California Water Data
CONSORTIUM



From Data to Resilience:
A Shared Vision for a Coordinated, Data-Informed Water Future in California

Supplemental Materials

Table of Contents

Supplemental Materials 1

Section I: Building Momentum and Strucutral Capabilties Solutions 2

Section II: AI- Specific Solutions 7



California Water Data CONSORTIUM

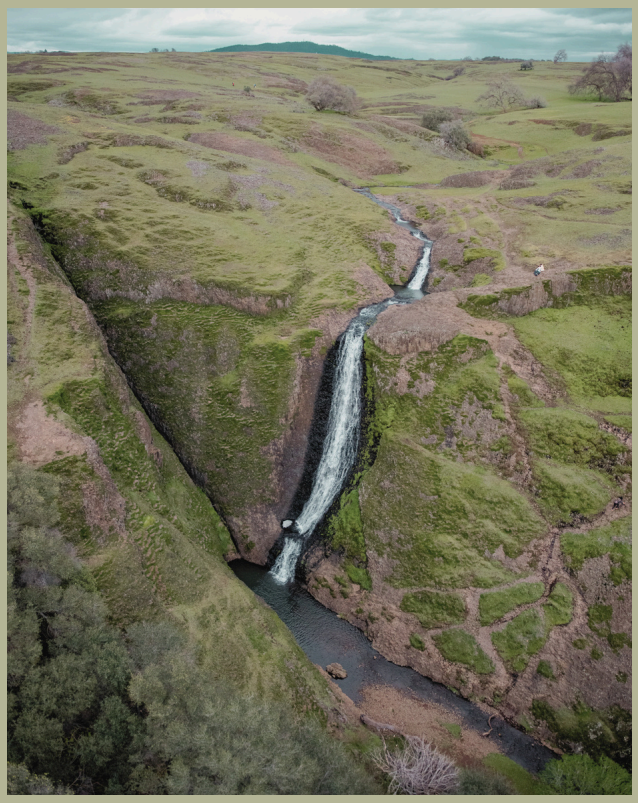
Supplemental Materials

Additional solutions and AI-specific opportunities for California's water information systems

About the Supplemental Materials

From Data to Resilience identifies interrelated solutions to modernize California's water information systems and advance the policy brief's outcomes. The main body emphasizes foundational solutions recommended for near-term implementation by the next administration and legislature. **Section I** provides additional detail on six solutions that complement these near-term priorities and support a broader pathway for sustained modernization and investment. **Section II** explores the opportunities and challenges of applying artificial intelligence (AI) to water management and presents eight additional AI-specific solutions not included in Figure 4 of the policy brief.

Each solution below, except those in the structural capabilities group, includes an indicative cost category to communicate scale. "**Low**" represents investments generally in the millions of dollars and "**Moderate**" represents investments in the tens of millions. These are not final estimates; actual costs will depend on scope, timing, implementation approach, existing investments, and opportunities to leverage federal, state, regional, and other resources.



Aerial view of Table Mountain showing Hallow Falls (Kelly M. Grow / California Department of Water Resources, 2022)

SECTION I
Building Momentum and Structural Capabilities Solutions

SECTION II
AI-Specific Solutions

SECTION I

Building Momentum and Structural Capabilities Solutions

Section I of these supplemental materials provides additional detail on six solutions that complement the policy brief's near-term priorities and support a broader pathway to sustained modernization and investment.

Solutions in this Section

6. Define minimum data standards
7. Support regionally-coordinated data hubs
8. Provide data literacy training and professional development pipelines
9. Integrated water resilience and accounting systems
10. Modernize open real-time monitoring and early-warning infrastructure
11. Integrated emergency decision-support and communications platform

Building Momentum Solutions

Three high-value solutions that build on the solutions included in Foundational Solutions. These actions are not quick fixes. They are feasible programmatic next steps expected to result in enhanced regional and cross-agency coordination, flexibility, and decision-making. They are also expected to build both state and local technical capacity and support technical workforce development in the water sector.

Structural Capabilities Solutions

Three actions to enable meaningful changes to natural and engineered systems, thereby positioning the state to proactively meet future needs, and support more legal and regulatory flexibility. These solutions rely on implementation of foundational and building momentum solutions, and are intended to create integrated information systems that can be used to manage water resources across multiple outcomes. Cost estimates are not provided for these solutions.

Building Momentum Solutions (2-3 Years)

Leadership and Governance

6

Define minimum data standards to ensure interoperability across agencies, programs, and jurisdictions. In partnership with the Water Insights Lead, define standardized parameter vocabularies, naming and identifier conventions, units, timestamps, and coordinates across state and local water entities. Establish standardized data quality and provenance categories that distinguish observed, modeled, extrapolated, and estimated data. Consider requiring Digital Object Identifiers to support dataset traceability and citation. Wherever possible, focus on unifying data definitions rather than dictating rigid file formats. Begin with highest-priority needs, including groundwater monitoring and extraction wells, surface water diversions, public water system sources and sampling points, and environmental compliance points.

HIGH VALUE

Improves data quality, accessibility, and integration, and makes it easier to synthesize and share data across state agencies and programs.

LOW COST

$4.5 million in one-time costs would support the initial development and implementation of the standards.



California Department of Water Resources Division of Integrated Science and Engineering staff members record data from fish collected from a fyke trap in the Yolo Bypass' Toe Drain in Yolo County (Nick Shockey / California Department of Water Resources, 2026)

Build Capacity

7

Support the development of regionally-coordinated data hubs centered on safe and affordable water supply, disaster response, and data-informed decision making. These hubs would coordinate more cost-effective regional data collection, governance, and sharing while also supporting timely and thorough reporting to state regulatory programs.

HIGH VALUE

Regionally-coordinated data hubs increase consistency across jurisdictions, build a common understanding of conditions, support regional flexibility, streamline reporting, and reduce total costs. They also build regional water resilience and capacity.

MODERATE COST

$60 million in one-time costs would support the initial development of regional data hubs, technical assistance, communications, and coordination. Ongoing funding of approximately $30 million annually would support state and regional personnel responsible for maintaining the hubs, facilitating coordination, and ensuring compliance with state reporting requirements. These investments - ideally from a mix of state and local sources - are expected to reduce long-term state agency technical assistance costs by strengthening regional capacity and streamlining data management and reporting.



The Russian River (jmoor17 / Getty Images Signature)

8

Provide data literacy training for staff at all levels to build awareness, understanding, and practical application of water and ecological data. Include establishing professional development pathways to help mid-career staff build skills in modern data management, analytics, and AI applications, and strengthen the capacity of data scientists and other technical staff to support decision making.

HIGH VALUE

Builds state and local agency technical capacity

LOW COST

$4 million annually would support personnel for data literacy and training in coordination with the Water Insights Lead, including communication, outreach, and ongoing development of curriculum and training materials.

Structural Capabilities Solutions (4+ Years)

Connecting information about water availability, infrastructure, ecosystems, climate risk, community vulnerability, and proposed investments would allow agencies and affected communities to evaluate options using a common evidence base, understand trade-offs, and track outcomes over time. Three solutions would advance this kind of integrated decision support:

9

Develop integrated water resilience and accounting systems. Build on California's significant investments in groundwater accounting to support open-source, integrated surface water–groundwater information and accounting systems across watersheds and basins. These systems would connect information now held in separate data systems, including water availability, rights and diversions, consumptive use, pumping, recharge, water quality, environmental flows, transfers, and allocations to provide a shared understanding of water resources, how they are used, and where risks and opportunities are emerging. They depend on the granular monitoring and telemetry described in Solution 10, and the two should be developed together. This integrated information would support more effective water management and transfers, protect community and environmental water needs, and help target investments and interventions before shortages or water quality challenges become emergencies. Integration with existing tools such as CalEnviroScreen, the state's environmental justice screening tool, would help incorporate environmental justice and equity considerations into risk assessments and resource prioritization.

Critically, these systems should use consistent accounting methods, common standards, and shared identifiers applied statewide, while retaining the flexibility to address regional priorities and needs. This would allow for a consistent, rolled-up understanding of water availability, use, and quality at both the watershed and statewide scale, supporting development of the California Water Plan and other state agency responsibilities.



A recharge project in Sacramento County, managed by the Omochumne-Hartnell Water District, that diverts water from two points of diversion along the Cosumnes River during high flow events (Sara Nevis / California Department of Water Resources, 2024)

10 **Modernize open real-time monitoring and early-warning infrastructure**, including expanded telemetry to support the accounting systems in Solution 9 and to improve local, regional, and state flood warning systems, drought response, post-fire watershed monitoring, hydrologic model development, and forecasting. Critical datasets needing additional monitoring and telemetry include stream gages, surface water diversions, and groundwater elevations. Critical monitoring gaps vary by watershed. Analysis should identify those gaps and the resources needed to fill them, as the 2023 California Stream Gaging Prioritization Plan did for stream gages. That plan now informs the California Stream Gage Improvement Program.

11 **Integrate emergency decision-support and communications capabilities.** Enhance local and state agencies' capability to integrate and share real-time water and environmental conditions, including streamflow, reservoir storage, groundwater conditions, and other critical data—alongside tools for multi-agency communication and coordination. Whether a new platform or just the improved integration of existing ones, the goal is to give local, regional, and state water and emergency managers access to a shared operating picture during disasters, ensuring critical information is visible beyond the agency that collects it and enabling faster, more coordinated emergency response.



A drone view of water from the Sacramento River overtop the Tisdale Weir in Sutter County, California (Andrew Nixon / California Department of Water Resources, 2024)

SECTION II

AI-Specific Solutions

Opportunities, risks, and applications of generative AI in water management

The three outcomes presented in the policy brief (safe, clean, and affordable water; lifesaving disaster preparedness and response; and sustainable agriculture and food production) depend on timely, contextualized, accurate, and usable information. Artificial intelligence (AI) tools can support delivering this information to a broader range of users, but only if the underlying data are well governed, appropriately structured, and thoroughly documented. The AI-specific solutions below therefore complement and build on the brief's solutions, including the statewide feature identity registry, minimum data standards, the Water Insights Lead, and local capacity building. They do not replace them. In addition to the policy brief and supplemental material's 11 solutions, this section presents six AI-specific solutions.

The term artificial intelligence covers a range of distinct technologies. Traditional machine learning has long supported water management in California and is broadly uncontroversial, though limited data and workforce capacity have constrained wider adoption. Generative AI (GenAI), by contrast, produces new text, code, data, or other content, most visibly through large language models (LLMs) such as Claude and ChatGPT. This technology is newer and far more contested, and this supplement focuses on it.

GenAI Tool Opportunities

These tools can now extract structured information from unstructured records, open up data through plain-language queries, fill gaps in monitoring networks, and model water availability and risk.

GenAI Tool Risks

These tools also carry global risks, from cybersecurity threats to misinformation, and water-specific ones: the energy and water demands of data centers, new vulnerabilities for water infrastructure, and the potential for results to be used without adequate human oversight.

Water Management Applications at the Nexus of Information Systems and GenAI

The Data to Resilience Initiative is not positioned to resolve broader debates about GenAI. As the technology proliferates, however, there is an urgent need and opportunity to identify safe and appropriate GenAI applications in water management. Testing GenAI on real water data use cases can show where GenAI tools add value, where they fall short, and where oversight and guardrails are needed, without assuming that broader tensions surrounding GenAI's development and governance are resolved.
The examples below show how GenAI could advance the brief's outcomes. The AI-specific solutions that follow build the data, governance, and evaluation capacity needed to pursue them safely.

Strengthening water data

Filling monitoring gaps. Generative geospatial models, combined with machine learning and in-situ, satellite, and sensor data, can estimate snowpack, streamflow, diversions, and groundwater pumping where direct measurements are limited.

Making data findable and usable. Semantic search can make state open data portals searchable in plain language, and genAI can extract usable data from records such as historic well completion reports that are too costly to digitize by hand.

Improving models and analyses

Climate-resilient reservoir operations. GenAI can generate large ensembles of plausible but never-before-observed hydrologic scenarios, letting operators test operating rules against conditions the historical record does not contain.

Faster, richer hydrologic modeling. GenAI emulators can approximate slow physically-based models in a fraction of the time, and generative techniques can fuse incomplete datasets to estimate conditions such as soil moisture or water quality.

Supporting decisions, dialogue, and public engagement

Plain-language access. Natural-language interfaces let water managers and community members query data and models without coding expertise, including in languages other than English.

Synthesizing public input. GenAI can help categorize large volumes of public comment in planning, permitting, and regulatory processes, surfacing key themes for decision-makers more quickly.

Realizing these opportunities will require water-specific benchmarks and evaluation frameworks that show where AI adds trustworthy value, where it does not, and what the lessons mean for GenAI in broader science and public-sector work. Building on the brief's Foundational Solutions, the following AI-specific solutions should be developed in coordination with the GenAI advisory group (Solution 4 in the brief) and integrated into existing statewide and interagency efforts to advance responsible AI use. Each AI-focused solution notes the Foundational, Building Momentum, and Structural Capability solution it builds on, and the examples it enables.

Priority Solutions (Year 1)

AI 1

Inventory critical datasets and make them AI-ready. In alignment with Solutions 3 and 6 in the policy brief and Section I of the supplemental materials (the statewide feature identity registry and minimum data standards), identify priority datasets and bring them up to AI-Ready standards. Initial datasets should be targeted to advance broadly useful outcomes, such as emergency response, endangered species protections, and water supply reliability (see the FAIR + AI-Ready box below). AI cannot correct inaccurate legacy records, such as historical diversion reports. The inventory would therefore assess each dataset's quality, document known limitations, and start with data reliable enough to support AI use. Where legacy data are weak, the findings would inform the audit and minimum data standards (Solutions 1 and 6 in the brief) so future submissions improve.

HIGH VALUE

Making critical datasets AI-Ready establishes the data foundation other GenAI applications rely on. It also complements the statewide feature identity registry.

MODERATE COST

$12 million in one-time costs would assess, reformat, and document priority datasets. About $1 million annually would support three dedicated staff positions to test and maintain AI-readiness standards as new datasets and technologies are added.

Builds on Solutions 1, 3, and 6 in the policy brief and supplemental part I. Enables all of the examples above.

AI 2

Add semantic search layers to open data portals this would let users search by meaning, in plain language, across datasets. It would likely entail natural language processing to support conversation-based search across disparate datasets, reducing the specialized knowledge currently required to find relevant data.

HIGH VALUE

A low-cost, high-visibility action that improves public and agency access to existing data assets

LOW COST

$4 million in one-time development costs, plus modest ongoing maintenance

Builds on Solution 6 in Section I of the supplemental materials. Enables "making data findable and usable."

Safely Leveraging AI to Build Momentum (2-3 Years)

For the past decade, California's water data work has focused on making data FAIR: findable, accessible, interoperable, and reusable. Since the passage of the Open and Transparent Water Data Act (AB 1755) in 2016, state agencies have made a large share of their water data public and easier to find. That progress is the foundation for what comes next.

But FAIR standards were not designed with AI in mind. They ensure that data can be found and shared, but not that AI tools can use it safely or appropriately. A dataset can be fully FAIR-compliant and still lack what natural language interfaces, semantic search, machine learning, and other AI-driven methods need:

- Standardized data structures (schemas) so tools can tell what each field means
- Machine-readable file formats in place of PDFs and one-off spreadsheets
- Consistent units and identifiers so stations, wells, and water bodies can be linked across datasets and agencies
- Embedded metadata that travels with the data and explains how it was collected and how it can be used

Good data governance, meaning accurate, quality-checked, and documented data with a clear owner, is needed whether or not AI is ever used. It lets software and new users read, understand, use, and link it to other sources and AI-ready data is not possible without it. The next decade of water data in California should build on FAIR by moving to "FAIR + AI-ready" datasets to support more timely emergency response, endangered species protections, and more reliable water supplies.



California Water Data Consortium attendees at the Putting Data to Work Summer Roundtable (Jason Alviar, 2025)

AI 3

Build natural language interfaces to critical datasets. Deploy LLM-based tools that let water managers, community members, and other stakeholders query priority datasets in plain language, including languages other than English, without coding or data science expertise.

HIGH VALUE

Expands practical access to water data for non-technical users and historically underserved communities.

MODERATE COST

$8-12 million in one-time development costs across priority datasets, then $3 million annually to maintain the tools, monitor their use, and review outputs for accuracy.

Builds on Solutions 5 and 8 in the brief. Enables “plain-language access.”

AI 4

Pilot self-hosted open-weight AI models for water management. Open-weight models are models whose underlying parameters are publicly available, so agencies can run them on their own infrastructure. The pilot would include strong technical and oversight safeguards and test which hosting approaches best balance cost, transparency, security, and performance for state agency needs. Specific to water management, the models would also be tested for real-world applicability in specialized hydrologic tasks and adherence to key physics-based principles. If the pilot shows value, the state could scale self-hosted deployments across agencies to further reduce dependence on the largest commercial AI developers. ”

HIGH VALUE

Provides transparency into where and how models run and the resources they consume, and reduces long-term reliance on the largest commercial AI developers. It also generates practice guidance on which self-hosting approaches are actually worth scaling, based on agencies’ existing infrastructure and workloads.

MODERATE COST

$12 million in one-time infrastructure and pilot development, then $4 million annually to host, evaluate, and refine candidate architectures.

Builds on Solution 4 in the brief. Provides hosting for all of the examples.

AI 5

Pilot generative geospatial (Earth foundation) models to fill gaps in water-cycle monitoring. Test emerging models that can estimate water conditions where direct measurements are missing, potentially in partnership with technology companies, national laboratories, or federal agencies such as the National Aeronautics and Space Administration and the United States Geological Survey.

HIGH VALUE

Could substantially reduce the cost of filling long-standing monitoring gaps if pilots demonstrate reliability at appropriate spatial and temporal scales.

LOW COST

$3 million for an initial pilot and evaluation phase. Scaling costs would depend on pilot results.

Builds on Solutions 9 and 10 in the brief and supplemental Part I. Enables "filling monitoring gaps."

AI 6

Fund initial research on the implications of GenAI for water. Direct grant funding to study GenAI's water use, energy demand, and cybersecurity implications; its effects on environmental justice in frontline communities; and the benchmarking, evaluation, and calibration research needed to assess and deploy foundation models reliably.

HIGH VALUE

Builds the evidence base needed to distinguish useful genAI applications from unhelpful or even harmful ones and provides the information necessary for constructive public dialogue about costs and benefits.

LOW COST

$7 million in initial grant funding and $1 million to administer the grant program, which could recur as research needs evolve.

Builds on Solution 4 in the policy brief. Provides the evidence base for the risks described above.



Lake Arrowhead, California (T M , 2020)

California Water Data
CONSORTIUM

In partnership with Earth Genome



cawaterdata.org