy Director

Michelle Banonis Assistant Chief Deputy Director

Christy Jones	Kathie Kishaba	Eric Koch	Joel Ledesma	Taryn Ravazzini
Deputy Director				

Erin Mellon Assistant Director, Public Affairs Kasey Shimke Assistant Director, Legislation Spencer Kenner Chief Counsel

DIVISION OF STATEWIDE INTEGRATED WATER MANAGEMENT Kamyar Guivetchi, Manager

This report was prepared under the supervision of:

Christina McCreadyIntegrated Data and Analysis Branch, Chief

by

Matt Correa	Senior Water Resource Engineer
Gary Darling	Operations Research Specialist III
Julie Haas	Senior Water Resource Engineer
Abdul Khan	Supervising Water Resource Engineer
Paul Shipman	Water Resource Engineer
Greg Smith	Program Manager II

Acknowledgements

California Department of Water Resources

Benjamin Brezing Tim Garza David Harris Kristopher Jones Rich Juricich William O'Daly Charlie Olivares Scott Olling Eric Senter Steven Springhorn

California Council on Science and Technology

Susan Hackwood Amber Mace

California Department of Fish and Wildlife

Steve Goldman Tom Lupo Karen Miner

Center for Collaborative Policy

Ariel Ambruster Alexandra Cole-Weiss

Delta Stewardship Council

George Isaac

Delta Conservancy

Shakoora Azimi-Gaylon

Government Operations

Stuart Drown Angelica Quirarte

Governor's Office of Planning and Research

Ken Alex Debbie Franco

Lawrence Berkeley National Laboratory

Debra Agarwal Susan Hubbard Peter Nico Charu Varadharajan

San Diego Super Computer Center

John Helly Michael Norman

S.D. Bechtel, Jr. Foundation

Joya Banerjee

State Water Resources Control Board

Greg Gearheart Rafael Maestu

University of California, Berkeley

Alida Cantor Ronan Kennedy Michael Kiparsky Meredith Lee

University of California, Merced

Roger Bales Leigh Bernacchi Martha Conklin

University of California, Santa Barbara

Jeff Dozier

Water Foundation

Mike Myatt

AB 1755 Technical Working Group

Stephen Abrams, University of California Deb Agarawal, Lawrence Berkeley Laboratory Roger Bales, University of California, Merced Martha Conklin, University of California, Merced Gary Darling, California Department of Water Resources Greg Gearheart, California Water Boards Kamyar Guivetchi, California Department of Water Resources Tony Hale, San Francisco Estuary Institute John Helly, University of California, San Diego Sara Larsen, Western States Water Council Don Sullivan, NASA Ames Matt Correa, California Department of Water Resources David Blodgett, U.S. Geological Survey Isaac George, California Delta Council Christina McCready, California Department of Water Resources Dwane Young, U.S. Environmental Protection Agency

Progress Report Reviewer's Guide -- January 31, 2018

This progress report on the implementation of Assembly Bill 1755, the Open and Transparent Water Data Act (AB 1755) includes background, engagement process, initial draft strategic plan, preliminary protocols, and by citation, supporting information, reports, and references used in its preparation. Comments received on this progress report by March 16, 2018, will inform the April 2018 progress report that will include a final strategic plan, evolving protocols and standards, governance and funding options, and an implementation plan.

How to Comment

Send comments to: ab1755@water.ca.gov Attn: Christina McCready Fax: 916-651-9289

What to Review

DWR Publications staff has not yet fully edited content for grammar, punctuation, style, consistency, accuracy, or other issues relating to readability or quality. The document will be edited for these issues by the Publications staff prior to releasing the April 2018 progress report.

Recommendations for what to focus on during your review are:

- **Relevance and Effectiveness**: Do you see your perspectives, issues, and challenges in the document? Do the documents speak to your constituents/members? Do the documents provide solutions that address your issues and challenges?
- **Completeness of information**: In general, does the text say all it should say? Is all information present that an average reader might need and presented appropriately?
- **Factual accuracy**: Is anything in the text incorrect? Does any information need additional attribution to a specific source?
- **Logical consistency**: Does the narrative build in a logical way and effectively tell the right story?

January 2018 Review Copy

Contents

Introduction	
Intent	
Communication and Engagement Process	
Open and Transparent Water Data Act	
Implementing the Open and Transparent Water Data Act	
Partner Agency Team.	
Initial Draft Strategic Plan	
Preliminary Protocols	
Data Initiatives Informing Implementation	
Use Cases	
Open Water Information Architecture	
Long-Term Governance and Funding	
Data Challenges and Test Bed	
Other Related Efforts	
Project Open Data	
Aspen Dialogue Series	
Open Water Web	
Enhancing the Vision for Managing California's Environmental Information	
California Council on Science and Technology	
Federation	
User Focus and Continuous Improvement	
Limitations	
Funding	
New Data Collection	
Next Steps	
Refinement of Strategic Plan	
Continuing Development of Protocols	
Implementation Plan	
References	

Figures

Figure 1 Typical Data Life Cycle	2
Figure 2 Implementing the Open and Transparent Water Data Act4	ł
Figure 3 Classification of Use Cases by Topic and Decision Objective	j

January 2018 Review Copy

Acronyms and Abbreviations

AB 1755	Assembly Bill 1755, the Open and Transparent Water Data Act
CCST	California Council on Science and Technology
CDFW	California Department of Fish and Wildlife
CNRA	California Natural Resources Agency
CWQMC	California Water Quality Monitoring Council
DWR	California Department of Water Resources
GovOps	Government Operations Agency
OWIA	Open Water Information Architecture
SWRCB	State Water Resources Control Board
WDAF	Water Data Administration Fund

Progress Report for Implementing the Open and Transparent Water Data Act with Initial Draft Strategic Plan and Preliminary Protocols

Introduction

This is a progress report on California Department of Water Resources' (DWR's) implementation of Assembly Bill 1755, the Open and Transparent Water Data Act (AB 1755), which directs DWR, in consultation with other State agencies, to improve the accessibility and usability, of water and environmental data.

Data are an important tool that can help build common understanding, allow for more informed decisionmaking, and improve efficiency and effectiveness. In concert with experience and intuition, data can help us to define, measure, learn from, and adjust outcomes. Data form a strong foundation for collaboration and coordination, which will be essential for achieving sustainable water management in the face of climate change and other pressures on our water resources.

This progress report includes background, communication and engagement process, initial draft strategic plan, preliminary protocols, and interim products to date. The next progress report is scheduled for April 2018, which will include a final strategic plan, evolving protocols and standards, long-term governance and funding options, and an implementation plan.

Intent

AB 1755 directs DWR, in consultation with several other State agencies, to improve the accessibility and usability of water and environmental data. Based on extensive stakeholder engagement, the intended outcome for AB1755 is a federated network of data portals with data used for decision making that are sufficient, accessible, useful, and used. It is understood that AB 1755 implementation is an evolving process that will adapt over time in response to both changing software capabilities and the needs of a broad array of data users and providers.

To capture the spirit of AB 1755, an emphasis is placed on use cases and federation. Use cases are a tool for assessing stakeholder data needs in specific decision contexts, and communicating those needs to technical developers. Employing use cases provides a direct relation between user needs and system requirements. And, because no one website or database could reasonably contain all of California's water-related data, federation offers a viable technical solution. In a federated open data system, each open data platform will be accessible through a federated data catalog, analogous to the inter-library loan system. As a starting point, DWR and its partners aim to federate and populate the California Government Operations Agency and California Natural Resources Agency open data portals to allow users access to available State water and ecological datasets.

This *Progress Report for Implementing the Open and Transparent Water Data Act with Initial Draft Strategic Plan and Preliminary Protocols* reflects a user-centered approach that is intended to build the relationships and investment necessary across State and federal agencies, and among stakeholders to assure that AB 1755 and the larger data vision will be sustainable over time. The aim is to transform data systems to assure that data are sufficient, accessible, and useful in a way that supports water resource decision-makers in achieving sustainable water management.

Open and Transparent Water Data Act

AB 1755 was signed into law by Governor Edmund G. Brown Jr. on September 23, 2016. It requires DWR, in consultation with the California Water Quality Monitoring Council (CWQMC), the State Water Resources Control Board (SWRCB), and the California Department of Fish and Wildlife (CDFW), to create, operate, and maintain a statewide integrated water data platform, and to develop a strategic plan to guide program implementation and protocols for data sharing, documentation, quality control, public access, and promotion of open-source platforms and decision support tools related to water data (California Water Code Section 12400 et seq.).

The integrated water-data platform must be operational with available water and ecological State agency datasets by September 1, 2019. The platform shall make available by August 1, 2020, certain water and ecological data related to California water supply and management held by identified federal agencies.

AB 1755 is focused primarily on the publication stage of the data-life cycle. A "data-life cycle" is a way of thinking about the flow of data from collection to end uses. Industry experts have characterized the number and names of life-cycle stages in different ways, yet agree that data are dynamic with changing risks and attendant needs at each stage. Figure 1 identifies how AB 1755 envisions addressing a representative data life cycle.



Figure 1 Typical Data Life Cycle adapted for AB 1755 from NSF DataOne Project

Data publication is a central focus when discussing open data. While statutory requirements of AB 1755 are most closely related to the publication stage, DWR and the Partner Agency Team have embraced the spirit of AB 1755, seeking to ensure a sustainable data future for California's citizens.

Communication and Engagement Process

In implementing the Open and Transparent Water Data Act, DWR has pursued an aggressive schedule of engagement intended to capture the needs of both data providers and data consumers. In 2017, DWR cohosted a series of interactive workshops focused on the development of use cases on February 9, May 8, and November 29. Workshops engaged data users and data providers from multiple sectors, including academia, research, state and federal agencies, NGOs, and the public. In addition, a technical working group was convened in July 2017 to develop functional and technical requirements for an interoperable system, based on decision-making outcomes identified in the use cases. DWR also convened a partner agency team to support communication among State agencies implementing AB 1755. To increase public involvement in the process and improve the quality of the final products, DWR developed an internet website (water.ca.gov/ab1755), AB 1755 Fact sheet, and a subscriber e-mail list. The outreach process and the accomplishments of our numerous contributing groups are expanded upon in the "Implementing the Open and Transparent Water Data Act" section.

Implementing the Open and Transparent Water Data Act

Partner Agency Team

Coordination and cooperation are vital to the successful implementation of AB 1755. Four State agencies and organizations are explicitly identified in the Open and Transparent Water Data Act —DWR, the SWRCB, CDFW, and the CWQMC. Each of these entities has its own policies and processes, and interagency collaboration is required for successful implementation of AB 1755. Successful implementation of AB 1755 will also require coordination with other agencies, specifically, the Governor's Office of Planning and Research, the California Natural Resources Agency, the Government Operations Agency, and the Delta Stewardship Council.

The AB 1755 Partner Agency Team, representing each of the eight entities noted above, was convened in July 2017 to facilitate communication among State agencies implementing AB 1755 and to support timely fulfillment of statutory requirements of the act. The Partner Agency Team meets regularly to plan, discuss progress, and offer high-level guidance in the multi-agency effort to fulfill requirements of AB 1755, including development of a strategic plan, protocols, and a statewide integrated water-data platform to publish water and ecological datasets.

Figure 2 provides a conceptual characterization of the State's approach to implementation of AB 1755.





With the passage of AB 1755 in 2016 has come a surge of momentum to converge related efforts and make open water data a reality. Researchers and water management professionals who have long been convinced of the importance and value of open data have banded together to make great strides in a relatively short time. DWR, in consultation with the Partner Agency Team, and others, has produced an initial draft strategic plan and preliminary protocols (represented as blue puzzle pieces in Figure 2). Both of these statutorily required documents have benefitted substantially from the insights and recommendations of other data initiatives (represented as green puzzle pieces). These parallel efforts, detailed in the "Data Initiatives" section, will also inform future implementation actions, as will stakeholder input.

Initial Draft Strategic Plan

AB 1755 requires the development of a strategic plan for data management (California Water Code Section 12410[a][1][A]).

Typically, a strategic plan expresses a vision for the future, and related goals, objectives, and actions. Pursuant to AB 1755, DWR has consulted with its partner agencies and others to develop an initial draft strategic plan that includes, (1) a vision for the future of water and ecological data in California, (2) goals that, when attained, will help to realize that future, (3) objectives that support attainment of the goals, (4) strategic actions associated with each objective, and (5) guiding principles to help make decisions and navigate obstacles that will undoubtedly emerge as implementation proceeds. Stakeholders are invited and encouraged to review the initial draft strategic plan presented in this progress report and provide feedback according to the reviewers' guide included at the beginning of this progress report.

Preliminary Protocols

AB 1755 requires the development of protocols for data sharing, documentation, quality control, public access, and promotion of open-source platforms and decision support tools related to water data (California Water Code Section 12406[a]). DWR has consulted with its partner agencies and others to produce preliminary protocols and recommendations for future adaptive management. These preliminary protocols are put forth as minimum procedural requirements to support the intended outcome of improved water resources management through integrated water and ecological data, as expressed in AB 1755. Stakeholders are invited and encouraged to review the preliminary protocols and provide feedback according to the reviewers' guide included at the beginning of this progress report.

Data Initiatives Informing Implementation

The State of California has long recognized the value of data and its importance in sustainable water management.

California Water Plan Update 2013 (www.water.ca.gov/waterplan/cwp/index.cfm) detailed two datarelated objectives and several related actions in its "Roadmap for Action." To begin implementation, DWR entered into partnership with the California Council on Science and Technology (CCST) and UC Water. CCST is a non-partisan, impartial, not-for-profit corporation established by the State Legislature to provide objective advice from California's best scientists and research institutions on policy issues involving science. The UC Water Security and Sustainability Research Initiative is focused on strategic research to build the knowledge base for better water resources management, applying innovative science, technology, and implementation strategy to surface and groundwater management.

Use Cases

During 2017, DWR, CCST, and UC Water co-hosted a series of workshops to engage, explore, and elevate the role of data in water management. The workshops were founded on the conviction that, to properly enable data-driven decisions, consideration should be given to "decision-driven data." In other words, there must be an examination of water management decisions, and the data necessary for decision-making, if users' needs are to be met. In these workshops and subsequent efforts, researchers and water management professionals engaged in the interactive development of 20 draft use cases — short examinations of how decision processes employ data — to inform a decision-driven water-data system, Figure 3 illustrates the variety of decision objectives represented by the 20 draft use cases. The approach of defining objectives and outlining data needs was iteratively refined over the course of the workshops and subsequent engagements with stakeholders. A report on this effort and a compendium of the 20 draft use cases are available at www.law.berkeley.edu/datafordecisions.



Figure 3 Classification of Use Cases by Topic and Decision Objective (Cantor 2018)

Open Water Information Architecture

DWR also partnered with CCST and its affiliate, UC San Diego Supercomputer Center, to outline an Open Water Information Architecture (OWIA), which is characterized as a formal structure for integration of the results of scientific research into the operational decision-making for water resources. The compilation of draft use cases described above helped to inform the development of OWIA functional and technical requirements with a focus on traceability and reproducibility of decisions. A technical working group, comprised of data and library scientists, and other thought leaders, was convened in July 2017 to review and refine the OWIA system requirements document prior to the release of an initial public draft in August 2017. The most recent version of the OWIA system requirements is presented in the appendix of the *Preliminary Protocols for Assembly Bill 1755, the Open and Transparent Water Data Act* report.

Long-Term Governance and Funding

DWR and its partners are cooperating with the S.D. Bechtel, Jr. Foundation and the Water Foundation in conducting an examination of governance and funding models for the long-term sustainability of the federated water-data system. In the near term, the Partner Agency Team provides a temporary governance structure, defining policies, core processes, and roles and responsibilities. In the future, the Partner Agency Team will evolve into a long-term governance body of the statewide integrated water-data platform, informed by this exploration of funding and governance models.

This effort is on an aggressive timeline, with plans to offer insights and recommendations for partner agency consideration by March 2018 to inform State implementation.

Data Challenges and Test Bed

Data challenges are a popular way to engage stakeholder talents and generate public enthusiasm in the application of data to today's issues. Plans are being made for a data challenge to take place in early 2018, focused on the topic of water quality. Other data challenges will follow. DWR and its partners will be keenly interested in the insights they offer.

Also in 2018, DWR aims to participate in a "test bed" environment, which will allow innovators in the private and non-governmental sectors to interact with a system prototype, including proposed protocols and standards, and share their real-time feedback.

These data challenges and test bed activities will help to solidify a feedback loop among data providers, data consumers, and platform operators. This exchange of information about what works and needed changes will support continuous improvement, which is vital to successful implementation of AB 1755, in both the short- and long-term.

Other Related Efforts

In addition to the initiatives described above, the partner agencies' implementation of AB 1755 is being informed by other key initiatives and work products.

Project Open Data

In 2013, the White House launched Project Open Data, based on the foundational belief that, "Data is a valuable national resource and a strategic asset to the U.S. Government, its partners, and the public. Managing this data as an asset and making it available, discoverable, and usable — in a word, open — not only strengthens our democracy and promotes efficiency and effectiveness in government, but also has the potential to create economic opportunity and improve citizens' quality of life."

Aspen Dialogue Series

The Aspen Institute Energy and Environment Program, in partnership with the Nicholas Institute for Environmental Policy Solutions at Duke University and Redstone Strategy Group, convened the Aspen Institute Dialogue Series on Water Data, hosting roundtable discussions around the United States in 2016 and 2017. Discussion outcomes have been captured in a white paper titled, *Internet of Water: Sharing and Integrating Water Data for Sustainability*.

Open Water Web

In a paper titled, *An Analysis of Water Data Systems to Inform the Open Water Data Initiative*, Blodgett et al., articulate four characteristics necessary for an open water web: (1) a catalog for the system, (2) data as a service, (3) enrichment of water data, and (4) community of water data and tools.

Enhancing the Vision for Managing California's Environmental Information

This white paper offers a synthesis of ideas explored during a June 2014 Data Summit based on initiatives, current programs, and the experiences of individuals from State and federal agencies, academia, private and non-governmental entities. Primary goals for the document include sustainable support and advancement of California's existing data systems, ensuring alignment with national technology trends, and laying the foundation for more consistent and robust access to data and metadata across organizational boundaries.

California Council on Science and Technology

In its 2014 document, *Achieving a Sustainable California Water Future through Innovations in Science and Technology*, the CCST, with participation from State, federal, and local agencies, academia, federal research laboratories, and the private sector, explored opportunities for innovation to help address California's water resource challenges, and offered both near-term and long-term recommendations.

DWR and its partners are profoundly grateful for the lessons and insights afforded by all of these data initiatives. Each has helped to place California on a path to making water resource information accessible, discoverable, and usable, thus fostering entrepreneurship, innovation, and scientific discovery.

Federation

DWR considers AB 1755 as a unique opportunity to contemplate and implement a more robust water data future in which public data are not only sufficient and available, but also useful and used in decision-making and innovation. Many governmental agencies already publish data, usually through program-specific webpages and web applications. This fragmented organization can make finding the data unduly difficult, often standing in the way of data being useful and used. To capture the innovative spirit of AB 1755, DWR and its partners will emphasize federation as a strategy.

No one website or data base could reasonably contain all of California's water-related data without a substantial increase in data management overhead by the State. But, there is a viable technical solution that links open data platforms under a federated water-data system. To understand federation, it may be helpful to think of open data platforms as libraries, which can be "federated" under an inter-library loan system. Each library has a different set of books and some have special collections. A customer can use the inter-library loan card catalog to find and request a book from any library in the federation. Similarly, in a federated open data system, each open data platform does not need to warehouse every dataset; instead, each open data platform will be accessible through a federated card catalog.

A federated approach will enhance consumer access through voluntary system expansion. Anyone can create an open data platform; and, provided that such a platform conforms to open data principles and defined AB 1755 protocols, it can be added to the federated system, further expanding the card catalog of available datasets.

California's Government Operations Agency (GovOps) hosts a statewide open data portal (www.data.ca.gov) to improve collaboration, expand transparency, and support innovation. In this context, "open data" is defined as public data collected by the State through its routine business activities, and published in a format that is easy to search, download, and combine with other datasets from other sources. The GovOps portal was designed to host open data from more than one agency. Several State agencies host their own open data portals.

The California Natural Resources Agency (CNRA), home of five constituent departments including AB 1755 implementation partners DWR and CDFW, will launch its own open data platform in the near future. This platform, like the GovOps portal, is aimed at increasing the transparency and sharing of data, specifically those collected and managed by the departments of the CNRA.

As a starting point, DWR, in consultation with its partners, aims to federate the GovOps and CNRA

portals to allow users access to the range of available existing water and ecological datasets held by State agencies. In time, additional portals will join the federation, bringing users greater access to available water and ecological data.

User Focus and Continuous Improvement

Success will be measured by the system's ability to evolve and improve over time in response to users' needs. Performance metrics and user feedback — both positive and suggestions for improvement — will be monitored and evaluated, and will guide evolution of the system in the long term. Decision-making about the system itself will be addressed by a governance structure to be determined.

Limitations

Funding

Progress in the implementation of AB 1755 will depend on the level of funding available. At present, the Legislature has authorized DWR, the SWRCB, and CDFW appropriation authority from the Water Data Administration Fund (WDAF) established by the Act, but has directed limited, one-time funding, in fiscal year 2017-18 to the WDAF.

New Data Collection

AB 1755 does not explicitly call upon the State to identify or address data gaps. But, over time, the collection and management of use cases will undoubtedly reveal gaps in data that must be addressed.

Next Steps

Refinement of Strategic Plan

Stakeholders are invited and encouraged to review the initial draft of *Assembly Bill 1755: Open and Transparent Water Data Act Strategic Plan* and provide feedback according to the reviewers' guide included at the beginning of this progress report. A final strategic plan is scheduled for release in April 2018.

Continuing Development of Protocols

Stakeholders are invited and encouraged to review the *Preliminary Protocols for Assembly Bill 1755, the Open and Transparent Water Data Act* and provide feedback according to the reviewers' guide included at the beginning of this progress report. An updated set of protocols is scheduled for release in April 2018.

Implementation Plan

While the strategic plan sets our sights on a long-term future for management and use of California's water and ecological data, it does not detail all the services and activities required for its implementation. An actionable implementation plan, more time-specific and tactical in nature, is necessary. Following release of a final strategic plan in April 2018, DWR, in consultation with its partners, will publish an implementation plan that expands upon the strategic actions with delivery timelines and metrics.

References

- Bales, Jerad. (2016). Featured Collection Introduction: Open Water Data Initiative. Journal of the American Water Resources Association (JAWRA) 52(4): 811–815. DOI: 10.1111/1752-1688.12439 [Available at http://onlinelibrary.wiley.com/doi/10.1111/1752-1688.12439/full]
- Blodgett, David, Emily Read, Jessica Lucido, Tad Slawecki, and Dwane Young. (2016). An Analysis of Water Data Systems to Inform the Open Water Data Initiative. Journal of the American Water Resources Association (JAWRA) 52(4):845–858. DOI: 10.1111/1752-1688.12417 [Available at http://onlinelibrary.wiley.com/doi/10.1111/1752-1688.12417/full]
- California Council on Science and Technology. (2014). Achieving a Sustainable California Water Future through Innovations in Science and Technology.
- Cantor A., Kiparsky M., Kennedy R., Hubbard S., Bales R., Cano Pecharroman L., Guivetchi K., McCready C., Darling G. 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. 54 pp. Available at: https://doi.org/10.15779/J28H01 or law.berkeley.edu/datafordecisions
- Delta Stewardship Council. (2015). Enhancing The Vision For Managing California's Environmental Information. Report from the Environmental Data Summit Organizing Committee under the leadership of the Delta Stewardship Council's Delta Science Program. [Available at http://deltacouncil.ca.gov/docs/enhancing-vision-managing-california-s-environmentalinformation-final]
- Dodd, B. (2016). Assembly Bill No. 1755. The Open and Transparent Water Data Act. Available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB1755 Cal. Water Code §12400
- Helly, J. (2017). Open Water Information Architecture System Requirements Document. Forthcoming. Available at: ftp://mae2.sdsc.edu/Download/Project-OWIA-Documentation/
- NSF DataONE (2017). Data Life Cycle. Available at https://www.dataone.org/data-life-cycle Accessed: Jan. 2018.
- Project Open Data (2017). Open Data Policy Managing Information as an Asset. Available at https://project-open-data.cio.gov/ Accessed: Jan. 2018
- The Aspen Institute. (2017). Internet of Water: Sharing and Integrating Water Data for Sustainability. Report from the Aspen Institute Dialogue Series on Water Data.

Attachment A

Initial Draft of the Strategic Plan for Assembly Bill 1755, the Open and Transparent Water Data Act

Who is this Strategic Plan for?

This strategic plan is intended to address the stakeholders interest in the Assembly Bill 1755, the Open and Transparent Water Data Act (AB 1755) of 2016 open water-data platform and water resources data as well as a plan for the State agencies charged with implementing the statutory requirements of AB 1755. Stakeholders are invited and encouraged to review the initial draft of the *Strategic Plan for Assembly Bill 1755, the Open and Transparent Water Data Act* and provide feedback according to the reviewers' guide. A final strategic plan is scheduled for release in Spring 2018.

Partner Agency Team

This strategic plan was developed by the California Department of Water Resources in consultation with the AB 1755 Partner Agency Team. The AB 1755 Partner Agency Team is comprised of four State agencies and organizations explicitly identified in the Open and Transparent Water Data Act — the California Department of Water Resources, the State Water Resources Control Board, the California Department of Fish and Wildlife, and the California Water Quality Monitoring Council. Additional agencies partnering to ensure successful implementation of AB 1755 include, the Governor's Office of Planning and Research, the California Natural Resources Agency, the Government Operations Agency, and the Delta Stewardship Council.

January 2018 Review Copy

Contents

C4 4 ' DI	e v 11	D'II 1955 (I O	1		4 1
Strategic Plan	tor accembly	KIII I / SS THE LI	nen and Franch	arent Water Data A	et l
Su augu i ian	IUI ASSUMULY	DIII 1755, the O	pun anu rransp	arone mater Data A	LL
a	•/	,			

Information as an Asset	1
Open and Transparent Water Data Act	1
Vision	3
Goals, Objectives, and Strategic Actions	3
Guiding Principles	3
Federation	3
Implementation Plan and Strategic Plan Updates	4
Responsibility for Implementing Strategic Actions	4
Goal 1: Data are Sufficient	6
Goal 2: Data are Accessible	8
Goal 3: Data are Useful	10
Goal 4: Data are Used	12
Glossary	15
References	16

Figure

Figure 1 Typical Data-Life Cycle for AB 1755 adapted from NSF DataOne Project2

Abbreviations

AB 1755	Assembly Bill 1755, the Open and Transparent Water Data Act
CNRA	California Natural Resources Agency
DWR	California Department of Water Resources
GovOps	Government Operations Agency
OWIA	Open Water Information Architecture

January 2018 Review Copy

Strategic Plan for Assembly Bill 1755, the Open and Transparent Water Data Act

Information as an Asset

Over the past decade or so, the perceived value of data has changed dramatically. Today, thanks to the prolific use of smart phones and tablets, people are quite accustomed to having the world at their fingertips. Available on demand are up-to-the-minute weather projections, driving routes that avoid traffic, and stock performance reports — all with little or no delay for the user. The ability of the average citizen to access and visualize data is at an all-time high.

This level of access is not quite the case when it comes to water-related data sets.

For California, water and ecological data is collected, managed, published, and analyzed by any number of sources. Water-resource-related datasets reside in numerous State, local, and federal agencies, academia, and non-governmental organizations, making it challenging to identify, access, and use data to inform water planning, management, and decision-making.

It stands to reason that decisions founded on timely data and science will be reliable and transparent. It also stands to reason that data collected using public funds are the property of the public and should be made publicly accessible.

Making data of documented quality both easily accessible and discoverable is a necessary first step to improving water management in California.

Open and Transparent Water Data Act

Assembly Bill 1755, the Open and Transparent Water Data Act (AB 1755), signed into law by Governor Edmund G. Brown Jr. on September 23, 2016, requires the California Department of Water Resources (DWR), in consultation with the California Water Quality Monitoring Council, the State Water Resources Control Board , and the California Department of Fish and Wildlife, to create, operate, and maintain a statewide integrated water data platform, and to develop a strategic plan to guide program implementation and protocols for data sharing, documentation, quality control, public access, and promotion of open-source platforms and decision support tools related to water data. (California Water Code Section 12400 et seq.)

The integrated water-data platform must be operational with available water and ecological State agency datasets by September 1, 2019. The platform shall make available by August 1, 2020, certain water and ecological data related to California water supply and management held by identified federal agencies.

AB 1755 is focused primarily on the publication stage of the data-life cycle. A *data-life cycle* is a way of thinking about the flow of data from collection to end uses. Industry experts have characterized the number and names of life cycle stages in different ways, yet agree that data are dynamic with changing risks and attendant needs at each stage. While statutory requirements of AB 1755 are most closely related

to the publication stage, DWR and the Partner Agency Team have embraced the spirit of AB 1755, seeking to ensure a sustainable data future for California's citizens. Consequently, we contemplate addressing the full data-life cycle in implementing the act. Figure 1 identifies how AB 1755 envisions addressing a representative data-life cycle.





Notes:

AB 1755 = Assembly Bill 1755, the Open and Transparent Water Data Act

Use Case Development: Description of who needs the data, and the form the data is needed, to make specific decisions. Plan: Description of the data that will be compiled, and how the data will be managed and made accessible throughout its lifetime. Collect: Observations are made either by hand or with sensors or other instruments and the data are placed into digital form. Assure: The quality of the data are assured through checks and inspections.

Describe: Data are accurately and thoroughly described using the appropriate metadata standards

Publish: Data are made publicly accessible for use by others, along with the relevant information about the data (metadata).

Archive: Data are submitted to an appropriate long-term archive (i.e. data center).

Information: Data are analyzed.

Data Driven Decisions: The practice of making decisions based on analysis of data rather than experience or intuition.

Vision

Useful data for sound, sustainable water resource management.

Goals, Objectives, and Strategic Actions

In support of the vision, four goals have been articulated, as follows:

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

These goals are, in turn, supported by a number of objectives and strategic actions, which are detailed on the following pages.

Guiding Principles

In the implementation of the AB 1755, the State of California embraces the following guiding principles:

- 1. Making data and information accessible, discoverable, and useful will foster new insights, innovation, and entrepreneurship, as well as enhance transparency and trust in decision-making. Data are a valuable resource, and should be made accessible to the widest range of users for the widest range of purposes to the extent permitted by law. At the same time, data are subject to privacy, security, and other valid restrictions.
- 2. A data-sharing platform should always support the accessibility, discoverability, and usefulness of data. Continuous improvement, based on feedback from end users and data producers, is the cornerstone of a successful platform.
- 3. Data producers and data users have distinct roles and responsibilities in open data. Data producers are responsible for sharing data of known quality, and documenting essential metadata; end users are responsible for determining data fitness for use, and documenting their data products.

In addition, we embrace the guiding principles set forth by Project Open Data and by the Aspen Institute's *Internet of Water: Sharing and Integrating Water Data for Sustainability.*

And finally, we acknowledge that State agencies, departments and boards involved in this effort may need to evaluate and refine some of their internal data management strategies to succeed in achieving the goals and vision of this plan.

Federation

The Partner Agency Team considers AB 1755 as a unique opportunity to implement a more robust waterdata future in which public data are not only sufficient and available, but also useful and used in decisionmaking and innovation. Many governmental and non-governmental agencies already publish water data, usually through program-specific web pages and web applications. This fragmented organization can make finding the data unduly difficult, often standing in the way of data being useful and used. To capture the innovative spirit of AB 1755, this plan will emphasize federation as a strategy.

No one website or data base could reasonably contain all of California's water-related data without a substantial increase in data management overhead by the State. But, there is a viable technical solution that links open data platforms under a federated water-data system. To understand federation, it may be helpful to think of open data platforms as libraries, which can be "federated" under an inter-library loan system. Each library has a different set of books and some have special collections. A customer can use the inter-library loan card catalog to find and request a book from any library in the federation. Similarly, in a federated open data system, each open data platform does not need to warehouse every dataset; instead, each open data platform will be accessible through a federated card catalog.

A federated approach will enhance consumer access through voluntary system expansion. Anyone can create an open data platform; and, provided that such a platform conforms to open data principles and defined AB 1755 protocols, it can be added to the federated system, further expanding the card catalog of available datasets.

California's Government Operations Agency (GovOps) hosts a statewide open data portal (www.data.ca.gov) to improve collaboration, expand transparency, and support innovation. In this context, "open data" is defined as public data collected by the State through its routine business activities and published in a format that is easy to search, download, and combine with other datasets from other sources. The GovOps portal was designed to host open data from more than one agency. Several State agencies host their own open data portals.

The California Natural Resources Agency (CNRA), home of five constituent departments including AB 1755 implementation partners DWR and the California Department of Fish and Wildlife, will launch its own open data platform in the near future. This platform, like the GovOps portal, is aimed at increasing the transparency and sharing of data, specifically those collected and managed by the departments of the CNRA.

As a starting point, DWR, in consultation with its partners, aims to federate the GovOps and CNRA portals to allow users access to the range of available existing water and ecological datasets held by State agencies. In time, additional portals will join the federation, bringing users greater access to available water and ecological data.

Implementation Plan and Strategic Plan Updates

While the strategic plan sets our sights on a long-term future for management and use of California's water and ecological data, it does not detail all the services and activities required for its implementation. An actionable implementation plan, more time-specific and tactical in nature, is necessary. Following release of a final strategic plan in Spring 2018, DWR, in consultation with its partners, will publish an implementation plan that expands upon the strategic actions with delivery timelines and metrics. The Partner Agency Team intends to update the AB 1755 Strategic Plan every 5 years in synchronization with the California Water Plan.

Responsibility for Implementing Strategic Actions

The Partner Agency Team, or future AB 1755 governance structure, is responsible for pursuing

implementation of strategic actions identified in this plan. In many cases, these strategic actions require participation from external partners, data providers and end users. These participants include State, federal, and local government; non-governmental-organizations, academia, tribes, for-profit companies, and citizens. It will take a village to develop and sustain a successful open water-data platform.

Goal 1: Data are Sufficient

To improve water management decisions, data must be available at sufficient spatial and temporal resolution while meeting minimum standards of quality. Use cases for water resources management decision-making will be used to determine sufficiency.

Objective 1.1: Utilize Use Cases to Identify the Data Needed for Water Management Decisions

Utilize the concept of use cases to build and maintain a robust inventory of datasets identified by data managers and data consumers. Synthesizing data needs from use cases will highlight important datasets, missing data (data gaps), and more broadly, general system characteristics necessary for an effective open water-data platform. Use cases will continue to be developed as implementation and needs progress.

Action Number	Strategic Action	Description	Time Frame
1	Use Cases	Develop an initial, extensible set of use cases to define and articulate stakeholders' data and information needs in terms that are translatable into platform development requirements. Use cases describe who needs the data, and the form the data is needed, to make specific decisions. For more information, see Data for Water Decision Making: Stakeholder Working Group Synthesis Report.	2018, First Quarter
2	Synthesize Data Needs	Identify and publish a list of water resources-related datasets from use cases.	2018, Third Quarter, Ongoing
3	Write Use Case Protocol	Develop protocol for writing use cases based on lessons learned from use case projects described in actions 1.1.1 and 4.1.1.	2019
4	Use Case Solicitation	Continually collect and publish use cases and data needs from data consumers and water resource managers.	2019 - Ongoing

Objective 1.2: Improve Coordination of Existing Data Collection Efforts

On a periodic basis, publish a catalog of existing data collection efforts at all levels of government, along with data needed for use cases. Prioritize existing datasets to publish. Support cost-efficient data collection efforts.

Action Number	Strategic Action	Description	Time Frame
1	Inventory Existing Datasets	Inventory and publish list of water resources related datasets. Include government agency that collected the data, along with the reason the information was collected.	2018, Third Quarter
2	Identify Data Gaps	Identify data needed, but not collected, for use cases identified in Objective 1.1.	2019, First Quarter
3	Prioritize Datasets	Prioritize existing datasets to publish on an open data platform.	2018, Fourth Quarter
4	New Programs	Support agency requests to develop new, or expand existing, data collection efforts to address data gaps.	2019, First Quarter
5	Improve Coordination	Improve data coordination efforts among all organizations.	2018, Third Quarter
6	New Data Collection Protocol	Develop protocols for creating or expanding data collection programs.	2019

Goal 2: Data are Accessible

Develop a federated open data platform for data stewards to publish data. Everyone would have access to the data, and the desired data would be easily discoverable. *Federation* refers to a system for data producers to maintain data, while allowing others to have access to the data through a shared catalog and agreed upon standards (Aspen Institute).

Objective 2.1: Develop Protocols

Protocols provide guidelines detailing how to complete certain tasks. Protocols also establish many of the minimum requirements for data to be available in the open data platform. Protocols are essential for ensuring interoperability across open data platforms and datasets. Developing and maintaining a robust set of protocols will be critical for the success of AB 1755.

Action Number	Strategic Action	Description	Time Frame
1	Initial minimum protocol requirements	Develop and publish minimum protocols for data publication. This includes required and optional metadata requirements, data dictionary if applicable, and optional free text description of the data collection, data management, and data qualification methods used by the data steward.	2018, First Quarter
2	Open Water Information Architecture (OWIA)	Leverage OWIA to develop processes data stewards <u>may</u> use to transform and standardize datasets for publication.	2018, Fourth Quarter
3	Develop governance structure and process for evolution of protocols	Develop governance structure and process to guide evolution of protocols.	2018, Second Quarter
4	Enhance protocols	Implement recommendation of the governance group, user feedback, and third-party efforts for the open data platform and the OWIA.	2019, Third Quarter, Ongoing
Objective 2.2: Develop Federated Open Water-Data Platform

The open water-data platform will be the gateway for access to all AB1755-related datasets. A federated platform with a single access point will significantly enhance user access to the correct data and make more data discoverable. This is not a once-and-done process. It will require constant iteration and user feedback to meet the needs of data consumers.

Action Number	Strategic Action	Description	Time Frame
1	Use Existing Open Data Platforms	Expand use of data.ca.gov, data.cnra.ca.gov, and other existing open data platforms to publish datasets including datasets identified in AB 1755.	2018, First Quarter, Ongoing
2	Develop Platform Functional Requirements	Develop functional and technical requirements for an open water- data platform for the State of California informed by Objective 4.1 use case pilot projects, and data consumer and data provider feedback.	2018, Fourth Quarter
3	Evaluate Existing Platforms	Determine if, and how, existing open data platforms and infrastructure can meet aspirational functional and technical requirements.	2019, First Quarter
4	Implement Platform Enhancements	Implement initial aspirational enhancements based on the functional requirements.	2019, Third Quarter
5	Support Platform	Continually enhance the open data platform based on input from data consumers, data providers, and AB 1755 governance.	2020
6	Publish Federal Datasets	Federate federal datasets identified in legislation on open data platform.	2020, Third Quarter
7	Federate with other Platforms	Federate with other open data platforms.	2020

Goal 3: Data are Useful

Objective 3.1: Improve Interoperability of Datasets

Decision making often involves comparing and integrating multiple datasets. To make datasets useful, all datasets should include documentation based on protocols and community-accepted standards. *Interoperability* refers to the formatting that allows computer systems to exchange information using specified data formats and communications protocols that enable data to be readily downloaded, uploaded, and exchanged. Data also need to be interoperable to establish some common information exchange reference — typically accomplished via data standards.

Action Number	Strategic Action	Description	Time Frame
1	Enhance Metadata	Enhance metadata documentation to improve discoverability based on use cases and user feedback.	2019, First Quarter, Ongoing
2	Controlled Vocabulary	Adopt subject-matter-controlled vocabulary to enhance dataset integration based on water community recommendations.	2020 - Ongoing
3	Geo-Referencing	Adopt best practices for geo-referencing data to enhance interoperability based on water community recommendations.	2020 - Ongoing
4	Organization	Develop guidance for organization of a dataset and its resources.	2019, First Quarter
5	Keywords	Maintain a master list of keywords and categories informed by user feedback and governance group.	2019, Second Quarter
6	Data Format	Collect information from consumers on improvements that could be made to the data, or data format, and implement recommendations as feasible.	2019

Objective 3.2: Transform Data into Information

Transforming data into useful information is the responsibility of the entire California water community. The open data platform will have limited analytical capability, mostly focused on helping the data consumer find the right data. Combining datasets, developing analytical tools, seeking innovative results, and sharing those results will be done outside of the open data platform, and by everyone interested in improving water management in California.

Action Number	Strategic Action	Description Time Fi	
1	Visualization	Develop simple visualization tools for the open data platform.	2019, Third Quarter
2	Reports	Develop annual report on dataset use from open data platforms for governance.	2019, Third Quarter
3	Download Tools	Develop tools for filtering and querying datasets, and support downloading subsets of data on open data platform.	2019, Third Quarter
4	Collaboration and Innovation	Collaborate with third parties to develop custom web applications and decision support tools. Provide a location to feature third-party tools on a rotational basis.	2019, First Quarter, Ongoing
6	Water Data for California Map Based Web Application	Develop map-based web application enabling users to spatially discover datasets across federated AB 1755 open data platforms.	2020
5	Analytics	Foster and support development of an analytical sandbox for the water community.	2021

Objective 3.3: Archive Datasets

Become the data archive for water-resource data in California.

Action Number	Strategic Action	Description	Time Frame
1	Archive Data	Archive all versions of datasets on the open data platform.	2020

Goal 4: Data are Used

The AB 1755 efforts (use case, open data platform, and data stewardship) should be used to inspire improved, sustainable water resource decisions in California, and foster a water community engaged in decision making at all levels of government.

Objective 4.1: Use Case Pilot Project

Pilot projects with programs that developed use cases should be conducted to illuminate the necessary next steps to improve water management decisions, after making the data accessible. These steps would include learning if the AB 1755 efforts were useful, learning how to improve the open data platform and interoperability of the data, documenting the responsibilities of data consumers, and republishing improved datasets employed in the use case.

Initial use cases prepared by UC Water should be used to evaluate the effectiveness of the open data platform.

Action Number	Strategic Action	Description	Time Frame
1	Pilot Projects and Interoperability Test Beds	Select a few use cases, protocols, and data standards, and use data published on the open data platform to address the issues raised.	2018, Third Quarter
2	Lessons Learned	Publish lesson learned from pilot projects to improve open data platform and protocols.	2019, First Quarter

Objective 4.2: Engage the Water Community

It is important to develop a vibrant water community to support AB 1755 efforts. To make the data meaningful, the efforts must continually evolve to reflect the needs of the water community. Community engagement is necessary for developing new use cases, recommending improvements to the open data platform, improving the usefulness of the data, using all of this to develop new analytical products, and sharing work with their colleagues.

Action Number	Strategic Action	Description Time	
1	Communication and Engagement	Develop and implement a communications and engagement strategy for the water community.	2018, Second Quarter
2	Data Literacy and Consumer Training	Develop training resources for data consumers to improve data literacy.	2019, Third Quarter
3	Data Stewardship Training	Develop training resources for Partner Agencies wishing to elevate their data to the open data platform.	2019, Third Quarter
4	User Group	Foster and support data steward user group.	2019, First Quarter
4	User Group	Foster and support data steward user group.	2019, First Quarter

Objective 4.3: Sustain the Open Water-Data Platform

The open data platform must be able to respond to changing needs and technology trends. To be sustainable, the system must continually adapt and respond to recommendations for improvement.

Action Number	Strategic Action	Description	Time Frame
1	Sustainable Governance and Funding	Work with the water community to develop a sustainable governance and funding structure for AB 1755 efforts.	2018, Second Quarter
2	Metrics	Platform and dataset usage tracking/statistics, and "success stories/case studies" to measure and highlight how the open data platform is being used for water management decisions.	2019, First Quarter
3	Improvements	Recommend open data platform, dataset documentation, analytical tool, and reporting improvements.	2019, Second Quarter
4	User Feedback	Solicit feedback from water managers, data consumers, and data stewards on platform and data performance.	2019, First Quarter

Objective 4.4: Use the Data

One measure of success will be when government agencies and the public begin to use data on the platform to create decision-support tools and inform decisions.

Action Number	Strategic Action	Description	Time Frame
1	Analytical Tools	Encourage citizens and government programs to develop decision-support tools and custom web application using datasets from the open data platform.	2019, Third Quarter
2	Government Publications	Encourage government publications to source and cite data published in the open data platform.	2020, Third Quarter
3	Civic Engagement Projects	Support civic engagement activities and events, like data challenges, application or visualization competitions and hackathons where citizens can meet and engage in collaborative computer programming towards developing innovative solutions for management of California's water resources. This includes challenges for third parties to make use of data on the open data platform to meet the needs of identified use cases with data gaps.	2020, Third Quarter

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: 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.

Federation: A group of data providers and users using jointly agreed-upon standards of operation in a collective fashion to ensure the interoperability of the resources they collectively hold and employ. The term may be used, for example, when describing the interoperation of distinct cyber infrastructure networks with different internal structures. The term may also be used when human groups agree to collectively manage cyberinfrastructure development and operation using commonly held, and managed, requirements, standards and conventions, and operating procedures to ensure the interoperability of distinct cyberinfrastructure resources.

Federated data system: 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.

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

Interoperability: The ability of diverse computer systems or software to exchange and make use of common input data.

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

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.

Protocols: Methods of implementing a set of objectives and requirements in a systematic way. In computing, protocols mean both specific implementations of methods such as HTTP and FTP and, more generally as described by the Internet Engineering Task Force, protocols are sequences of processing steps that are also referred to as procedures.

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.

References

- Dodd, B. (2016). Assembly Bill No. 1755. The Open and Transparent Water Data Act. Available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB1755 Cal. Water Code §12400
- Cantor A., Kiparsky M., Kennedy R., Hubbard S., Bales R., Cano Pecharroman L., Guivetchi K., McCready C., Darling G. 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. 54 pp. Available at: https://doi.org/10.15779/J28H01 or law.berkeley.edu/datafordecisions
- Helly, J. (2017). Open Water Information Architecture System Requirements Document. Forthcoming. Available at: ftp://mae2.sdsc.edu/Download/Project-OWIA-Documentation/
- The Aspen Institute. (2017). Internet of Water: Sharing and integrating water data for sustainability. Report from the Aspen Institute Dialogue Series on Water Data.
- Blodgett, David, Emily Read, Jessica Lucido, Tad Slawecki, and Dwane Young. (2016). An Analysis of Water Data Systems to Inform the Open Water Data Initiative. Journal of the American Water Resources Association (JAWRA) 52(4):845–858. DOI: 10.1111/1752-1688.12417 [Available at http://onlinelibrary.wiley.com/doi/10.1111/1752-1688.12417/full]
- Bales, Jerad. (2016). Featured Collection Introduction: Open Water Data Initiative. Journal of the American Water Resources Association (JAWRA) 52(4): 811–815. DOI: 10.1111/1752-1688.12439 [Available at http://onlinelibrary.wiley.com/doi/10.1111/1752-1688.12439/full]
- California Council on Science and Technology. (2014). Achieving a Sustainable California Water Future through Innovations in Science and Technology.
- Delta Stewardship Council. (2015). Enhancing The Vision For Managing California's Environmental Information. Report from the Environmental Data Summit Organizing Committee under the leadership of the Delta Stewardship Council's Delta Science Program. [Available at http://deltacouncil.ca.gov/docs/enhancing-vision-managing-california-s-environmentalinformation-final]
- NSF DataONE (2017). Data Life Cycle. Available at <u>https://www.dataone.org/data-life-cycle</u> Accessed: Jan. 2018.

Project Open Data (2017). Open Data Policy - Managing Information as an Asset. Available at https://project-open-data.cio.gov/ Accessed: Jan. 2018

Attachment B

Preliminary Protocols for Assembly Bill 1755, the Open and Transparent Water Data Act

Contents

Preliminary Protocols for Assembly Bill 1755, the	Open and Transparent Water Data Act1
---	--------------------------------------

Introduction	1
Foundational Concepts	2
Protocols	2
Data Standards	2
Interoperability	2
Initial Minimum Protocols	3
Identify a Data Steward	3
Publish and Document on an Open Data Platform	4
Metadata Requirements	4
Data Dictionary Requirements	5
Guideline for Optional Descriptive Text	5
Access Data	6
Continuing Protocol Development - Use Cases and the Open Water Information Architecture	6
Continuing Development of Protocols – Long Term	8
Interoperability Testbeds	9
References	9
Glossary	10
Appendix A Open Water Information Architecture System Requirements Document	
Appendix B Preliminary Open Data Metadata and Data Dictionary Requirments	

Tables

Table 1 Three Initial Protocols Developed to Support Early Implementation of AB 17553
Table 2 A Sample Technical Workflow for Publishing a Dataset on data.ca.gov or the California Natural Resources Agency Open Data Platform
Table 3 A Sample Workflow for Accessing Data on an Open Data Portal
Table 4 Summary of the Functional and Technical Requirement Categories Specified in the OWIAto Protocol Categories Identified by AB 1755
Table 5 Mapping of 41 Functional and Technical Requirement identified in OWIA to Protocol Categories Identified by AB 1755

Figure

Figure 1	Proposed Proc	ess for Protocol	Development	8
----------	----------------------	------------------	-------------	---

Acronyms and Abbreviations

AB 1755	Assembly Bill 1755, The Open and Transparent Water Data Act
DWR	California Department of Water Resources
CCST	California Council on Science and Technology
IETF	Internet Engineering Task Force
OWIA	Open Water Information Architecture
Progress Report	Progress Report for Implementing the Open and Transparent Water Data Act with Initial Draft Strategic Plan and Preliminary Protocols

Preliminary Protocols for Assembly Bill 1755, the Open and Transparent Water Data Act

Introduction

In the wake of the most recent drought, the Legislature passed Assembly Bill 1755, the Open and Transparent Water Data Act (AB 1755) with the goal of improving water resources management through development of an open source platform that integrates existing water and ecological data. AB 1755 highlights the value of accessible, discoverable, and usable water data for both water managers and users; and a desire for increased transparency and collaboration among State agencies. AB 1755's specific requirements for protocol development are as follows:

The (California Department of Water Resources [DWR]), in consultation with the California Water Quality Monitoring Council, the state board, and the Department of Fish and Wildlife, **shall develop protocols for data sharing**, **documentation**, **quality control**, **public access**, **and promotion of open source platforms and decision support tools related to water data** (emphasis added). The department shall develop and submit to the Legislature, in compliance with Section 9795 of the Government Code and by January 1, 2018, a report describing these protocols. The report shall be developed in collaboration with the California Water Quality Monitoring Council, the state board, the Department of Fish and Wildlife, relevant federal agencies, and interested stakeholders, including, but not limited to, technology and open data experts and water data users. (California Water Code Section 12406 [a])

As described further in the *Progress Report for Implementing the Open and Transparent Water Data Act with Initial Draft Strategic Plan and Preliminary Protocols* (Progress Report), DWR has worked in consultation with its partner agencies, the State Water Resources Control Board, California Department of Fish and Wildlife, California Water Quality Monitoring Council, California Council on Science and Technology (CCST), UC Water, and others on development of protocols in response to AB 1755.

This report maps out AB 1755 protocol development in stages. The first stage identifies minimum initial protocols to support early implementation of an open data portal ("Initial Minimum Protocols" section). The second stage develops additional protocols identified by the Open Water Information Architecture (OWIA) and use cases ("Continuing Protocol Development – Use Cases and the Open Water Information Architecture" section). The third stage implements a mechanism for allowing protocols and standards to be created and modified ("Continuing Development of Protocols – Long Term" section). At each stage of the protocol development interoperability testbeds will be utilized to determine whether identified protocols and standards are effective in specific applications ("Interoperability Testbeds" section).

Foundational Concepts

The process of developing protocols for AB 1755 began with defining a few key concepts. Work on protocol development could not occur without a common understanding of the terms "protocols" and "interoperability." Likewise, much debate occurred over the need to include data standards. Ultimately, the following foundational concepts prevailed.

Protocols

AB 1755 provides clear direction for the development of protocols. But, AB 1755 does not define "protocols," which can connote different meaning depending on context. The following working definition of protocols was used in the development of this document and will be revisited in later phases of protocols development:

Protocols are methods of implementing a set of objectives and requirements in a systematic way. In computing, protocols mean both specific implementations of methods such as HTTP and FTP and, more generally as described by the Internet Engineering Task Force, protocols are sequences of processing steps that are also referred to as procedures.

Data Standards

Data standards will be integral to many of the protocols developed to meet AB 1755 requirements. The following working definition is adapted from the U.S. Geological Survey (USGS):

Data standards are sets of rules by which data or processes are described and recorded.

Standards are critical to sharing, exchanging, and understanding data in a meaningful way. The importance of data standards is well-stated by the U.S. Bureau of Land Management: "Standards provide data integrity, accuracy and consistency, clarify ambiguous meanings, minimize redundant data, and document business rules." To evolve a system that allows for meaningful exchange of data between groups it is necessary to agree on common data standards.

The USGS highlights issues that may arise from not implementing data standards: "If different groups are using different data standards, combining data from multiple sources is difficult, if not impossible. Utilizing data standards allows the agency to move from 'project-based' data files to 'enterprise' data files - and vice versa. In other words, the data become usable to more than just the project or person that created the data, because you know the data will be in an expected format and you know what is represented by the data."

When standards are executed correctly they become a cost-efficient way to assure interoperability between those who produce data and those who use data, across organizational boundaries. It is anticipated that future phases of protocol development will necessarily address data standards.

Interoperability

Interoperability is the ability of diverse computer systems, or software, to exchange and make use of common input data. Interoperability is critical to supporting decision-making, as it allows different

relevant datasets to be analyzed together. Increasing the ability for data sets and decision support tools to interoperate will, over time, yield better-informed decisions for water management. Many protocols and standards developed for AB 1755 will directly relate to, or support, increased interoperability. Over time, consensus on protocols and standards will lead to a higher level of interoperability.

Initial Minimum Protocols

As discussed in both the Progress Report and the Initial Draft Strategic Plan, the approach to implementation of AB 1755 involves starting with accessible products and adapting in response to user feedback, changing program needs, and policy decisions. To support the initial implementation of AB 1755, DWR has consulted with the partner agencies to outline three initial minimum protocols, consistent with available open data platforms, to guide early implementation of the program. The intent is to develop only what is necessary to facilitate early implementation to avoid creating barriers to sharing of data through an open data portal. These protocols will necessarily adapt over time in response to both changing software capabilities and the needs of the users of the open data portals to support a more efficient and transparent use of data. The section "Continuing Development of Protocols – Long Term," highlights a tentative process by which these protocols, and others developed in the interim, might be changed. Three initial protocols are presented in Table 1.

Protocol	Business Requirement
Identify a Data Steward	All datasets published by partner agencies on the open platform have partner Agency "owners," whom are responsible for maintaining and curating them for users.
Publish and Document on an Open Data Platform	All datasets published by partner agencies on the open platform have a place where they can be discovered.
Access Data	All datasets published by partner agencies on the open platform are machine readable, well documents and accessible to users.

Table 1	Three	Initial	Protocols	Develo	ned to	Supp	ort Farly	/ Im	plementatio	n of Al	B 1755
	111166	minai	1 10100013	Develo	peu io	Jupp		,	prementatio		0 17 33

DWR, in consultation with others, has identified these three protocols as being critical to enable early implementation of AB 1755. Each protocol is described below.

Identify a Data Steward

To facilitate dissemination of information and avoid orphaned datasets, each dataset on the open data platform must have a data steward assigned to it from the appropriate agency. The data steward is responsible for the data and for meeting any related data requests. This protocol allows for multiple levels of data stewardship, such as a data creator or author (originator of the data), data caretaker (inheritor, or external sponsor of the data), data sub-steward (person responsible for a subset of the data), and other roles beyond what is defined here. This protocol does not define specific roles for data stewards, it simply indicates the need to have at least one accessible person identified, and prescribes minimum required information for each data steward:

- Name of steward.
- Contact information.
- Organization.
- Roles.
- Dataset(s).

Only data stewards can publish, update, maintain, or remove datasets published on the platform, and each dataset that is published must be assigned to an active data steward from the appropriate agency. The next protocol addresses publication in more detail.

Publish and Document on an Open Data Platform

For data to be discoverable, it must be published to, or made available to, an open data platform. To be published, all data must meet the minimum documentation standards outlined in this section, including the metadata standard, the data dictionary requirements, and the guidelines for optional descriptive text. Requiring minimum documentation helps ensure these items can be found by users of an open data portal, and once a user has found the dataset, that sufficient documentation on the dataset is available to answer most of the users' questions. A sample technical workflow is provided in Table 2 as guidance for a data steward trying to publish their dataset on one of the existing open data portals.

 Table 2 A Sample Technical Workflow for Publishing a Dataset on data.ca.gov or the California

 Natural Resources Agency Open-Data Platform

Step	Activity	Actor	Required
1.	Log In	Data Steward	Always
2.	Ask system to create a new dataset	Data Steward	Always
3.	System creates new metadata template	System	Always
4.	Name dataset	Data Steward	Always
5.	Complete metadata for dataset	Data Steward	Always
	See Metadata Requirements, Machine Readability Requirements, and Guidelines for Optional Descriptive Text		
6.	Identify category for dataset	Curator	Always
7.	Upload data as resource(s) for dataset	Data Steward	Always
8.	Identify keywords for resource(s)	Data Steward	Always
9.	Complete data dictionary	Data Steward	If applicable
	See Data Dictionary Requirements		
10.	Identify keywords for dataset	Data Steward	Always
11.	Create API for dataset	System	If applicable
11.	Test API is functional	Data Steward	If applicable
12.	Notify curator that the dataset is published	Data Steward	Always
13.	Confirm dataset and resources appear and Test API	Curator	If applicable
14.	Check category	Curator	Always
15.	Check keywords	Curator	Always

Note: API = application programming interface

Metadata Requirements

To support initial implementation of AB 1755 on data.ca.gov and the California Natural Resources Agency (CNRA) Open Data Platform, the metadata requirements are to complete the metadata requirements identified by those portals. As additional necessary metadata elements are identified they will be added to the existing metadata requirements using a block structure format with the appropriate block elements, depending on the type of data. These requirements are included in the Appendix.

Data Dictionary Requirements

Similar to the metadata requirements, the data dictionary requirements are to follow those required by the respective open data portals, data.ca.gov and the CNRA Open Data Platform. These requirements are included in the Appendix.

Machine Readable Data Requirement

All tabular datasets published on an open platform must be machine readable. The Office of Management and Budget (OMB) describes machine readable format in Circular A-11 Part 6 as: "a standard computer language (not English text) that can be read automatically by a web browser or computer system. (e.g.; xml). Traditional word processing documents, hypertext markup language (HTML) and portable document format (PDF) files are easily read by humans but typically are difficult for machines to interpret. Other formats such as extensible markup language (XML), [JavaScript Object Notation] (JSON), or spreadsheets with header columns that can be exported as comma separated values (CSV) are machine readable formats. It is possible to make traditional word processing documents and other formats machine readable but the documents must include enhanced structural elements." (Project Open Data)

Guideline for Optional Descriptive Text

To better meet user needs and reduce quantity of inquiries related to the data, data stewards are encouraged to provide optional descriptive text using the following guidelines:

For all datasets:

- Purpose
- Public license
- Distribution and Reuse Conditions
- Version
- Applicable temporal range
- Temporal accuracy
- When data was collected or produced
- Applicable spatial range
- Spatial accuracy
- Management procedures
- Data quality procedures
 - Records
 - Dataset
- Explanation of all controlled vocabulary used
- Explanation of all field domains used

For observations:

- Observation methods
- Instruments
 - Instrument calibration
 - Instrument accuracy

For derivative products and datasets:

- Lineage
 - o Dataset
 - Reference
- Methods
 - Statistical formulae applied
 - Spatial aggregations
 - Temporal
 - aggregations
 - Models
- Consistency
- Completeness

Access Data

Well documented, published data with an appropriate data steward is not useful to the larger water community unless it is also accessible to the user of that data. To support this need, a sample workflow for user access to a dataset has been created in Table 3. While this workflow pertains to the user accessing the data set, it has significant implications on how state agencies should build platforms and organize data to support accessibility.

Step	Activity	Actor	Required
1.	Access open data portal via internet	Data Consumer	Always
2.	Search using keywords or tags	Data Consumer	Always
3.	Generate results list sorted by relevance to search terms	Open Data Platform	Always
4.	Select desired dataset from search results	Data Consumer	Always
5.	Take user directly to data or to data location	Open Data Platform	Always
6.	Query and visualize results using basic in-browser tools	Data Consumer	If applicable
7.	Download full or relevant queried portion of the dataset	Data Consumer	If applicable
8.	Connect to dataset directly via API	Data Consumer	If applicable

Table 3 A Sample Workflow for Accessing	Data on an Open Data Portal
---	-----------------------------

Note: API = application programming interface

Continuing Protocol Development - Use Cases and the Open Water Information Architecture

The next stage of protocol development will revolve around use cases and the OWIA. Use cases, as described in Data for Water Decision Making are "short examinations of how decision processes employ data – to inform a decision-driven water data system." In February and May 2017, DWR, CCST, and UC Water co-hosted workshops to engage stakeholders in the development of use cases for AB 1755. As a result of the workshops, 20 draft use cases were developed (Cantor 2018).

The OWIA, in turn, applied the use cases to develop a list of functional and technical requirements. The OWIA document addresses the intended outcomes (functional requirements) and system details (technical requirements) to ensure that both executives and engineers remain aligned in common purpose. The OWIA outlines the protocols, procedures, resources, governance, and minimum standard of technology required to meet the needs of California's water community, while also promoting greater levels of openness, transparency, and comparability for the information needed to manage water-related resources more effectively.

Each of the 41 functional and technical requirements identified in the OWIA is expected to result in a corresponding protocol. Protocols developed to meet the OWIA functional and technical requirements are mapped to AB 1755 protocol requirements in summarized form in Table 4. Table 5 includes the full list of functional and technical requirements as presented in the OWIA report. For more information about the functional and technical requirements, see the attached OWIA report.

OWIA	Required Protocol Category Identified by AB 1755								
Functional/Technical Requirement Categories	Data Documentation Sharing		Quality Public Control Access		Open Source Platforms	Decision Support Tools			
Data Acquisition	Х	•							
Quality Control	Х	Х	Х		Х	Х			
Publication	Х	Х		Х					
Data Traceability	Х	Х		Х					
System Portability					Х	Х			
External Interfaces	Х				Х	Х			

Table 4 Summary of the Functional and Technical Requirement Categories Specified in the OWIA to Protocol Categories Identified by AB 1755

Notes: AB 1755 = Assembly Bill 1755, The Open and Transparent Water Data Act; OWIA = Open Water Information Architecture

Table 5 Mapping of 41 Functional and Technical Requirement Identified in OWIA to Protocol Categories Identified by AB 1755 (Helly 2017)

Name	Data Sharing	Documentation	Quality Control	Public Access	Open-source platforms and decision support tools
Data Acquisition	Х				
*-Manual-	Х				
*-Automated-	Х				
Quality Control-*-	Х		Х		
*-Verification-	Х		Х		
--Documentation	Х	Х	Х		
--Reproducibility	Х		Х		
--Data Traceability	Х		Х		
*-Standardization-	Х	Х	Х		Х
--File-naming Conventions	Х	Х	Х		Х
*-Interoperable Transformation-	Х		Х		Х
--Separation of Data and Computation	Х		Х		Х
--Data Interoperability	Х	Х	Х		Х
--Products or Resources	Х		Х		Х
Publication-*-	Х	Х		Х	
*-Cross-Referencing-Service-	Х	Х		Х	
--Assignment of Digital Object Identifiers	Х	Х		Х	
*-Packaging-	Х			Х	
--Compression Methods	Х			Х	
--Archive File Formatting	Х			Х	
*-Archival-	Х			Х	
--Open Access Distribution	Х			Х	
Data Traceability-*-	Х	Х		Х	
*-Metadata Production-	Х	Х		Х	
*-Intellectual Property Rights Management-	Х	Х		Х	
*-Public Law Compliance-	Х	Х		Х	
*-Licensing-	Х	Х		Х	
*-Liability-	Х	Х		Х	
*-Searching-	Х			Х	
--Cross-referencing System Integration	Х			Х	
--Search Engine Optimization	Х			Х	
*-Version Control-	Х	Х			
--Binary Data	Х	Х			
--Non-Binary Data	Х	Х			
*-Anomaly Reporting-	Х	Х			
System Portability-*-					Х
*-Backup and Restore-					Х
*-Platform Portability-					Х
External Interfaces-*-	Х				Х
*-Data and Metadata Acquisition-	Х				Х
*-Data and Metadata Distribution-	X				X

Notes: AB 1755 = Assembly Bill 1755, The Open and Transparent Water Data Act; OWIA = Open Water Information Architecture

Continuing Development of Protocols – Long Term

The implementation of AB 1755 is based on the idea of iterative improvement as outlined in the *Initial Draft Strategic Plan.* To facilitate iterative improvement, a mechanism for adopting new protocols and standards is needed. To that end, the partner agencies are considering implementation of a modified form of the Internet Engineering Task Force (IETF) Internet Standards Process.

The IETF concept of working groups is particularly well-suited to AB 1755. Briefly, from the IETF website: "Working Groups (WGs) are the primary mechanism for development of IETF specifications and guidelines, many of which are intended to be standards or recommendations." Leveraging the IETF process, the partner agencies are considering the creation of a governance structure that could support the creation of a fluid set of groups to develop and adopt protocols and standards to support interoperability and smooth functioning of the AB 1755 open data portal.

Conceptually, this might take the form of three interactive groups, a policy group consisting of an AB 1755 governance structure, a stakeholder working group consisting of people working with the data or utilizing the data for decision making, and a technical working group consisting of the people collecting and providing the data. These groups would interact through the lens of the use cases, which define "WHO needs WHAT data in WHAT form for WHAT decisions." For example, the policy group could put forth a request to the stakeholder working group and technical working group to develop a standard to support interoperability. If the technical working group and stakeholder working group agree that the requested standard is needed, the two working groups would decide which group will develop the standard. Because of the technical nature of this particular request, the technical working group would develop the interoperability standard and pass it to the stakeholder working group for review. If the stakeholder working group agreed the proposed standard was feasible, it would then be returned to the policy group. If approved by the policy group, they would then formally codify the standard. A simple diagram of what this process might look like is provided in Figure 1.



Figure 1 Proposed Process for Protocol Development

Interoperability Testbeds

During the development and implementation of protocols and standards, there is a need for testing to see if the requirements imposed achieve the business requirements identified. This section discusses the concept of testing protocols against use cases, an activity termed interoperability testbeds.

An interoperability testbed allows innovative users to test proposed protocols and standards to address a specific use case. Conducting a series of testbeds is critical to the development and vetting of protocols in different applications of any given use case. Interoperability testbeds would include, at a minimum, State and federal agencies responsible for providing data under AB 1755. There are additional groups, not mentioned in the bill, who could be helpful in the formation of a long-term well-functioning water data system. These groups include other federal and State agencies, the research sector, local governments, non-governmental organizations, groundwater sustainability agencies, and the private sector.

Once participants in interoperability testbeds are selected, they would work together as a group to select data formats, quality assurance/quality control levels, update frequencies, and exchange protocol choices. It is also necessary to assure that relevant data is properly inventoried and available for the selected use cases. Conducting these testbeds will help to ensure that proposed protocols and standards utilized by the groups are beneficial in achieving the goals of the Strategic Plan, that data are accessible, sufficient, useful, and used. Each testbed would conclude with a documentation of lessons learned and analysis of beneficial, or non-beneficial, protocols and standards as well as recommendations for future actions.

References

- Cantor A., Kiparsky M., Kennedy R., Hubbard S., Bales R., Cano Pecharroman L., Guivetchi K., McCready C., Darling G. 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. 54 pp. Available at: https://doi.org/10.15779/J28H01 or law.berkeley.edu/datafordecisions
- Dodd, B. (2016). Assembly Bill No. 1755. The Open and Transparent Water Data Act. Available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB1755 Cal. Water Code §12400
- Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing, RFC 7230 ISSN: 2070-1721, IETF, June 2014.

Internet Engineering Task Force (IETF), The File Transfer Protocol, Rfc 354, IETF, 1972.

- Open Water Information Architecture AB 1755 Protocol Specification DRAFT Version 0 OWIA Technical Working Group 5, OWIA Technical Working Group, July 23, 2017
- Open Water Information Architecture Technical Working Group, OWIA System Requirements Document, Version 2.0, OWIA Technical Working Group, January 7, 2018

Project Open Data. 2017. Applying the Open Data Policy to Federal Awards. Viewed online at: https://project-open-data.cio.gov/federal-awards-faq/. Accessed: Jan. 2018

The Internet Standards Process – Revision 3, Tech. Report RFC 2026, IETF, October 1996.

U.S. Geologic Survey. 2018. USGS Data Management. Viewed online at: https://www2.usgs.gov/datamanagement/plan/datastandards.php. Accessed: Jan. 2018

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.

API: An application programming interface is a set of programming library calls and supporting compile and run-time libraries. These exist on both the client-side and server-side of a computer application although they are usually asymmetrical in terms of what the server implements versus what the client implements. The purpose for providing APIs is to standardize and simplify the programming required to add functionality to a software application and to enhance the portability and interoperability of software across both platforms and data.

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.

Federation: A federation is a group of data providers and users using jointly agreed-upon standards of operation in a collective fashion to ensure the interoperability of the resources they collectively hold and employ. The term may be used, for example, when describing the interoperation of distinct cyber infrastructure networks with different internal structures. The term may also be used when human groups agree to collectively manage cyberinfrastructure development and operation using commonly held, and managed, requirements, standards and conventions, and operating procedures to ensure the interoperability of distinct cyberinfrastructure resources.

Federated data system: 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.

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 diverse computer systems or software to exchange and make use of common input data.

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.

Procedures: An established or official way of doing something.

Protocol: Protocols are methods of implementing a set of objectives and requirements in a systematic way. In computing, protocols mean both specific implementations of methods such as HTTP and FTP and, more generally as described by the Internet Engineering Task Force, protocols are sequences of processing steps that are also referred to as procedures.

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.

Appendix A

Open Water Information Architecture System Requirements Document

Open Water Information Architecture System Requirements Document Version 2.0

OWIA Technical Working Group

January 24, 2018

Contents

1	Intr	roduction												
2	Proi	niect Management Approach												
	2.1	The Differences Between Objectives, Functional Requirements Technical Requirements												
		and Design Alternatives												
		č												
3	Con	oncept of Operation												
	3.1	Definition of the OWIA System												
	3.2	Governance												
		3.2.1 Technical Working Group (TWG)												
		3.2.2 Stakeholder Working Group (SWG)												
4	Fun	Functional Requirements												
	4.1	FR-100-100: Data Acquisition												
		4.1.1 FR-100-110: Manual												
		4.1.2 FR-100-120: Automated												
	4.2	FR-200-100: Quality Control												
		4.2.1 FR-200-110: Verification												
		4.2.1.1 FR-200-120: Documentation												
		4.2.1.2 FR-200-130: Reproducibility												
		4.2.1.3 FR-200-140: Data Traceability												
		4.2.2 FR-200-150: Standardization												
		4.2.2.1 FR-200-160: Metadata Conventions												
		4.2.2.2 FR-200-160: File-naming Conventions												
		4.2.3 FR-200-170: Interoperable Transformation												
		4.2.3.1 FR-200-180: Separation of Data and Computation												
		4.2.3.2 FR-200-190: Data Interoperability												
		4.2.3.3 FR-200-200: Products or Resources												
	4.3	FR-300-100: Publication												
		4.3.1 FR-300-110: Cross-Referencing-Service												
		4.3.1.1 FR-300-120: Assignment of Digital Object Identifiers												

		4.3.2	FR-300-130: Packaging	6
			4.3.2.1 FR-300-140: Compression Methods	7
			4.3.2.2 FR-300-150: Archive File Formatting	7
		4.3.3	FR-300-160: Archival	7
			4.3.3.1 FR-300-170: Open Access Distribution	7
	4.4	FR-40	0-100: Data Traceability	7
		4.4.1	FR-400-110: Metadata Production	7
		4.4.2	FR-400-120: Intellectual Property Rights Management	7
		4.4.3	FR-400-130: Public Law Compliance	7
		4.4.4	FR-400-140: Licensing	7
		4.4.5	FR-400-150: Liability	7
		4.4.6	FR-400-160: Searching	7
			4.4.6.1 FR-400-170: Cross-referencing System Integration	7
			4.4.6.2 FR-400-180: Search Engine Optimization	7
		4.4.7	FR-400-190: Version Control	8
			4.4.7.1 FR-400-200: Binary Data	8
			4.4.7.2 FR-400-210: Non-Binary Data	8
		4.4.8	FR-400-220: Anomaly Reporting	8
	4.5	FR-50	0-100: System Portability	8
		4.5.1	FR-500-110: Backup and Restore	8
		4.5.2	FR-500-120: Platform Portability	8
	4.6	FR-60	0-100: External Interfaces	8
		4.6.1	FR-600-110: Data and Metadata Acquisition	8
		4.6.2	FR-600-120: Data and Metadata Distribution	8
_				_
5	Tech	nnical R	Requirements	8
5	Tech 5.1	nnical R TR-10	Requirements 10-100-00100: Data Acquisition Methods	8 10
5	Tech 5.1	nnical R TR-10 5.1.1	Requirements 00-100-00100: Data Acquisition Methods TR-200-200-00100: Data Transformation Methods	8 10 10
5	Tech 5.1	TR-10 5.1.1 5.1.2	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming Languages	8 10 10 10
5	Tech 5.1	TR-10 5.1.1 5.1.2 5.1.3	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data Interoperability	8 10 10 10 10
5	Tech 5.1	TR-10 5.1.1 5.1.2 5.1.3 5.1.4	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)	8 10 10 10 10 10
5	Tech 5.1	TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data Traceability	8 10 10 10 10 10 10
5	Tech 5.1	TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Reproducibility of Data Products	8 10 10 10 10 10 10 10
5	Tech 5.1	TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Reproducibility of Data Products00-100-00100: Standardization	8 10 10 10 10 10 10 10 10 10 10 10 10 10
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products0-100-00100: StandardizationTR-400-100-00200: Metadata	8 10 10 10 10 10 10 10 10
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products0-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts	 8 10
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Reproducibility of Data Products0-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00400: Controlled Vocabulary	 8 10
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products00-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00500: Climate and Forecasting Conventions	 8 10
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-300-00100: Programming LanguagesTR-200-5000-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products00-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00500: Climate and Forecasting Conventions5.2.1.4TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospa-	8 10 10 10 10 10 10 10 10 10 10 10
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products0-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00500: Climate and Forecasting Conventions5.2.1.4TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines	 8 10 <l< td=""></l<>
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products00-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00500: Climate and Forecasting Conventions5.2.1.4TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines5.2.1.5TR-400-100-00700: Open-Geospatial Consortium Standards and Guide-	 8 10 <l< td=""></l<>
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Reproducibility of Data Products00-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00500: Climate and Forecasting Conventions5.2.1.4TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines5.2.1.5TR-400-100-00700: Open-Geospatial Consortium Standards and Guidelines	 8 10 <l< td=""></l<>
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1	Requirements00-100-00100: Data Acquisition MethodsTR-200-200-00100: Data Transformation MethodsTR-200-300-00100: Programming LanguagesTR-200-400-00100: Data InteroperabilityTR-200-5000-00100: Products (List Derived Products traceable to Objectives)TR-300-100-00100: Data TraceabilityTR-300-100-00100: Data TraceabilityTR-300-100-00200: Reproducibility of Data Products00-100-00100: StandardizationTR-400-100-00200: Metadata5.2.1.1TR-400-100-00300: Resolution of Metadata conflicts5.2.1.2TR-400-100-00400: Controlled Vocabulary5.2.1.3TR-400-100-00500: Climate and Forecasting Conventions5.2.1.4TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines5.2.1.5TR-400-100-00700: Open-Geospatial Consortium Standards and GuidelinesTR-400-100-00800: Mapping Standards	 8 10 11
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1 5.2.2 5.2.2 5.2.3	Requirements 00-100-00100: Data Acquisition Methods TR-200-200-00100: Data Transformation Methods TR-200-300-00100: Programming Languages TR-200-400-00100: Data Interoperability TR-200-5000-00100: Products (List Derived Products traceable to Objectives) TR-300-100-00100: Data Traceability TR-300-100-00100: Data Traceability TR-300-100-00200: Reproducibility of Data Products 00-100-00100: Standardization TR-400-100-00200: Metadata 5.2.1.1 TR-400-100-00200: Metadata 5.2.1.2 TR-400-100-00300: Resolution of Metadata conflicts 5.2.1.3 TR-400-100-00500: Climate and Forecasting Conventions 5.2.1.4 TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines 5.2.1.5 TR-400-100-00700: Open-Geospatial Consortium Standards and Guidelines TR-400-100-00800: Mapping Standards TR-400-100-00800: Mapping Standards	 8 10 11 11
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1 5.2.2 5.2.2 5.2.3 5.2.4	Requirements 00-100-00100: Data Acquisition Methods TR-200-200-00100: Data Transformation Methods TR-200-300-00100: Programming Languages TR-200-400-00100: Data Interoperability TR-200-5000-00100: Products (List Derived Products traceable to Objectives) TR-300-100-00100: Data Traceability TR-300-100-00100: Data Traceability TR-300-100-00200: Reproducibility of Data Products 00-100-00100: Standardization TR-400-100-00200: Metadata 5.2.1.1 TR-400-100-00300: Resolution of Metadata conflicts 5.2.1.2 TR-400-100-00300: Controlled Vocabulary 5.2.1.3 TR-400-100-00500: Climate and Forecasting Conventions 5.2.1.4 TR-400-100-00500: Climate and Forecasting Conventions 5.2.1.5 TR-400-100-00700: Open-Geospatial Consortium Standards and Guide- lines	 8 10 11 11 11
5	Tech 5.1	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1 5.2.2 5.2.2 5.2.3 5.2.4 5.2.5	Requirements 00-100-00100: Data Acquisition Methods TR-200-200-00100: Data Transformation Methods TR-200-300-00100: Programming Languages TR-200-400-00100: Data Interoperability TR-200-5000-00100: Products (List Derived Products traceable to Objectives) TR-300-100-00100: Data Traceability TR-300-100-00100: Products (List Derived Products traceable to Objectives) TR-300-100-00200: Reproducibility of Data Products 00-100-00100: Standardization TR-400-100-00200: Metadata 5.2.1.1 TR-400-100-00200: Metadata 5.2.1.2 TR-400-100-00300: Resolution of Metadata conflicts 5.2.1.3 TR-400-100-00500: Climate and Forecasting Conventions 5.2.1.4 TR-400-100-00500: Climate and Forecasting Conventions 5.2.1.5 TR-400-100-00700: Open-Geospatial Consortium Standards and Guide- lines	 8 10 11 11 11 11
5	Tech 5.1 5.2 5.3	nnical R TR-10 5.1.1 5.1.2 5.1.3 5.1.4 5.1.5 5.1.6 TR-40 5.2.1 5.2.2 5.2.2 5.2.3 5.2.4 5.2.5 TR-60	Requirements 00-100-00100: Data Acquisition Methods TR-200-200-00100: Data Transformation Methods TR-200-300-00100: Programming Languages TR-200-400-00100: Data Interoperability TR-200-5000-00100: Products (List Derived Products traceable to Objectives) TR-300-100-00100: Data Traceability TR-300-100-00100: Data Traceability TR-300-100-00200: Reproducibility of Data Products 00-100-00100: Standardization TR-400-100-00200: Metadata 5.2.1.1 TR-400-100-00200: Metadata 5.2.1.2 TR-400-100-00300: Resolution of Metadata conflicts 5.2.1.3 TR-400-100-00500: Climate and Forecasting Conventions 5.2.1.4 TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines 5.2.1.5 TR-400-100-00700: Open-Geospatial Consortium Standards and Guidelines TR-400-100-00800: Mapping Standards TR-400-100-00900: Numerical Accuracy and Precision Standards TR-400-100-0100: Measures of Uncertainty TR-400-100-01100: File Naming Convention 00-100-00100: Data Publication 00-100-00100: Data Publication	 8 10 11 11 11 11 11

		5.3.1.1 TR-600-100-00300: Assignment of Digital Object Identifiers	11
	5.3.2	TR-600-100-00400: Metadata Production	11
:	5.3.3	TR-600-100-00500: Open Access Distribution	11
:	5.3.4	TR-600-100-00600: Intellectual Property Rights Management	11
		5.3.4.1 TR-600-100-00700: Public Law Compliance	11
		5.3.4.2 TR-600-100-00800: Licensing	11
:	5.3.5	TR-600-100-00900: Discovery	11
		5.3.5.1 TR-600-100-01000: Cross-referencing System Integration	11
		5.3.5.2 TR-600-100-01100: Search Engine Optimization	11
:	5.3.6	TR-600-100-01200: Packaging	11
		5.3.6.1 TR-600-100-01300: Compression Methods	11
		5.3.6.2 TR-600-100-01400: Archive File Formatting	11
:	5.3.7	TR-600-100-01500: Version Control	11
		5.3.7.1 TR-600-100-01600: Binary Data	11
		5.3.7.2 TR-600-100-01700: Non-Binary Data	11
:	5.3.8	TR-600-100-01800: Anomaly Reporting	11
5.4	TR-60)-100-00100: System Interoperability	11
:	5.4.1	TR-600-100-00200: Backup and Restore	11
:	5.4.2	TR-600-100-00300: Platform Portability	11
Appendix	x A S	tandards and Conventions	12
A.1	Operat	ing Systems (OS)	12
A.2	Metada	ata Schema	12
A.3	Contro	lled Vocabulary	12
A.4	Georef	erencing	12
A.5	Intelle	ctual Property Rights	12
A.6 ′	Trustee	Archives	13
A.7	Digital	Objects	13
A.8	File Fo	rmats and Data Encodings	13
A.9	Cross-	referencing Services	15
A.10	Comm	ercial Search Services	15
Appendix	x B N	letadata Schema	15
Appendix	x C S	upport for AB1755	17
Appendix	хDТ	raceability Tables	19

List of Figures

1	Relationship between system baseline documents and operations	1
2	Conceptual representation of the system engineering process. <i>Reproduced from [19].</i>	3

3	Illustration of the OWIA federation concept with a triumvirate governance structure of part-	
	ners (OWIA-GP) supported by interacting stakeholder working group (SWG) and the tech-	
	nical working group (TWG). The federation is comprised of dedicated OWIA system imple-	
	mentations to enable individual data providers to independently integrate the OWIA system	
	into their existing methods and procedures within their operations. Shared OWIA system	
	implementations provide the flexibility for the harvesting of non-compliant data sources into	
	an OWIA system implementation that will support the OWIA federation without insisting	
	that the producers be OWIA compliant.	4
4	OWIA system (cf. section 3.1) functional block diagram. Parenthetical references point to	
	the governing functional requirements.	9

List of Tables

1	Definition of OWIA classes of datasets.	9
2	Classification of file formats, content and related interoperability features. References are	
	supplied in brackets and listed in the bibliography.	14
3	Traceability of AB1755 objectives (columns) to OWIA SOPs by use case (rows).	18
4	Traceability Table: Objective O-1100-1000 to Functional Requirements. This is an example	
	of what subordinate Technical Requirements might resolve to and is meant only to charac-	
	terize what <i>Resolution</i> of Functional Requirements might look like in a Technical Proposal.	20

1 Introduction

This document contains the functional and technical requirements for the *Open Water Information Architecture (OWIA)* and is called the *OWIA System Requirements Document (SRD)*. It has within it an Appendix:B Standards and Conventions that contains narrative explanations that are referred to within individual requirements where appropriate. This is done because the requirements are meant to be terse, declarative, testable statements that are not overloaded with narrative exposition. There are two companion documents to the SRD: (1) the subordinate document *OWIA Standard Operating Procedures (SOPs)* and the (2) parent document *California Council for Science and Technology (CCST) Stakeholder Use Case* document.

The SOPs are compliant with the requirements specified here yet written at a more detailed level of abstraction with examples of programming code or sometimes pseudo-code to exemplify the implementation details important to developers as well as precisely documenting the processing steps (i.e., procedures) used to operate on data. It is meant to be analogous to an *OWIA Programmer's Guide* and, as the OWIA implementation proceeds, there will be open-source code repositories with *minimal working examples (MWE)* for use in improvements and innovations to current procedures.

Each of these documents is intended for a technical audience although it is hoped that they are comprehensible to a motivated non-technical reader. There is a glossary in the back of the SRD to aid in navigating the technical language and as an effort to disambiguate some of the terms for which there may be competing and inconsistent def-



Figure 1: Relationship between system baseline documents and operations.

initions. In addition to these two, there is a third document that contains the stakeholder use cases used to develop the stakeholder objectives from each use case. These objectives are being used to define and constrain the requirements contained in the SRD and the procedures for satisfying them defined in the SOPs.

The SRD and SOPs are designed to provide a foundation for a community-based *OWIA* development of a federated set of cyberinfrastructure resources (i.e., computers, networks, data, metadata, and standards and conventions) that are interoperable and highly-automated to minimize labor as well as idiosyncratic anomalies. We therefore refer to them as the *baseline documents* (Figure 1). The objective of these baseline documents is to establish a framework for sustainable water resource management and to formalize that framework to a degree exemplified by other systems of standard methods such as those found in [7].

The federated nature of the OWIA extends to its (1) human governance structure as well as its (2) cyberinfrastructure (cf. Section 3 and Figure 3). Therefore we speak of the OWIA federation as including both these aspects and will differentiate the two parts contextually when using the term. The *open* aspect means open-access, open-source and open-architecture: encouraging innovation and automation while precluding the siloing and stove-piping that occurs when proprietary software and systems pose restrictive technology dependencies and requirements. The planning horizon is open-ended although intended to provide for a near-term operational system with an initial operating capability (IOC) within 1-2 years evolving to a final operating capability (FOC) over five (5) years that is operationally sustainable while responsive to technology innovation and risk minimization (i.e, cost, schedule, technical and operational) over its lifetime.

The approach is to follow standard system engineering practices [32] that: (1) define stakeholder ob-

jectives and, from these, (2) enumerate functional requirements in terms of functional components and major interfaces both of which are implementation-independent, and (3) enumerate technical requirements which specify fundamental technical features such as network transfer rates, storage capacities, reliability, maintainability and availability (RMA), interface dependencies and contingencies and similar quantitative or qualitative requirements at a level of specificity (or abstraction) that is more detailed than the functional requirements on which they are based. It is also designed to present an initial evaluation of some of the obvious design trade-studies to explicate and focus on the key risk areas related to technical, schedule, cost and operational risks.

This is an interative and recursive, hierarchical design approach (Figure 2) which prioritizes *Stakeholder Objectives, Functional Requirements, and Technical Requirements* respectively and cross-correlates them to each other via a *traceability matrices* (Section 4) to ensure that there are no *widows or orphans* in the sense that there are no unsupported Objectives or Functional Requirements (i.e., widows) as well as no lower-level design features that are not specified in the Functional Requirements (i.e., orphans). As a development methodology, the system engineering method used here is sometimes contrasted with the agile development method. Every methodology has pros and cons and the reason we use this approach for the OWIA is because we already know a great deal about what is needed to improve access-to and reuse-of the collective set of water resource data and the OWIA focus is on the data content. This is not primarily a process of discovery and prototyping of software applications. For a broader discussion of the pros and cons of alternative software development approaches, the reader is encouraged to consider the discussions provided in [27] and [32].

Finally, some historical perspective is helpful. This document is meant to integrate the thinking on water resource information broadly and digital data about water resources specifically. The OWIA concept developed independently of the AB1755 legislation [1][30] that is currently, as of this writing, driving many efforts across the State of California to comply with its mandates and schedule. Fortuitously, the development of the OWIA and the activation of AB1755-related efforts overlap strongly such that AB1755 requirements are a subset of the broader OWIA requirements. The implementation of the OWIA will satisfy the requirements of AB1755 and support the Sustainable Groundwater Management Act (SGMA) in such a way that we can treat AB1755 as an OWIA use-case as described in Appendix C. The OWIA concept is a reflection and integration of a wide range of on-going efforts especially those in the UC WATER Security and Sustainability Research Initiative and CITRIS [9], California Council on Science and Technology (CCST), the Center for Western Weather and Water Extremes (CW3E)[38], the San Diego Supercomputer Center (SDSC) [4, 16, 12, 18, 15, 41, 35, 14, 31, 5, 11, 36, 17, 6, 2, 13, 20] and the UC Santa Barbara Bren School. We expect to grow this community to include private California universities, national laboratories and private sector partners as we go.

2 Project Management Approach

Figure 2 depicts the overall management approach used for this project and graphically summarizes it in the context of a *system engineering framework*. The system engineering framework is a set of methods, procedures, standards and conventions, for specifying design constraints to minimize the risk that a sought-after system implementation will successfully perform its intended functions. The output of this process is a system design and implementation that is *verifiably compliant with the functional and technical requirements* for the system and which can be *validated against the stakeholder objectives*.

2.1 The Differences Between Objectives, Functional Requirements, Technical Requirements and Design Alternatives

In designing and building any type of system the first step is to describe *what it is that you want to accomplish by building the system. These are the* **objectives**. Sometimes these are called *stakeholder*



Figure 2: Conceptual representation of the system engineering process. *Reproduced from [19]*.

objectives. They should be stated as simple declarative sentences focused on what the stakeholder wants the system to do. The language should be as non-technical as possible in order for the broadest understanding and consensus across the lay stakeholders who typically have diverse backgrounds and experience. On the other hand, the *functional requirements* are the translation of the objectives into engineering terms (i.e., functions using more precise technical language) describing how the objectives will be met. This is the first level of abstraction in specifying how the system will be implemented (Figure 2).

The articulation of objectives is often a stumbling block for stakeholders and developers alike because it poses a bit of chicken and egg or floor versus ceiling ambiguity. The way to get past this is focus on the description, often through workshops of stakeholders, of examples of usage without worrying very much whether something is an objective or a functional requirement. These can be re-factored by a *technical working group* once they are articulated. The most important thing is to articulate and document examples of how the system will be used. Once this process has an initial iteration, the system requirements document (SRD), can be maintained through configuration management of versions over time to provide accountability, via traceability tables, to the stakeholders as well as a path forward for developers and a contractually-applicable basis for acceptance testing for contract managers.

3 Concept of Operation

The OWIA system will be operated to produce standardized data of documented quality needed by stakeholders, as defined by the objectives, such as (1) figures, (2) tables, (3) statistics, and (4) analyses.

3.1 Definition of the OWIA System

The *OWIA system* is a federated collection of data systems, cooperating through a shared governance process, standards and conventions, standard interfaces (i.e., APIs, network protocols), procedures and algorithms, and verifiable compliance with requirements. The system structure supports adherence to open-data standards and principles and guided by a set of functional requirements such as accessibility, interoperability, discoverability, and traceability.

Open-data means that the data are free to use, re-use, and redistribute with no restrictions on their use [43, 39]. A federated system does not require a central catalogue or a single interface to such a catalogue. Rather, the federation permits the formation of one or more common catalogues (cf. Figure 4, Curated Collections) as an implementation dependent on the requirement for standardized, published metadata.



Figure 3: Illustration of the OWIA federation concept with a triumvirate governance structure of partners (OWIA-GP) supported by interacting stakeholder working group (SWG) and the technical working group (TWG). The federation is comprised of dedicated OWIA system implementations to enable individual data providers to independently integrate the OWIA system into their existing methods and procedures within their operations. Shared OWIA system implementations provide the flexibility for the harvesting of non-compliant data sources into an OWIA system implementation that will support the OWIA federation without insisting that the producers be OWIA compliant.

3.2 Governance

The OWIA governance structure is modeled on that of the Internet Engineering Task Force (IETF) [24]. The IETF governance structure is modified to reflect the exigencies of the needs of California stakeholders and the mission agencies responsible for management leadership: the OWIA Governance Partners (OWIA-GP).

The OWIA-GP are responsible for the direction of the system, policy, prioritization and resourcing of work, and curation of the OWIA system baseline. This group cooperates with Stakeholder Working Group (SWG) and the Technical Working Group (TWG) to ensure that the OWIA federation.

3.2.1 Technical Working Group (TWG)

The *Technical Working Group (TWG)* is responsible for the identifying, adopting, approving data standards, data publication approaches, and controlling functional and technical requirements.

3.2.2 Stakeholder Working Group (SWG)

The *Stakeholder Working Group (SWG)* is responsible for informing the OWIA Governance Partners of the intended uses of the OWIA system, and providing iterative feedback on the effectiveness of the system to meet their requirements.

4 Functional Requirements

4.1 FR-100-100: Data Acquisition

Manual and automated methods shall be provided for data acquisition. Data at the acquisition stage of OWIA processing shall be referred to as *Level 0* data.

4.1.1 FR-100-110: Manual

Manual data acquisition methods shall provide metadata conforming the the OWIA minimum metadata standard.

4.1.2 FR-100-120: Automated

Automated data acquisition methods shall provide metadata conforming to the OWIA minimum metadata standard.

4.2 FR-200-100: Quality Control

No data transformation shall require the use of proprietary software, methods or special-purpose computing platforms for data processing and transportation. Data that has received quality control processing according to OWIA standards and conventions shall be referred to as *Level 1* data.

4.2.1 FR-200-110: Verification

Data verification shall be accomplished according to OWIA standards and conventions (cf. Appendix A).

4.2.1.1 FR-200-120: Documentation Documentation shall be provided according to OWIA standards and conventions (cf. Appendix A).

4.2.1.2 FR-200-130: Reproducibility All data products shall be verifiably reproducible by an anonymous second-party from the input data, metadata and the processing methods used to produce the data product.

4.2.1.3 FR-200-140: Data Traceability All data products shall be traceable to their parent data sources to the extent that a data product composed of multiple input data sources shall be decomposable and traceable to its parents.

4.2.2 FR-200-150: Standardization

All data products shall conform to the OWIA standards and conventions (cf. Appendix A).

4.2.2.1 FR-200-160: Metadata Conventions Metadata shall be provided according to OWIA standards and conventions (cf. Appendix A).

4.2.2.2 FR-200-160: File-naming Conventions File-name shall be performed according to OWIA standards and conventions (cf. Appendix A).

4.2.3 FR-200-170: Interoperable Transformation

All data transformations shall be achievable with open-source, non-proprietary software, non-proprietary data formats and commodity computers.

4.2.3.1 FR-200-180: Separation of Data and Computation Data and computation shall be separated between data files and stored procedures.

4.2.3.2 FR-200-190: Data Interoperability All data products shall be interoperable across OWIA-supported computing platforms and be able to be operated on using non-proprietary, open-source software and commodity computers and communications systems to operate on them or transport them.

4.2.3.3 FR-200-200: Products or Resources Data products shall be developed in accordance with the objectives as per section **??**.

4.3 FR-300-100: Publication

Data shall be published according to OWIA standards and conventions (cf. Appendix A).

4.3.1 FR-300-110: Cross-Referencing-Service

Data objects shall be registered with a cross-referencing service.

4.3.1.1 FR-300-120: Assignment of Digital Object Identifiers A digital object identifier (DOI) shall be acquired for each Level 1 digital object according to the OWIA standards and conventions (cf. Appendix A).

4.3.2 FR-300-130: Packaging

Packaging shall conform to OWIA standards and conventions (cf. Appendix A).
4.3.2.1 FR-300-140: Compression Methods Compression methods shall be non-lossy and conform to OWIA standards and conventions (cf. Appendix A).

4.3.2.2 FR-300-150: Archive File Formatting Archive file formats shall be only those conforming with OWIA standards and conventions.

4.3.3 FR-300-160: Archival

Data shall be archived in trusted data archives with external interfaces to provide for data access and transportation to end-users and applications.

4.3.3.1 FR-300-170: Open Access Distribution All data products shall be accessible using OWIA standard protocols or transportable external media where network transport is impossible or impractical.

4.4 FR-400-100: Data Traceability

Data traceability shall be provided according to OWIA standards and conventions (cf. Appendix A).

4.4.1 FR-400-110: Metadata Production

All data products shall have metadata provided with them sufficient to meet the OWIA minimum metadata standard.

4.4.2 FR-400-120: Intellectual Property Rights Management

Metadata shall be produced according to the OWIA standards and conventions (cf. Appendix A).

4.4.3 FR-400-130: Public Law Compliance

All data and metadata products shall comply with relevant public law requirements.

4.4.4 FR-400-140: Licensing

Licensing of data and metadata shall conform to OWIA standards and conventions (cf. Appendix A).

4.4.5 FR-400-150: Liability

Liability limitations shall be declared with each data object through the metadata in conformance with OWIA standards and conventions (cf. Appendix A).

4.4.6 FR-400-160: Searching

Minimal metadata shall be provided to meet OWIA standards and conventions for search and discovery.

4.4.6.1 FR-400-170: Cross-referencing System Integration Cross-referencing system integration shall be based on a digital objects DOI.

4.4.6.2 FR-400-180: Search Engine Optimization Search-engine optimization shall be based on the metadata associated with the DOI as a minimum.

4.4.7 FR-400-190: Version Control

Source code, data and metadata shall be version-controlled in conformance with OWIA standards and conventions (cf. Appendix A).

4.4.7.1 FR-400-200: Binary Data A version control method shall be provided for binary data products in accordance with OWIA standards and conventions (cf. Appendix A).

4.4.7.2 FR-400-210: Non-Binary Data A version control method shall be provided for non-binary data products in accordance with OWIA standards and convetions.

4.4.8 FR-400-220: Anomaly Reporting

There shall be a method for reporting of anomalies detected in the data products and there shall be a method of tracking the anomalies for resolution and notification of those that have obtained the anomalous data products that an anomaly has been reported and a method for determining the resolution of the anomaly.

4.5 FR-500-100: System Portability

System portability shall be provided such that any implementation of the OWIA system is portable to other platforms in conformance with OWIA standards and conventions (cf. Appendix A).

4.5.1 FR-500-110: Backup and Restore

Backup and restore capability shall be provided using interoperable procedures and systems according to OWIA standards and conventions (cf. Appendix A).

4.5.2 FR-500-120: Platform Portability

Platform portability shall be provided in conformance with OWIA standards and conventions (cf. Appendix A).

4.6 FR-600-100: External Interfaces

External interfaces shall be provided for data acquisition and open-access to data products.

4.6.1 FR-600-110: Data and Metadata Acquisition

External interfaces shall be provided supporting the data sources in Appendix ??.

4.6.2 FR-600-120: Data and Metadata Distribution

External interfaces to end-users and applications shall be provided supporting those listed in ??.

5 Technical Requirements

The technical requirements are subordinate and traceable to the functional requirements above and have a higher-level of engineering detail and more precise, more technical language. This is the first level sufficiently specific to require decision-making about engineering trade-offs and what types of hardware, software and data representations qualify to be included in an OWIA-node.



Figure 4: OWIA system (cf. section 3.1) functional block diagram. Parenthetical references point to the governing functional requirements.

Dataset Class	Provenance	Transformation	Quality Control	Metadata	Published
Level 0	cf. Table 2	N/A	N/A	N/A	N/A
Level 1	Traceable to Level 0 parent	Interoperable	SOPs	Listing 1	DOI, Archived
Level 2	Traceable to multiple Level 1 parents (composite, decomposable)	Interoperable	SOPs	Listing 1	DOI, Archived

 Table 1: Definition of OWIA classes of datasets.

5.1 TR-100-100-00100: Data Acquisition Methods

5.1.1 TR-200-200-00100: Data Transformation Methods

All data transformations shall be accomplished with open-source, non-proprietary software and commodity computers for data processing and transportation.

5.1.2 TR-200-300-00100: Programming Languages

Data processing shall be realized through the use of stored procedures written in the GNU programming languages.

5.1.3 TR-200-400-00100: Data Interoperability

All data products shall be interoperable across OWIA-supported computing platforms and be able to be operated on using non-proprietary, open-source software and commodity computers and communications systems to operate on them or transport them.

5.1.4 TR-200-5000-00100: Products (List Derived Products traceable to Objectives)

5.1.5 TR-300-100-00100: Data Traceability

All data products shall be traceable to their parent data sources to the extent that a data product composed of multiple input data sources shall be decomposable and traceable to its parents.

5.1.6 TR-300-100-00200: Reproducibility of Data Products

All data products shall be verifiably reproducible by an anonymous second-party from the input data and the method used to produce the data product.

5.2 TR-400-100-00100: Standardization

5.2.1 TR-400-100-00200: Metadata

All data products shall have metadata provided with them sufficient to meet the OWIA minimum metadata standard.

5.2.1.1 TR-400-100-00300: Resolution of Metadata conflicts All conflicts in metadata standards and conventions shall be subject to the determination of the TWG.

5.2.1.2 TR-400-100-00400: Controlled Vocabulary All metadata shall be populated with controlled vocabulary drawn from the following standards.:

5.2.1.3 TR-400-100-00500: Climate and Forecasting Conventions Climate and Forecasting conventions shall be used when suitable.

5.2.1.4 TR-400-100-00600: Federal Geographic Data Committee (FGDC) Geospatial Metadata Standards And Guidelines FGDC standards and guidelines shall be used when suitable.

5.2.1.5 TR-400-100-00700: Open-Geospatial Consortium Standards and Guidelines Open-Geospatial Consortium standards and guidelines shall be used when suitable.

- 5.2.2 TR-400-100-00800: Mapping Standards
- 5.2.3 TR-400-100-00900: Numerical Accuracy and Precision Standards
- 5.2.4 TR-400-100-01000: Measures of Uncertainty
- 5.2.5 TR-400-100-01100: File Naming Convention
- 5.3 TR-600-100-00100: Data Publication
- 5.3.1 TR-600-100-00200: Methods
- 5.3.1.1 TR-600-100-00300: Assignment of Digital Object Identifiers
- 5.3.2 TR-600-100-00400: Metadata Production
- 5.3.3 TR-600-100-00500: Open Access Distribution

5.3.4 TR-600-100-00600: Intellectual Property Rights Management

All data products shall be accessible using OWIA standard protocols or transportable external media where network transport is impossible or impractical.

5.3.4.1 TR-600-100-00700: Public Law Compliance

- 5.3.4.2 TR-600-100-00800: Licensing
- 5.3.5 TR-600-100-00900: Discovery
- 5.3.5.1 TR-600-100-01000: Cross-referencing System Integration
- 5.3.5.2 TR-600-100-01100: Search Engine Optimization
- 5.3.6 TR-600-100-01200: Packaging
- 5.3.6.1 TR-600-100-01300: Compression Methods
- 5.3.6.2 TR-600-100-01400: Archive File Formatting
- 5.3.7 TR-600-100-01500: Version Control
- 5.3.7.1 TR-600-100-01600: Binary Data
- 5.3.7.2 TR-600-100-01700: Non-Binary Data
- 5.3.8 TR-600-100-01800: Anomaly Reporting
- 5.4 TR-600-100-00100: System Interoperability
- 5.4.1 TR-600-100-00200: Backup and Restore
- 5.4.2 TR-600-100-00300: Platform Portability

Appendix A Standards and Conventions

This appendix is a narrative description of the standards and conventions referred to in the functional and technical requirements. The Internet Engineering Task Force is used as a reference and as the default for standards and conventions that are not otherwise superseded by those identified here. For the subset of standards and conventions that pertain only to web-related services and interfaces, we refer to the World-Wide Web Consortium (W3C) unless explicitly superseded in this Appendix.

A.1 Operating Systems (OS)

All conventional operating systems may be employed. If there is an exceptional consideration or doubt about a particular OS or version, it should be submitted as a review item for explicit consideration by the TWG.

A.2 Metadata Schema

There are many important metadata standards that bear consideration. Examples include:

- 1. Ecological Metadata Language ([22], [31]),
- 2. Open Geospatial Metadata
- 3. FGDC

For that reason, the metadata schema recommended here is referred to as a multilateral metadata convention that supports the production of arbitrary metadata files to support compliance with the current and future complement of metadata interfaces: it is designed to be independent of any particular standard but compatible with most.

The metadata schema in Appendix B, Listing 1 is the default schema. It may be augmented. It is intended to be the basis of all metadata interoperability schemas derived from it for integration purposes as required to integrate with other systems and catalogues. This metadata schema is dependent on the controlled vocabulary standard described in section A.3.

A.3 Controlled Vocabulary

Controlled vocabulary is referenced to existing community standards where they exist. The OWIA will conform to the community standards listed here and these will be superseded by the OWIA Controlled Vocabulary when there is a conflict.

- (1) CF Conventions and Metadata: Standard Names,
- (2) World Meteorological Organization Practices,
- (3) Open Geospatial Consortium WaterML 2.0,
- (4) OWIA Standard Names (TBD).

A.4 Georeferencing

Most data within the OWIA federation will require georeferencing. The default reference for definitions of map projections, use of datums and related geospatial standards and conventions will be [40]. Snyder and EPSG, PROJ.4, GDAL.

A.5 Intellectual Property Rights

Data published by the OWIA federation is governed by at least one of the following licensing mechanisms: (1) GNU Public License

(2) Creative Commons

A.6 Trusted Archives

Trusted archives are digital object repositories where data published within the OWIA community can be reliably found. These may not be the only locations but they are considered to be the primary authoritative sources of copies of digital objects. The criteria used to determine trusted archive status are those of the

- (1) USGS Acceptable Digital Repositories for USGS Scientific Publications and Data,
- (2) CoreTrustSeal,
- (3) DIN 31644 Information and documentation Criteria for trustworthy digital archives, and
- (4) ISO 16363:2012 Space data and information transfer systems Audit and certification of trustworthy digital repositories.

OWIA trusted archives include:

- (1) University of California (in discussion),
- (2) California Department of Water Resources (TBD).

A.7 Digital Objects

Digital objects are anything that can be stored and retrieved from within the file system of an operating system. Streaming data presented to a display device are not considered to be digital objects since the data contained in the stream is not stored and cannot be directly used in reproducible analyses or unambiguously referred to or re-used. A suggested approach to employing data of this type is to checkpoint the stream into a file which can be used as a stored digital object.

A.8 File Formats and Data Encodings

Recommended file formats and data encodings are summarized in Table 2. The default standard for character encodings is UTF-8 [25] with extended ASCII as a secondary alternative.

OWIA Class	File Type	Structure	Encoding	Georeferencing	Controlled Vocabulary	Interoperability Tools
Level 1	Comma-separated Values (*.csv)	Record-oriented, Scalar (Integer, Float), Text	ASCII, UTF-8	EPSG [21], DAU-County, HUC10/12 [46]	OWIA, CF [3]	Any
	Spread-sheet (*.ods [45])	Record-oriented, Scalar (Integer, Float), Text	ASCII, UTF-8	EPSG	OWIA, CF	Any
	Geospatial (GDAL-supported)	Raster	Binary	EPSG	OWIA, CF	GDAL
		Vector	Any	EPSG	OWIA, CF	GDAL [8], ogr2ogr, QGIS [37], GRASS [10], GMT [47]
	NetCDF [44]	Multi-dimensional, self-documenting	Binary	EPSG [21], DAU-County, HUC10/12	OWIA, CF	NetCDF API, NCL, NCO [48], GMT
	Text-processing	Rich Text Format (rtf), free-text	Binary	N/A	OWIA, CF	OpenOffice, LibreOffice, rtf2latex, rtf2html
		TeX [42]	ASCII, UTF-8, human-readable	N/A	OWIA, CF	latex2rtf, latex2html, tex4ht
Level 0	Microsoft Excel Spreadsheets (*.xlsx, *.xls)	Cell, Worksheet	Binary	No	No	OpenOffice [34], LibreOffice [29]
	Microsoft Word Documents (*.docx, *.doc)	Free-text	Binary	No	No	OpenOffice, LibreOffice
	DBMS Export	Human-readable	ASCII, UTF-8 (*.txt)	No	No	None
	ESRI Geodatabase	Proprietary	Binary	No	No	QGIS (GDAL-enabled)
	NetCDF	Multi-dimensional, self-documenting	Binary	Any	Any	NetCDF API, NCL, NCO, GMT
Inadmissib	le DBMS	Database Structure and Schema	Any	Any	Any	None
	Proprietary w/o Interoperability Tools	Proprietary	Any	Any	Any	None

OWIA System Requirements Document

A.9 Cross-referencing Services

Cross-referencing services are used to support global searching for digital objects published using the OWIA standards and conventions. The default system is the University of California's EZID service.

A.10 Commercial Search Services

Commercial search services using the WWW are typified by Google.

Appendix B Metadata Schema

Listing 1: The current metadata schema.

```
# Canonical Collection
 OWIA_CanonicalCollection_ArchivistEmail,"VARCHAR(50000)","Required","1","OWIA","Manager Email"
 OWIA_CanonicalCollection_ArchivistInstitution, "VARCHAR(50000)", "Required", "1", "OWIA", "Manager Institution"
OWIA_CanonicalCollection_ArchivistInstitution, VARCHAR(50000), Required, T, OWIA, Manager Institution
OWIA_CanonicalCollection_ArchivistName, "VARCHAR(50000)", "Required", "1", "OWIA", "Manager Name"
OWIA_CanonicalCollection_ArchivistPhone, "VARCHAR(50000)", "Required", "1", "OWIA", "Manager Phone"
OWIA_CanonicalCollection_CollectionIdentifier, "VARCHAR(50000)", "Required", "1", "OWIA", "Collection Identifier"
OWIA_CanonicalCollection_ControlledVocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "Controlled Vocabulary"
OWIA_CanonicalCollection_ControlledVocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "Controlled Vocabulary"
OWIA_CanonicalCollection_Creator, "VARCHAR(50000)", "Required", "1", "OWIA", "Controlled Vocabulary"
 OWIA_CanonicalCollection_Description,"VARCHAR(50000)"," Required","1","OWIA"," Thorough Description of collection"
OWIA_CanonicalCollection_Description, VARCHAR(50000)", Required ", "1", "OWIA", "Infordign Description of OWIA_CanonicalCollection_MTFVersion, "VARCHAR(50000)"," Required ", "1", "OWIA", "MTFVersion"
OWIA_CanonicalCollection_Ontology, "VARCHAR(50000)"," Required ", "1", "OWIA", "OWIA"," Owtoology
OWIA_CanonicalCollection_Ontology, "VARCHAR(50000)"," Required ", "1", "OWIA"," Owtoology"
OWIA_CanonicalCollection_Publisher, "VARCHAR(50000)"," Required ", "1", "OWIA"," Publisher of collection"
 OWIA_CanonicalCollection_Subject,"VARCHAR(50000)"," Required ","1","OWIA"," General Subject area of collection"
 OWIA_CanonicalCollection_Title, "VARCHAR(50000)", "Required", "1", "OWIA", "Specific Title of collection"
 # Canonical ADO
 OWIA_CanonicalADO_ADOIdentifier, "VARCHAR(50000)", "Required", "1", "OWIA", "ADO Identifier"
OWIA_CanonicalADO_ADOVersion, "VARCHAR(50000)", "Required", "1", "OWIA", "ADO Version"
OWIA_CanonicalADO_AccessControl, "VARCHAR(50000)", "Required", "1", "OWIA", "Access Control for this object"
OWIA_CanonicalADO_Author, "VARCHAR(50000)", "Required", "1", "OWIA", "Author"
 OWIA_CanonicalADO_BlockTypes, "VARCHAR(50000)", "Required", "1", "OWIA", "Major data block types"
 OWIA_CanonicalADO_Children, "VARCHAR(50000)", "Required", "1", "OWIA", "Children"
OWIA_CanonicalADO_Children, "VARCHAR(50000)", "Required", "1", "OWIA", "Children"
OWIA_CanonicalADO_CollectionIdentifier, "VARCHAR(50000)", "Required", "1", "OWIA", "Collection Identifier"
OWIA_CanonicalADO_ContentFilenames, "VARCHAR(50000)", "Required", "1", "OWIA", "Content Filenames"
OWIA_CanonicalADO_Contributor, "VARCHAR(50000)", "Required", "1", "OWIA", "Contributor of this upload"
OWIA_CanonicalADO_Contributor, "VARCHAR(50000)", "Required", "1", "OWIA", "Controlled Vocabulary"
OWIA_CanonicalADO_ControlledVocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "Controlled Vocabulary"
OWIA_CanonicalADO_Coverage, "VARCHAR(50000)", "Required", "1", "OWIA", "Coverage min max lat lon"
OWIA_CanonicalADO_Creator, "VARCHAR(50000)", "Required", "1", "OWIA", "Original creator of object"
OWIA_CanonicalADO_DOI, "VARCHAR(50000)", "Required", "1", "OWIA", "Digital Object Identifier"
OWIA_CanonicalADO_DI, "VARCHAR(50000)", "Required", "1", "OWIA", "Digital Object Identifier"
OWIA_CanonicalADO_DOI, "VARCHAR(50000)", "Required", "1", "OWIA", "Digital object Identifier"
OWIA_CanonicalADO_DOI, "VARCHAR(50000)", "Required", "1", "OWIA", "Digital object Identifier"
OWIA_CanonicalADO_DOI, "VARCHAR(50000)", "Required", "1", "OWIA", "Digital object Identifier"
 OWIA_CanonicalADO_Description, "VARCHAR(50000)", "Required", "1", "OWIA", "Description including importance"
OWIA_CanonicalADO_Description, "VARCHAR(50000)"," Required", "1", "OWIA", "Description in
OWIA_CanonicalADO_ExpertLevel, "VARCHAR(50000)"," Required", "1", "OWIA", "Description in
OWIA_CanonicalADO_Filesize, "NUMERIC", "Required", "1", "OWIA", "Filesize"
OWIA_CanonicalADO_Format, "VARCHAR(50000)"," Required", "1", "OWIA", "Format MIME type"
OWIA_CanonicalADO_Keywords, "VARCHAR(50000)"," Required", "1", "OWIA", "Keywords"
OWIA_CanonicalADO_Keywords, "VARCHAR(50000)"," Required", "1", "OWIA", "Language"
OWIA_CanonicalADO_Language, "VARCHAR(50000)"," Required", "1", "OWIA", "Language"
OWIA-CanonicalADO-LatitudeEnd, "REAL", "Required", "1", "OWIA", "Latitude End"
OWIA-CanonicalADO-LatitudeNorth, "REAL", "Required", "1", "OWIA", "Latitude End"
OWIA-CanonicalADO-LatitudeNorth, "REAL", "Required", "1", "OWIA", "Latitude Northernmost"
OWIA-CanonicalADO-LatitudeSouth, "REAL", "Required", "1", "OWIA", "Latitude Southernmost"
 OWIA_CanonicalADO_LatitudeStart, "REAL", "Required", "1", "OWIA", "Latitude southernmost"
OWIA_CanonicalADO_LatitudeStart, KEAL, "Required", "1", "OWIA", "Latitude at Start of object"
OWIA_CanonicalADO_LongitudeEast, "REAL", "Required", "1", "OWIA", "Longitude Easternmost"
OWIA_CanonicalADO_LongitudeEnd, "REAL", "Required", "1", "OWIA", "Longitude at End of object"
OWIA_CanonicalADO_LongitudeStart, "REAL", "Required", "1", "OWIA", "Longitude at Start of object"
OWIA_CanonicalADO_LongitudeWest, "REAL", "Required", "1", "OWIA", "Longitude at Start of object"
OWIA_CanonicalADO_LongitudeWest, "REAL", "Required", "1", "OWIA", "Longitude Westernmost"
OWIA_CanonicalADO_LongitudeWest, "REAL", "Required", "1", "OWIA", "Longitude Westernmost"
OWIA_CanonicalADO_MD5SUM, "VARCHAR(50000)", "Required", "1", "OWIA", "Verifier (md5sum)"
 OWIA_CanonicalADO_MIFVersion, "VARCHAR(50000)"," Required", "1", "OWIA"," MetaData Content Version"
```

OWIA_CanonicalADO_MTFVersion,"VARCHAR(50000)"," Required","1","OWIA"," MetaData Template File Version" OWIA_CanonicalADO_ONTOVERSION, VARCHAR(50000), "Required", "1", "OWIA", "MetaData Template The Version OWIA_CanonicalADO_Parent, "VARCHAR(50000)", "Required", "1", "OWIA", "Ontology" OWIA_CanonicalADO_PhysicalStorageLocation, "VARCHAR(50000)", "Required", "1", "OWIA", "Publisher" OWIA_CanonicalADO_PhysicalStorageLocation, "VARCHAR(50000)", "Required", "1", "OWIA", "Publisher" OWIA_CanonicalADO_Relation, "VARCHAR(50000)", "Required", "1", "OWIA", "Relation" OWIA-CanonicalADO-Relation, VARCHAR(50000)", Required ", 1", OWIA", Relation OWIA-CanonicalADO-Siblings, "VARCHAR(50000)", "Required ", "1", "OWIA", "Link to rights statement" OWIA-CanonicalADO-Siblings, "VARCHAR(50000)", "Required ", "1", "OWIA", "Siblings" OWIA-CanonicalADO-Source, "VARCHAR(50000)", "Required ", "1", "OWIA", "Source of object for Dublin Core" OWIA-CanonicalADO-SourceFileName, "VARCHAR(50000)", "Required ", "1", "OWIA", "Source of object for Dublin Core" OWIA_CanonicalADO_SourceFileName, "VARCHAR(50000)", "Required", "1", "OWIA", "Source File Name" OWIA_CanonicalADO_Subject, "VARCHAR(50000)", "Required", "1", "OWIA", "Subject area of object" OWIA_CanonicalADO_TimeEnd, "DATE", "Required", "1", "OWIA", "End Date Time of object" OWIA_CanonicalADO_TimeStart, "DATE", "Required", "1", "OWIA", "Start Date Time of object" OWIA_CanonicalADO_TimeStart, "DATE", "Required", "1", "OWIA", "Start Date Time of object" OWIA_CanonicalADO_Title, "VARCHAR(50000)", "Required", "1", "OWIA", "Title to identify object in specific detail" OWIA_CanonicalADO_Type, "VARCHAR(50000)", "Required", "1", "OWIA", "Type of Dublin Core resource" OWIA_CanonicalADO_URL, "VARCHAR(50000)", "Required", "1", "OWIA", "Universal Resource Locator" * ********************** Documentation OWIA_Documentation_MTFVersion, "VARCHAR(50000)", "Required", "1", "OWIA", "MTFVersion" OWIA_Documentation_ADOIdentifier, "VARCHAR(50000)", "Required", "1", "OWIA", "ADOIdentifier" OWIA_Documentation_ControlledVocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "ADOIdentifier" OWIA_Documentation_ControlledVocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "ControlledVocabulary" OWIA_Documentation_Ontology, "VARCHAR(50000)", "Required", "1", "OWIA", "Ontology" OWIA_Documentation_Description, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Document Description or Title" OWIA_Documentation_Format, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Format" OWIA_Documentation_ObjectType, "VARCHAR(50000)", "Required", "1", "OWIA", "Data Object Type" OWIA_Documentation_Type, "VARCHAR(50000)"," Arbitrary ","1","OWIA"," Document Type' # Products OWIA_Products_MTFVersion, "VARCHAR(50000)", "Required", "1", "OWIA", "MTFVersion" OWIA_Products_ADOIdentifier, "VARCHAR(50000)", "Required", "1", "OWIA", "ADOIdentifier" OWIA_Products_ControlledVocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "ControlledVocabulary" OWIA_Products_Ontology, "VARCHAR(50000)", "Required", "1", "OWIA", "Ontology" OWIA_Products_Description,"VARCHAR(50000)","Arbitrary","1","OWIA","Data Product Description" OWIA_Products_Description, "Archar(50000)", "Arbitrary", "1", "OWIA", "Data Product Description OWIA_Products_Format, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Data Production Method" OWIA_Products_ObjectType, "VARCHAR(50000)", "Required", "1", "OWIA", "Data Object Type" OWIA_Products_Type, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Data Product Type" * **** # CKAN per Greg Smith 2017-11-01 OWIA_CKAN_MTFVersion, "VARCHAR(50000)", "Required", "1", "OWIA", "MTFVersion" OWIA_CKAN_ADOIdentifier, "VARCHAR(50000)", "Required", "1", "OWIA", "ADOIdentifier" OWIA_CKAN_ADOIdentifier, VARCHAR(50000)", Required ", 1", "OWIA", "ADOIdentifier" OWIA_CKAN_Controlled Vocabulary, "VARCHAR(50000)", "Required", "1", "OWIA", "Controlled Vocabulary" OWIA_CKAN_Ontology, "VARCHAR(50000)", "Required", "1", "OWIA", "Ontology" OWIA_CKAN_Description, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Data Product Description" OWIA_CKAN_Format, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Format" OWIA_CKAN_Method, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Data Production Method" OWIA_CKAN_ObjectType, "VARCHAR(50000)", "Required", "1", "OWIA", "Data Object Type" OWIA_CKAN_Type, "VARCHAR(50000)", "Arbitrary", "1", "OWIA", "Data Product Type" OWIA_CKAN_title, "VARCHAR(50000)", "Required", "OWIA", "Title" OWIA_CKAN_description, "VARCHAR(50000)", "Required", "OWIA", "Description" OWIA_CKAN_keyword, "VARCHAR(50000)", "Required", "OWIA", "Tags" OWIA_CKAN_modified, "VARCHAR(50000)", "Required", "OWIA", "Last Update" OWIA_CKAN_publisher, "VARCHAR(50000)", "Required", "OWIA", "Publisher" OWIA_CKAN_publisher, VARCHAR(50000), Required , OWIA , Publisher OWIA_CKAN_contactPoint, "VARCHAR(50000)", "Required", "OWIA", "Contact Name and Email" OWIA_CKAN_identifier, "VARCHAR(50000)", "NA", "OWIA", "Unique Identifier" OWIA_CKAN_accessLevel, "VARCHAR(50000)", "Required", "OWIA", "Public Access Level" OWIA_CKAN_bureauCodeUSG, "VARCHAR(50000)", "NA", "OWIA", "Bureau Code" OWIA_CKAN_programCodeUSG, "VARCHAR(50000)", "NA", "OWIA", "Bureau Code" OWIA_CKAN_programCodeUSG, "VARCHAR(50000)", "NA", "OWIA", "Program Code" OWIA_CKAN_license, "VARCHAR(50000)", "NA", "OWIA", "License" OWIA_CKAN_rights, "VARCHAR(50000)", "NA", "OWIA", "Rights" OWIA_CKAN_spatial, "VARCHAR(50000)", "Required", "OWIA", "Spatial" OWIA_CKAN_temporal, "VARCHAR(50000)", "Required", "OWIA", "Temporal" OWIA_CKAN_distribution ,"VARCHAR(50000)","NA","OWIA"," Distribution" OWIA_CKAN_@type,"VARCHAR(50000)","NA","OWIA"," Metadata Type" OWIA_CKAN_accrualPeriodicity, "VARCHAR(50000)", "NA", "OWIA", "Frequency"

OWIA_CKAN_conformsTo, "VARCHAR(50000)", "NA", "OWIA", "Data Standard"
OWIA_CKAN_dataQualityUSG, "VARCHAR(50000)", "NA", "OWIA", "Data Quality"
OWIA_CKAN_describedBy, "VARCHAR(50000)", "NA", "OWIA", "Data Dictionary"
OWIA_CKAN_describedByType, "VARCHAR(50000)", "NA", "OWIA", "Data Dictionary Type"
OWIA_CKAN_ispartOf, "VARCHAR(50000)", "NA", "OWIA", "Collection"
OWIA_CKAN_issued, "VARCHAR(50000)", "NA", "OWIA", "Release Date"
OWIA_CKAN_language, "VARCHAR(50000)", "NA", "OWIA", "Language"
OWIA_CKAN_landingPage, "VARCHAR(50000)", "NA", "OWIA", "Homepage URL"
OWIA_CKAN_primaryITInvestmentUIIUSG, "VARCHAR(50000)", "NA", "OWIA", "Related Documents"
OWIA_CKAN_systemOfRecordsUSG, "VARCHAR(50000)", "NA", "OWIA", "Category"

Appendix C Support for AB1755

The OWIA provides complete support for the open-data and transparency requirement of the AB1755 legislation. Table 4 summarizes the relationship between the functional requirments and the objectives stated in the AB1755 bill.

Identifier	Name	Data Sharing	Documentation	Quality Control	Public Access	Open-source platforms and decision support tools
FR-100-100	Data Acquisition	Х				
FR-100-110	*-Manual-	Х				
FR-100-120	*-Automated-	Х				
FR-200-100	Quality Control-*-	Х		Х		
FR-200-110	*-Verification-	Х		Х		
FR-200-120	*-*-Documentation	Х	Х	Х		
FR-200-130	*-*-Reproducibility	Х		Х		
FR-200-140	*-*-Data Traceability	Х		Х		
FR-200-150	*-Standardization-	Х	Х	Х		Х
FR-200-160	*-*-File-naming Conventions	Х	Х	Х		Х
FR-200-170	*-Interoperable Transformation-	Х		Х		Х
FR-200-180	*-*-Separation of Data and Computation	Х		Х		Х
FR-200-190	*-*-Data Interoperability	Х	Х	Х		Х
FR-200-200	*-*-Products or Resources	Х		Х		Х
FR-300-100	Publication-*-	Х	Х		Х	
FR-300-110	*-Cross-Referencing-Service-	Х	Х		Х	
FR-300-120	*-*-Assignment of Digital Object Identifiers	Х	Х		Х	
FR-300-130	*-Packaging-	Х			Х	
FR-300-140	*-*-Compression Methods	Х			Х	
FR-300-150	*-*-Archive File Formatting	Х			Х	
FR-300-160	*-Archival-	Х			Х	
FR-300-170	*-*-Open Access Distribution	Х			Х	
FR-400-100	Data Traceability-*-	Х	Х		Х	
FR-400-110	*-Metadata Production-	Х	Х		Х	
FR-400-120	*-Intellectual Property Rights Management-	Х	Х		Х	
FR-400-130	*-Public Law Compliance-	Х	Х		Х	
FR-400-140	*-Licensing-	Х	Х		Х	
FR-400-150	*-Liability-	Х	Х		Х	
FR-400-160	*-Searching-	Х			Х	
FR-400-170	*-*-Cross-referencing System Integration	Х			Х	
FR-400-180	*-*-Search Engine Optimization	Х			Х	
FR-400-190	*-Version Control-	Х	Х			
FR-400-200	*-*-Binary Data	Х	Х			
FR-400-210	*-*-Non-Binary Data	Х	Х			
FR-400-220	*-Anomaly Reporting-	Х	Х			
FR-500-100	System Portability-*-					Х
FR-500-110	*-Backup and Restore-					Х
FR-500-120	*-Platform Portability-					X
FR-600-100	External Interfaces-*-	Х				Х
FR-600-110	*-Data and Metadata Acquisition-	Х				Х
FR-600-120	*-Data and Metadata Distribution-	Х				Х

Table 3: Traceability of AB1755 objectives (columns) to OWIA SOPs by use case (rows).

OWIA System Requirements Document

Appendix D Traceability Tables

The traceability tables for *stakeholder objectives to functional requirements* and functional requirements to technical requirements are listed below in Tables 4 and **??**, respectively. These tables are provided to assist in the evaluation of change proposals and design approaches in order to understand more conveniently how any proposed change may ripple through the OWIA in unintended way and to provide a sound basis for engineering analysis of the interdependencies of the requirements both functional and technical as they bear upon project management and design decisions.

Table 4: Traceability Table: Objective O-1100-1000 to Functional Requirements. This is an example of what subordinate Technical Requirements might resolve to and is meant only to characterize what *Resolution* of Functional Requirements might look like in a

ED 100 100		
FR-100-100	Data Acquisition-*-NULL	
FR-100-110	*-Manual-NULL	Level 0: H1 P scraping (cf. Table ?? for UC001)
FR-100-120	*-Automated-NULL	Level 0: Stored procedures for updating
FR-200-100	Quality Control-*-NULL	
FR-200-110	*-Verification-NULL	Stored programs and transformation of Level 0 sources to OWIA standards, Compute checksums and version control a list of the checksums.
FR-200-120	*-*-Documentation	OWIA Standard Formats
FR-200-130	*-*-Reproducibility	Stored procedures and input data with descriptive metadata.
FR-200-140	*-*-Data Traceability	OWIA Level 0 metadata generation, OWIA standard Level 0 processing
FR-200-150	*-Standardization-NULL	OWIA Level 0 standard processing (verification of contents, anomaly detection, missing value coding)
FR-200-160	*-*-File-naming Conventions	OWIA Level 0 naming convention
FR-200-160	*-*-File-naming Conventions	Level 0 verification of data access and reproduction of quality control and standardization
FR-200-170	*-Interoperable Transformation-NULL	Level 0 metadata verification
FR-200-180	*-*-Separation of Data and Computation	
FR-200-190	*-*-Data Interoperability	EZID (External Interface)
FR-200-200	*-*-Products or Resources	(1) The water manager must identify potential source(s) of water, and for each determine the quantity and timing of water available for recharge and its cost. (2) To determine where the project should be located, the water manager must examine different options based on basin capacity and suitability of recharge areas; parcel data indicating available land and land values; and water quality implications based on current or past land use and the design of the project. (3) To determine the best method for recharge, basin characteristics such as subsurface characteristics, soil types, topography, current and planned land use, and basin capacity must be taken into account.
FP 300 100	Publication * NULL	Lavel On Matedata Production
FR-300-100	* Cross Deferencing Service NULL	Ever of Metadata Frouenon
FR-300-110 FR 300 120	* * Assignment of Digital Object Identifiers	Identifier Assignment (a grant) via External Interfore
FR-300-120	* Packaging NULL	ABITS
FR-300-130	* * Compression Methods	Losslass
FP 300 150	* * Archive File Formatting	Lussics
FR-300-150	*- Archival-NULL	a_{12} , a_{12}
FR-300-170	*_*_Open Access Distribution	In bits revue see a fin a dusted active for access and derivery using OWTA-compliant external interfaces.
FR-400-100	Data Traceability.*-NULL	Via DOI for parents and siblings
FR-400-110	*-Metadata Production-NULL	
FR-400-120	*-Intellectual Property Rights	Attribution 4.0 International (CC BY 4.0). Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
1 K-400-120	Management_NULI	Autoduon 4.0 incluational (CC D I 4.0), Autoduon VonCommercial 4.0 incluational (CC D I 4.0)
FR-400-130	*-Public Law Compliance-NULI	AR1755
FR-400-140	*-Licensing-NULL	Compute checksums and version control a list of the checksums
FR-400-150	*-Liability-NULL	OWIA standard version control system
FR 400-150	* Searching NULL	OWIA-statical version control system
FR-400-170	*_*_Cross_referencing System Integration	Crossee DataCite
FR_400_180	*_*-Search Engine Ontimization	Goode bot
FR_400-100	*-Version Control-NULI	Oren. Source systems verified on Linux. Windows: OSX
110-170		Open source systems vermed on Emux, windows, OSA

Resolution

Label

FR-400-200	*-*-Binary Data	Naming convention.
FR-400-210	*-*-Non-Binary Data	ASCII-based version control systems (e.g., git, svn, mercurial)
FR-400-220	*-Anomaly Reporting-NULL	Curatorial email address
FR-500-100	System Portability-*-NULL	Open-source operation on major operating systems.
FR-500-110	*-Backup and Restore-NULL	Rsync-based
FR-500-120	*-Platform Portability-NULL	Demonstrated operation across major platforms: Linux, OSX, Windows
FR-600-100	External Interfaces-*-NULL	Uniquely identified per the Interface Control Appendix.
FR-600-110	*-Data and Metadata Acquisition-NULL	Compliant with OWIA standards and conventions
FR-600-120	*-Data and Metadata Distribution-NULL	Compliant with OWIA standards and conventions

Glossary

Algorithm [28] The modern meaning for algorithm is quite similar to that of recipe, process, method, technique, procedure, routine, rigamorole, except that the word "algorithm" connotes something just a little different. Besides merely being a finite set of rules that gives a sequence of operations for solving a specific type of problem, an algorithm has five important features: 1. Finiteness, 2. Definiteness, 3. Input, 4. Output, 5. Effectiveness 22

algorithms Algorithm 3

API An API (Application Programming Interface) is a set of programming library calls and supporting compile and run-time libraries. These exist on both the client-side and server-side of a computer application although they are usually asymmetrical in terms of what the server implements versus what the client implements. The purpose for providing APIs is to standardize and simplify the programming required to add functionality to a software application and to enhance the portability and interoperability of software across both platforms and data Application Programming Interface. 22

Application Programming Interface see API 22

federated See Federation 1

federation A federation is a group of data providers and users using jointly agreed-upon standards of operation in a collective fashion to ensure the interoperability of the resources they collectively hold and employ. The term may be used, for example, when describing the interoperation of distinct cyberinfrastructure networks with different internal structures. The term may also be used when human groups agree to collectively manage cyberinfrastructure development and operation using commonly held, and managed, requirements, standards and conventions (e.g., metadata and APIs), and operating procedures to ensure the interoperability of distinct cyberinfrastructure resources (cf. Wikipedia Definition). 1, 22

Federation See federation 22

- **interoperability** The ability of diverse computer systems or software to exchange and make use of common input data. 22
- procedures An established or official way of doing something (Oxford English Dictionary). 1, 3, 22
- **protocol** Protocols are methods of implementing a set of objectives and requirements in a systematic way. In computing, protocols mean both specific implementations of methods such as HTTP [26] and FTP [23] and, more generally as described by the *Internet Engineering Task Force*, protocols are sequences of processing steps that are also referred to as procedures [24]. 7

specification cf. Specification 22

Specification Specification is the articulation of a one or more system requirements. It may also refer to a document (i.e., a 'spec') containing a set of requirements specifications. In the project management context, specification pertains to the expression of one or more requirements used to minize cost, schedule, technical and operational risk on a contract. 22

standards cf. Standards 3, 22

Standards Standards are testable criteria used to constrain system designs in the interest of reducing cost, schedule, technical and operational risk. As described in [33], *standards provide a proven basis for establishing common technical requirements across a program or project to avoid incompatibilities and ensure that at least minimum requirements are met. Common standards can also lower implementation cost as well as costs for inspection, common supplies, etc. Typically, standards (and <i>specifications*) are used throughout a product life cycle to establish design requirements and margins, materials and process specications, test methods, and interface specifications. Standards are used as requirements (and guidelines) for design, fabrication, verification, valdidation, acceptance, operations, and maintenance. Consider also the International Standards Organization and IETF as examples of standards organizations. 22

References

- [1] California Legislature, *AB-1755 The Open and Transparent Water Data Act. (2015-2016)*, Assembly Bill 1755, California Legislature, 2015.
- [2] Dan Cayan and John Helly, *The Wireless Watershed at the Santa Margarita Ecological Reserve*, Southwest Hydrology **September/October** (2003).
- [3] CF Governance Conventions Committee, *CF Conventions and Metadata*.
- [4] B. Chadwick, P.F. P.F. Wang, J. Helly, M. Brand, R. Flick, A. Young, W. O'Reilly, P. Bromirski, W. Crampton, R. Guza, J. Helly, T. Nishikawa, S. Boyce, M. Landon, M. Martinez, I. Canner, and B. Leslie, A Methodology for Assessing the Impact of Sea Level Rise on Representative Military Installations in the Southwestern United States, Technical Report RC-1703, SPAWAR Systems Center Pacific, 2014.
- [5] CODATA-ICSTI Task Group on Data Citation Standards and Practices, Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data, Data Science Journal 12 (2013), 1–75.
- [6] B. W. Eakins, S.P. Miller, J. Helly, and B. Zelt, *The fully electronic IODP Site Survey Data Bank*, Scientific Drilling **2** (2006), 40–42.
- [7] Eugene W. Rice and Roger B. Baird and Andrew D. Eaton and Lenore S. Clesceri (ed.), *Standard Methods for the Examination of Water and Wastewater*, vol. 22nd Edition, American Public Health Association, American Water Works Association, Water Environment Federation, 2012.
- [8] Geospatial Data Abstraction Library, Geospatial Data Abstraction Library.
- [9] Steven Glaser, Martha Conklin, and Roger Bales, *Building an Intelligent Water Information System American River Prototype*, Technical report, CITRIS, 2013.
- [10] GRASS Development Team, *Geographic Resources Analysis Support System (GRASS GIS) Software*, Open Source Geospatial Foundation, 2012.
- [11] Kay Gross, Edie Allen, Caroline Bledsoe, Robert Colwell, Paul Dayton, Megan Dethier, John Helly, Robert Holt, Nancy Morin, William Michener, Steward T. Pickett, and Susan Stafford, *Report of the Committee on the Future of Long-term Ecological Data (FLED)*, Tech. report, Ecological Society of America, Washington, D. C., 1995.
- [12] J. Helly, T. T. Elvins, D. Sutton, D. Martinez, S Miller, S. Pickett, and A. M. Ellison, *Controlled Publication of Digital Scientific Data*, CACM 45 (2002), no. 5, 97–101.
- [13] J. J. Helly, *Visualization of ecological and environmental data*, LTER Network Office, University of New Mexico, Albuquerque, New Mexico, 1998.
- [14] John Helly, New concepts of publication, Nature 393 (1998).
- [15] _____, *Digital libraries in hydrology*, Hydroinformatics: Data Integrative Approaches in Computation, Analysis, and Modeling (Praveen Kumar, Mike Folk, Momcilo Markus, Jay C Alameda, and Peter Bajcsy, eds.), Taylor and Francis, London, England, 2005, p. 552.
- [16] John Helly and Maggi Kelly, Collaborative Management of Natural Resources in San Diego Bay, Coastal Management 29 (2001), 16.

- [17] John Helly, Hubert Staudigel, and Anthony Koppers, *Scalable models of data sharing in Earth sciences*, Geochemistry Geophysics Geosystems **4** (2003).
- [18] John J. Helly, *Cyberinfrastructure for data authorship, publication and application interoperability*, no. 1501928, AGU, AGU, December 2012.
- [19] _____, Systems engineering for scalable cyberinfrastructure: Data, software, systems, DRAFT, 2017.
- [20] John J. Helly and Kevin T. Herbinson, Visualization of the salinity plume from a coastal ocean water desalination plant, Water Environment Research 66 (1994), no. 5, 753–758.
- [21] International Association of Oil & Gas Producers, International Association of Oil & Gas Producers.
- [22] International Standards Organization (ISO), *EML Ecological Metadata Language*, http://www.dcc.ac.uk/resources/metadata-standards/ eml-ecological-metadata-language (2017).
- [23] Internet Engineering Task Force (IETF), The File Transfer Protocol, RFC 354, IETF, 1972.
- [24] _____, The Internet Standards Process Revision 3, Tech. Report RFC 2026, IETF, October 1996.
- [25] _____, UTF-8, a transformation format of ISO 10646, Internet standard, IETF, November 2003.
- [26] _____, Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing, RFC 7230 ISSN: 2070-1721, IETF, June 2014.
- [27] ISTQB EXAM CERTIFICATION, What is Agile model advantages, disadvantages and when to use it?, http://istqbexamcertification.com/what-is-agile-model-advantages-disadvantages-and-when-to-use-it (2018).
- [28] Donald Ervin Knuth, *The Art of Computer Programming*, Fundamental Algorithms, Addison-Wesley Series in Computer Science and Information Processing, vol. 1, Addison-Wesley, 1981.
- [29] LibreOffice, *LibreOffice*.
- [30] J. Gage Marchini, Connecting the "Drops" of California Water Data: Chapter 506: The Open and Transparent Water Data Act, The University of the Pacific Law Review (2017), 785–799.
- [31] William K. Michener, James W. Brunt, John J. Helly, Thomas B. Kirchner, and Susan G. Stafford, *Nongeospatial metadata for the ecological sciences*, Ecological Applications 7 (1997), no. 1, 330– 242.
- [32] NASA, NASA Systems Engineering Handbook, Tech. Report NASA-SP-2007-6105-Rev-1-Final-31Dec2007, NASA, 2007.
- [33] NASA-2016, NASA Systems Engineering Handbook, NASA/SP-2016-6105 Rev2 supersedes SP-2007-6105 Rev 1 dated December, 2007. NASA/SP-2016-6105 REV 2, NASA, 2016.
- [34] OpenOffice, *OpenOffice*.
- [35] Paul F. Uhlir Rapprteur, For Attribution Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop, Report ISBN 978-0-309-26728-1, National Acadeies Press, 2012.

January 24, 2018

- [36] J. Pundsack, R. Belland, D. Broderson, G.C. Fox, J. Dozier, J. Helly, W. Li, P. Morin, M. Parsons, A. Roberts, C. Tweedie, and C. Yang., *Report on Workshop on Cyberinfrastructure for Polar Sciences. St. Paul, Minnesota. University of Minnesota Polar Geospatial Center.*, Report, National Science Foundation, 2013.
- [37] Quantum GIS Development Team, *Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.*
- [38] F. M. Ralph, K. A. Prather, D. Cayan, J. R. Spackman, P. DeMott, M. Dettinger, C. Fairall, R. Leung, D. Rosenfeld, S. Rutledge, D. Waliser, A. B. White, J. Cordeira, A. Martin, J. Helly, and J. Intrieri, *CalWater Field Studies Designed to Quantify the Roles of Atmospheric Rivers and Aerosols in Modulating U.S. West Coast Precipitation in a Changing Climate.*, Bulletin of the American Meteorological Society (2015).
- [39] SmartH20, Standards and open data report, Tech. Report smarth20.D82.SUPSI.WP8.V1.1, Coordinator: SUPSI Contractors: POLIMI, UoM, SETMOB, EIPCM, TWUL, SES, MOONSUB, July 31 2015.
- [40] John P. Snyder, *Map Projections: A Working Manual*, USGS Numbered Series: Professional Paper (1987), no. 1395, ix, 385 p.; Plate: 36.82 x 29.15 inches.
- [41] Hubert Staudigel, John Helly, Anthony A. P. Koppers, Henry F. Shaw, William F. McDonough, Albrecht W. Hofmann, Charles H. Langmuir, Kerstin Lehnert, Baerbel Sarbas, Louis A. Derry, and Alan Zindler, *Electronic data publication in geochemistry*, Geochem. Geophys. Geosyst. 4 (2003), no. 3.
- [42] TeX Users Group, TeX Users Group.
- [43] The Aspen Institute, Internet Of Water: Sharing And Integrating Water Data For Sustainability: A Report From The Aspen Institute Dialogue Series On Water Data, https://assets.aspeninstitute.org/content/uploads/2017/05/Internet-of-Water-Report-May-2017.pdf ISBN: 0-89843-665-6, The Aspen Institute, One Dupont Circle, NW Suite 700, Washington, DC 20036, 2017.
- [44] Unidata Community Programs, NetCDF.
- [45] United States Library of Congress, *Sustainability of Digital Formats: Planning for Library of Congress Collections*.
- [46] USGS, Hydrologic Unit Maps.
- [47] P. Wessel and W. H. F. Smith, Free software helps map and display data, EOS Trans. AGU 72 (1991), 441.
- [48] Charlie Zender, netCDF Operator (NCO).

Appendix B

Preliminary Open Data Metadata and Data Dictionary Requirements

Preliminary Open Data Metadata Requirement

Metadata Field	Requirement	Usage Description
Dataset Level		
Dataset Title	Required	Descriptive name for dataset
Dataset Description	Required	Summary explanation of dataset contents, purpose, origination,
		methods and usage guidance. Avoid jargon where possible.
Tags	Required	Enter descriptive keywords which describe the subject groups for the
		dataset and help it to be found in searches.
Organization	Required	Agency, department, board or commission publishing the dataset.
		Also known as Publisher.
Contact Name/Program	Required	Enter the name of the contact who maintains the dataset.
Contact Email	Required	The email for the dataset maintainer.
Public Access Level	Required	Whether this info could ever be made public. (Public, Restricted, Non-Public)
License	Required	List any restrictions on use of the data. Most often "Public Domain"
Program	Optional	Program or cross functional team name.
Spatial Coverage	Optional	Name of defined area or geometry of area data describes.
Temporal Coverage	Optional	Start and end time of events described in data.
Frequency	Optional	How often data needs to be updated.
Language	Optional	Most often English
Торіс	Optional	Select a subject area from the defined list.
Homepage URL	Optional	URL for the page with useful information on the program creating the
		data.
Limitations	Optional	Appropriate usage notes, disclaimers and conditions of use.

Resource Distribution Fields			
File Title	Required	Descriptive name of the file.	
File Description	Optional	Summary explanation of file contents, purpose, origination, methods and usage guidance. Avoid jargon where possible. Include for all in dataset or none.	
Download URL	Optional	Optional if uploading	
Format	Optional	File format such as CSV, PDF, XML, SHP or JSON	
Data Standard	Optional	Established file structure defined for a particular use.	

Preliminary Data Dictionary Requirement

Data Dictionary Element	Description
column	The name of the field from the data table
type	(text, numeric, timestamp)
label	Common English title for the data contained in this column. Please avoid abbreviations if possible.
description	Full description of the values included in the column. If the value is a date, document the time zone of recording, e.g. PDT (Pacific Daylight Time). If the column is a category, such as age group, then all categories or levels should be listed. If the values are calculated, the source of raw data and calculation method should be included.

