INTERNET OF WATER:
SHARING AND INTEGRATING WATER DATA FOR SUSTAINABILITY

A REPORT FROM THE ASPEN INSTITUTE
DIALOGUE SERIES ON WATER DATA

THE ASPEN INSTITUTE

For all inquiries, please contact:

Energy & Environment Program
The Aspen Institute
One Dupont Circle, NW
Suite 700
Washington, DC 20036
Phone: 202.736.3576

Nicholas Institute for Environmental Policy Solutions
P.O. Box 90335
Duke University
Durham, NC 27708
Phone: 919.613.8709
nicholasinstitute@duke.edu

Redstone Strategy Group LLC
3223 Arapahoe Avenue, #210
Boulder, CO 80303
Phone: 303.606.87100
redstonestrategy.com

The Aspen Institute
One Dupont Circle, NW
Suite 700
Washington, DC 20036

Published in the United States of America in 2017 by The Aspen Institute

Printed in the United States of America
Publication Number: 17/011
ISBN: 0-89843-665-6

INTERNET OF WATER: Sharing and Integrating Water Data for Sustainability. A Report from the Aspen Institute Dialogue Series on Water Data. 2017. Lauren Patterson, Policy Associate, Water Policy Program, Nicholas Institute for Environmental Policy Solutions at Duke University; Martin Doyle, Director, Water Policy Program, Nicholas Institute for Environmental Policy Solutions at Duke University; Kathy King, Associate Principal, Redstone Strategy Group, LLC; and David Monsma, Executive Director, Energy and Environment Program, The Aspen Institute.

INTERNET OF WATER:
SHARING AND INTEGRATING WATER DATA FOR SUSTAINABILITY

A REPORT FROM THE ASPEN INSTITUTE
DIALOGUE SERIES ON WATER DATA

THE ASPEN INSTITUTE

THE ASPEN INSTITUTE

The Aspen Institute is an educational and policy studies organization based in Washington, D.C. Its mission is to foster leadership based on enduring values and to provide a nonpartisan venue for dealing with critical issues. The Institute has campuses in Aspen, Colorado, and on the Wye River on Maryland's Eastern Shore. It also maintains offices in New York City and has an international network of partners.

THE ASPEN INSTITUTE ENERGY & ENVIRONMENT PROGRAM

The Aspen Institute **Energy and Environment Program** (EEP) is an active and prominent convener of non-partisan policy dialogue and neutral forums focused on key energy and environmental topics. The Program's mission is to take-up the enduring questions about nature and society, and to prompt new thinking about the security and wellbeing of both. Not as an adjunct of doing well by doing good but for the purpose of deliberately testing assumptions and policies about energy, climate change and conservation. The Program promotes values-based dialogue by bringing together a cross-section of thought leaders from industry, government, non-governmental organizations, and academia to address complex energy and environmental policy challenges in a collegial atmosphere. These gatherings provide fertile ground to allow deliberation, creativity, collaboration, and compromise to flourish among diverse participants in a discussion about environmental sustainability in a technological society. Like the Aspen Institute as a whole, the Energy and Environment program seeks to inspire and explore new ideas and provoke action in the real world. *www.aspeninstitute.org*

Duke NICHOLAS INSTITUTE FOR ENVIRONMENTAL POLICY SOLUTIONS

The Nicholas Institute for Environmental Policy Solutions at Duke University improves environmental policymaking worldwide through objective, fact-based research to confront the climate crisis, clarify the economics of limiting carbon pollution, harness emerging environmental markets, put the value of nature's benefits on the balance sheet, develop adaptive water management approaches, and identify other strategies to attain community resilience. The Nicholas Institute is part of Duke University and its wider community of world-class scholars. This unique resource allows the Nicholas Institute's team of economists, scientists, lawyers, and policy experts not only to deliver timely, credible analyses to a wide variety of decision makers, but also to convene these decision makers to reach a shared understanding regarding this century's most pressing environmental problems. *www.nicholasinstitute.duke.edu*

REDSTONE

Redstone Strategy Group, LLC helps philanthropies, non-profits, and governments solve the world's most urgent social problems. Redstone's work is informed by a deep appreciation of their clients' expertise, rigorous thought, and more than a decade of experience. *www.redstonestrategy.com.*

TABLE OF CONTENTS

PREFACE

The Aspen Institute Dialogue Series on Sharing and Integrating Water Data for Sustainability was convened to address one of the country's most pressing challenges: how to improve our water data infrastructure to enable us to more sustainably manage our water resources.

To address this challenge, this Dialogue Series, hosted by the Aspen Institute Energy and Environment Program in partnership with the Nicholas Institute for Environmental Policy Solutions and the Redstone Strategy Group, aimed to formulate a national water data and information policy framework for sharing, integrating, and disseminating public data to characterize and forecast the quantity, quality, and uses of water across the United States. Participants included experts and representatives from the federal, state, and local government agencies, the private sector, academia and non-governmental organizations.

This document represents the group's findings as developed over the course of three scoping sessions and two roundtable dialogues convened in 2016-2017. Here we provide a summary of the major findings, shared principles and action-oriented recommendations toward creating a national water data and information policy framework for sharing and integrating open water data. The intended audience for this report is threefold:

- Policymakers for whom this document can inform their thinking and strategic approaches;
- The community of practitioners focused on water data and management issues and their implications for water sustainability and innovation; and
- Funders who will support the implementation process and initiatives put forth in this report.

The Aspen Institute team is grateful for the generous support from the S.D. Bechtel, Jr. Foundation, Kingfisher Foundation, Walton Family Foundation, and Pisces Foundation that made this Dialogue Series possible. Our thanks as well to Martin Doyle of the Nicholas Institute for Environmental Policy Solutions and Kathy King of Redstone Strategy Group for all of their guidance, support and hard work in the development and shaping of this dialogue and report. A special thanks also to

our rapporteur Lauren Patterson of the Nicholas Institute, who fully captured the complexity of this conversation throughout the dialogue series, and Nikki DeVignes, our project manager for the dialogue whose proficiency and attention to detail is exceptional.

This report is issued under the auspices of the Aspen Institute. Not all views expressed in the report are unanimous and not all comments reflect individual expectations or understanding of the dialogue. The experts identified in this report participated in and are affiliated with the dialogue in their individual capacity. Their titles and affiliations are included for identification purposes only and their organizations are not responsible for the report's content. We will continue to support these important discussions regarding the sharing and integration of open water data across the U.S. and its broader institutional implications. We hope that this report proves useful to the water data community in strengthening efforts to collaborate on shared standards and expectations that guide data sharing and overcome institutional norms and barriers.

David Monsma
Executive Director
Energy and Environment Program
The Aspen Institute

VISION

The Aspen Institute Dialogue Series on Sharing and Integrating Water Data for Sustainability was designed to address one of the country's most pressing challenges: *how to improve our water data infrastructure to enable us to more sustainably manage our water resources.* Currently, we are unable to answer fundamental questions about our water systems in a timely way:

- How much water is there?
- What is its quality?
- How is it used (*i.e.*, withdrawn, consumed or returned for different purposes)?

The data needed to answer these questions often exist, although collected by multiple agencies across different scales of government and non-government organizations for different purposes. Since data are scattered across multiple platforms with different standards, much of it cannot be re-used beyond the primary purpose for which it was collected and is not used or ever transformed into information that supports real-time decision-making, identifying trends and patterns, or forecasting future conditions at a larger scale.

To address this challenge, *the Dialogue Series aimed to formulate a national water data and information policy framework for sharing, integrating, and disseminating public data to support the sustainable management of water*. As a result, we focused attention on the data needed to create water budgets (*i.e.*, characterize historic and current quantity, quality, and uses of water in watersheds across the United States). This Dialogue Series builds on existing state, regional and national efforts. For example, the Open Water Data Initiative consists of federal actors (*e.g.*, USGS, EPA, NOAA, USDA) and non-federal actors (*e.g.*, CUAHSI) that have made significant progress in addressing many of the technical barriers, and have developed and introduced standards for water data. Individual states are also integrating water data, through efforts such as Colorado's Decision Support System and California's recent passage of the Open and Transparent Data Act. Other efforts have focused on integrating state data. For example, WaDE is a collaboration among western state agencies to allow sharing of water use data in a more streamlined and cost-effective way. Each of these projects focuses on a portion of the water cycle or a specific geographic region.

However, river basins don't follow political boundaries and water budgets consist of a variety of data types; requiring coordination between ongoing data collection efforts.

The 2015 Aspen-Nicholas Water Forum: Data Intelligence for 21st Century Water Management found there were significant institutional barriers and norms that discouraged data sharing and prevented these efforts from reaching scale. In response, the Aspen Institute Dialogue Series convened stakeholders across water sectors to outline key principles to encourage and promote making data open, sharing data, and connecting ongoing data sharing efforts through an "Internet of Water". **The Internet of Water follows the organizational structure of the Internet with a backbone organization that provides support and governance structures to ongoing data sharing communities; connecting these communities to one another.**

In the following report, we articulate a policy framework addressing institutional barriers to scaling the integration of water data and information to support sustainable water management. The framework is organized around three core findings of the Dialogue Series. Each finding is further specified through principles and accompanied by recommendations for actions to advance the development of a national water data and information policy.

FINDINGS AND PRINCIPLES

FINDING 1: THE VALUE OF OPEN, SHARED, AND INTEGRATED WATER DATA HAS NOT BEEN WIDELY QUANTIFIED, DOCUMENTED, OR COMMUNICATED

Principle 1.1: A user-based approach will maximize the value of water data.

FINDING 2: MAKING EXISTING PUBLIC WATER DATA OPEN IS A PRIORITY

Principle 2.1: All public water data needed to characterize and forecast water budgets should be open by default, discoverable, and digitally accessible.

Principle 2.2: Water data standards to promote interoperability, efficiency, and user-flexibility will evolve in response to user demand.

Principle 2.3: Data producers are responsible for sharing data of known quality and documenting essential metadata; end users bear final responsibility for determining whether the data is fit for use.

Principle 2.4: Data should be shared as openly as possible, consistent with the principle that any security and privacy risks associated with sharing need to be balanced with the potential benefits.

FINDING 3: THE APPROPRIATE ARCHITECTURE FOR AN "INTERNET OF WATER" IS A FEDERATION OF DATA PRODUCERS, HUBS, AND USERS

Principle 3.1: Control and responsibility over data is best maintained by data producers.

Principle 3.2: A federated system of public water data hubs provides scalability and financial stability to better meet the diverse needs of data users.

Principle 3.3: A backbone organization should link data hubs and facilitate governance of the system, but not govern the production or use of data.

RECOMMENDATIONS

ACTION 1:	**ARTICULATE A VISION**
Recommendation 1.1:	Articulate a vision of sustainable water resource management and stewardship enabled by open, shared, and integrated public water data.
Recommendation 1.2:	Initiate an Internet of Water through regional pilots that solve near-term water management problems for key stakeholders through shared and integrated water data.
ACTION 2:	**ENABLE OPEN WATER DATA**
Recommendation 2.1:	Develop water data catalogs that identify all existing public water data maintained by states.
Recommendation 2.2:	Develop tools for opening existing, public water data and enable the use of those tools by producers and users.
Recommendation 2.3:	Bind regulation, management practices, permitting, and funding to the provision of open data.
ACTION 3:	**CREATE AN INTERNET OF WATER**
Recommendation 3.1:	Existing water data hubs should be stabilized and further resourced.
Recommendation 3.2:	A backbone organization should be formed to structure and enable a system of federated data
Recommendation 3.3:	The backbone organization should be a non-profit organization but with a cooperative agreement with a federal, non-regulatory agency.
Recommendation 3.4:	Develop proof-of-concept on integrating data from multiple hubs to advance a water budget.

PART 1: FINDINGS AND PRINCIPLES

The Dialogue Group developed three major findings around the challenges and opportunities of creating a national water data and information policy framework. The key findings of the group are that:

(1) water data are undervalued;

(2) there is a priority on simply making public water data open; and

(3) water data could be most effectively integrated through a federated system of local, regional and national data hubs supported by a national backbone organization.

Through this report, we refer to this federated system as the Internet of Water (IOW). While many elements of the IOW currently exist across the nation, the findings and recommendations below put forward a strategy to rapidly accelerate its development over the coming years.

Progress towards a national water data and information policy requires coordinated action of a large number of public and private sector actors across local, state, regional, and national scales. Given the need to coordinate action across numerous and diverse stakeholders, the group chose to adopt a principles-based approach to data sharing and integration. A principles-based approach outlines standards and expectations to guide data-sharing efforts while allowing variation and flexibility on implementation. In addition, a principles-based approach to the IOW allows individual organizations to continue their specific missions while simultaneously enabling a broader network of sharing and integration. These principles are a starting point and should be revisited over time as technology advances and the primary needs around water data evolve.

The Dialogue focused around the need for open water data to address one of the country's most pressing challenges: *how to improve our water data infrastructure to enable us to more sustainably manage our water resources*. To sustainably manage any resource, there needs to be a good accounting system. Currently, we are unable to answer fundamental questions, referred to as a "water budget", about our water systems in a timely way:

- How much water is there?
- What is its quality?
- How is it used (*i.e.*, withdrawn, consumed or returned for different purposes)?

Examples of water data that address quantity, quality and use are provided in Appendix 1.

FINDING 1: THE VALUE OF OPEN, SHARED, AND INTEGRATED WATER DATA HAS NOT BEEN WIDELY QUANTIFIED, DOCUMENTED, OR COMMUNICATED

The clear, tangible, and potentially transformative benefits of new approaches to managing water data need to be articulated for public agencies and organizations to motivate them to invest in opening public water data. Data producers often bear the financial burden of data collection and management while users realize the benefits. Articulating the value of data sharing to both data producers and users is necessary for there to be growing investment in, and acceptance of, water data sharing.

Some of these values include:

- **A common foundation for negotiation and decision-making.** By providing a common foundation of facts, better availability and use of public data can reduce litigation, build trust, and support consensus-building between stakeholders.

- **Improved analysis allowing better planning and decision-making for sustainable management.** Data on water budgets are necessary to assess whether water systems are being managed sustainably, and to guide real-time decisions, along with rapid forecasting of future conditions. Integrated data sharing reduces the time analysts spend locating, cleaning, and estimating data, allowing them to instead spend time and resources on analysis more directly related to management decisions and forecasting.

- **Increased precision across sectors and purposes.** Better data improves the information base for the myriad of decisions affecting water, making the decisions and resulting actions more effective for their intended purposes. In the same way that better field-level data allowed the transformation of "precision ag," better water data could lead to precision rate-setting and precision infrastructure (*e.g.*, point of use) by utilities, potentially enabling efficient water markets.

- **Creating space for innovation.** Better integration of data about water systems supports innovation in water management and its associated technologies. Data should be considered necessary infrastructure for 21st century water management, creating opportunities for innovation in ways that cannot be anticipated at the outset.

- **Public engagement and education.** Data visualizations and interfaces can be a powerful tool for public engagement, making water issues relevant and resonant. There is great potential of consumer-facing data products for the public, which would build appreciation for water systems. For example, communicating to the public how water gets from the environment to their sink requires data from multiple different sources that are typically dispersed.

Principle 1.1: A user-based approach will maximize the value of water data.

A user-centered approach advances water data sharing and integration, as value is created by specific organizations and institutions while using those data to improve sustainable water management. **A user-centered approach maximizes the realized value of data sharing and integration. Determining how the data are used—including analysis, synthesis, or modeling—should be driven by the users themselves. Users are best able to determine which data are needed, and how those data can advance their own purposes.** Data users may also develop analytical or modeling approaches quite different from that envisioned or previously used.

Identifying anchor tenants will help advance data sharing and secure the needed continuation of integration efforts. **Anchor tenants are sectors, organizations, and specific institutions who would significantly benefit from the use of shared and integrated water data, and therefore champion an open data initiative by demanding the data from the public sector.** By growing the IOW in response to the specific needs of anchor tenants, the clear value of the investments can be identified and articulated by those who benefit most immediately.

As more public water data become open, and the sheer size of the IOW grows, data that were initially integrated in response to the priorities of anchor tenants will be put to novel uses by new users. In addition, private industry will enter the field to further tailor data services to the needs of specific user groups. While some of these will likely be fee-based, others will likely develop which will enable general search or synthesizing functions for low or no cost. The Weather Channel, which packages and distributes already free National Weather Service data, provides an example of this evolution with meteorological data.

FINDING 2: MAKING EXISTING PUBLIC WATER DATA OPEN IS A PRIORITY

Making public water data open and accessible is the most necessary step in using water data for sustainability. Without open data we cannot realize or demonstrate the value of water data (see Finding 1) or integrate water data to answer fundamental questions about our water systems. Open data typically refers to data that are available, discoverable, and accessible. Many follow-on activities, from private sector analytics to inclusion of new types of data such as crowd-sourced or citizen science data, will be predicated on the availability of open water data. Thus, the Dialogue consensus was to promote any steps that advance open water data. The clear message from the group was simply "just get the data out there."

Principle 2.1: All public water data needed to characterize and forecast water budgets, should be open by default, discoverable, and digitally accessible.

The initial focus should be on making *already existing public water data* open and accessible. There are volumes of public data that already exist because they are collected by public agencies or because they are mandated to be reported for regulatory purposes. Much of these data can inform real-time water management decisions as well as provide a base for long-term planning purposes moving towards sustainability. Yet these data are often diffused and difficult to access or synthesize; and so much of the data cannot readily be re-used beyond the primary purpose of the data producer. Simply making these data open and accessible could transform water management.

Open data means that the data are free to use, re-use, and redistribute with no restrictions on their use. The federal government along with many states and municipalities have adopted open data policies, making open data the default for all public data; this practice by definition includes all public water data nationally. In setting openness as the default, for reasons of privacy or security, it may not be appropriate for certain water data to be shared openly (see Principle 2.4). In these circumstances, however, it is still important for others to know that the data exist, and as result, the existence of these data should still be discoverable.

To be usable, public water data should be first and foremost *discoverable*, meaning easily found by users via a search engine, within an application, or on a publicly accessible website. This principle is based on the notion that it is critical to know what water data exist. A common taxonomy of terms would help improve **data discoverability**, as would linking data to geographic locations. *Digitally accessible* refers to the data and associated metadata being accessible through an online platform that is *machine-readable* (*i.e.*, formatted in a standard computer language that can be read automatically by a web browser or computer system).

Principle 2.2: Water data standards to promote interoperability, efficiency, and user-flexibility will evolve in response to user demand.

Water data becomes increasingly valuable as they become more *interoperable*, both in a technical and a pragmatic sense. Technical **interoperability** allows computer systems to exchange information using communication protocols that enable data to be readily downloaded, uploaded, and exchanged. Pragmatic interoperability is typically accomplished via data standards that establish some common information exchange reference to remove ambiguity such as collection methods and data units.

Ensuring interoperability requires establishing data standards. *Data standards* refer to documented agreements on the representation, format, definition, structuring, tagging, transmission, manipulation, use, and/or management of data and associated metadata. Data standards reduce errors introduced in merging data from different sources and increase the discoverability of open data, and so are vital to realize the full value of open data. The water community has a growing set of accepted data standards (see Box 1), but given the number and diversity of data collectors in the water field, these standards are far from universal.

Water data have been collected by different users for different purposes, with each having their own units, terminology, resolution, etc. As a result, setting rigorous standards for open data at the outset will keep many valuable data sources closed and inaccessible. **Rather, the priority should be on increasing the potential use and integration of existing public data, regardless of the standards under which the**

BOX 1: ALREADY EXISTING DATA STANDARDS

Some standards for water data already exist and will increasingly be adopted over time. For example, the Open Geographic Consortium (OGC) developed spatial standards adopted by the federal government and large private data industries such as Google, Oracle and Esri. WaterML2 is a data exchange standard in hydrology for time series data used by public and private sectors including the USGS, NOAA, and CUAHSI. The Water Quality Exchange (WQX) provides a schema of standard water quality data and formats used by 400 federal, state, tribal and watershed organizations. The National Hydrograph Dataset (NHDplus) provides a geospatial framework for 2.7 million stream reaches across the nation that can serve as the common geo-referencing foundation for all water data (essentially providing the address). By including NHDplus geospatial referencing in WaterML2 and WQX water data records, the utility of the data is greatly enhanced by enabling data records to be linked with locations on surface hydrology features such as rivers, lakes and wetlands. The most appropriate standards are best developed by the communities of practice, and will evolve toward increased standardization over time.

data were initially collected. Over time, data producers and users will iteratively arrive at data standards and protocols that are appropriate to meet their respective needs. There was a strong consensus in the Dialogue Series that absence of widely accepted and used data standards in the water community should not delay efforts to make existing water data more open and accessible.

Principle 2.3: Data producers are responsible for sharing data of known quality and documenting essential metadata; end users bear final responsibility for determining whether the data is fit for use.

Data quality should be maintained through a chain of responsibility from those collecting data to end users. Consistent with a user-centered approach to data sharing and integration (see Principle 1.1), there was agreement in the Dialogue Series that the data users bear responsibility for whether data are fit for use – that is, whether the data are of sufficient quality to be used in their specific analysis or application. The group also observed that decision-makers are often highly constrained by the availability of data and would likely benefit from having *some* data of known quality rather than very limited data of the highest quality. **Over time, data quality will improve and become optimized as users increasingly use or demand data at a resolution and quality that meet their needs or products.** It is possible, and even likely, that data resolution or precision will decrease for some parameters as users indicate that their needs do not require a previously assumed level of precision.

To empower end-users to assess fitness for use, they need information about the quality of the data they are using. Data producers need to document the quality of data as part of the associated metadata, and these data and metadata must be maintained with a high fidelity by all data managers.

Consistent with Principle 2.2, the dimensions used to document data quality should not be mandated at the outset, but instead allow for flexible descriptions of data quality, which may include attributes such as: completeness, uniqueness, timeliness, validity, accuracy, resolution, precision, and consistency. When data quality is known but cannot be fully characterized, the data producers should provide a qualitative level of confidence in the data. When data are of unknown quality, the data should still be made available, but should be clearly identified as having unknown quality.

In the long-term, standardization of metadata is incredibly important for harmonization among datasets and across geographies to occur. Development of metadata standards and models will require development of tools such as: controlled vocabularies, taxonomies, thesauri, and data dictionaries. It is assumed that over time, a codification of methods and metadata for different types of data will emerge. Having good metadata is immensely critical as it impacts *data discoverability.*

Principle 2.4: Data should be shared as openly as possible, consistent with the principle that any security and privacy risks associated with sharing need to be balanced with the potential benefits.

Public data should be open by default. That said, it is possible and potentially appropriate to restrict access to certain public data as a result of legal restrictions, or in measured response to justified reasons of privacy, confidentiality, or security concerns. Reasons for not sharing public data need to be understood and balanced with the potential benefits of sharing the data more openly.

Data privacy refers to interests that an individual or institution may have in restricting access to particular kinds of information *about* them. Privacy considerations most clearly apply to personally identifiable information (PII) and proprietary information. The federal government has set standards for the protection of PII, and these standards need to be respected in the management of water data. That said, privacy considerations are frequently invoked in data sharing discussions to cover a wide range of interests and concerns, and the experience from other fields is that people are often willing to relinquish their data in favor of other goods. As a result, it is vital to understand the motivations behind privacy concerns, including the potential harm(s) that the individual or institution anticipates as a result of increased access to particular information. With better understanding of the underlying interests, and potential harms, it is often possible to identify management approaches (*e.g.,* masking or blurring) to release the data, and protect legitimate interests (see Box 2).

The notion of private and non-identifiable information is an evolving concept, as evolving technologies, such as remote sensing, may in the near future give access to information that is currently thought to be private or non-identifiable.

BOX 2: BLURRING AND MASKING WATER DATA

Instances where data are geo-referenced, aggregation (*e.g.,* at a HUC 12) or other data blurring approaches may be required to ensure that the data are de-identified. For example, the value of water use data for a water budget may largely be realized on a small watershed scale (*e.g.*, approx. 40 square miles, HUC 12 scale). This may require the identification of a non-regulatory 3rd party with established privacy protocols to become an aggregator and distributor of raw data identified as sensitive. Water use data may also be aggregated by industrial or customer class for benchmarking purposes, noting that some classes (*e.g.,* UIC or NDPES) do not have their privacy protected because they are already in the public domain. Ideally, as a culture of data openness emerges, the data will become less anonymized to enable greater precision in sustainably managing water resources.

Data confidentiality is similar to data privacy in that it is concerned with limiting access to information. That said, confidentiality is not focused on limiting the ability to trace information back to particular individuals or institutions, but rather limiting access to information to limit the risks of particular harms. The more confidential the information, the more restricted the access to the information. In the water system, for example, confidentiality is a concern for data relating to the safety ratings of the nation's dam systems.

When broad sharing of public data is not possible, because of confidentiality concerns, controlled and tiered access may be appropriate rather than complete inaccessibility. *Controlled access* allows only a limited set of users to access the data in any form. Users may be required to meet certain characteristics and agree to terms of use. *Tiered access* provides differential access to the data by multiple user groups, and users may be required to meet certain criteria and agree to specified terms of use. Additionally, raw data may be available for pre-approved users, while aggregated data may be more widely available.

Data security refers to the protective measures established to prevent unauthorized access to data and protecting data from corruption. Data should be secured through current best practices in the information technology sector.

FINDING 3: THE APPROPRIATE ARCHITECTURE FOR AN "INTERNET OF WATER" IS A FEDERATION OF DATA PRODUCERS, HUBS, AND USERS

Making public water data open does not unleash the power of data; rather it is in the integration of data, the inter-accessibility of data, and the iterative process between end-users and data producers that are the necessary components of making data powerful. The Dialogue Series convened around a shared vision of an Internet of Water (IOW) in which open public water data would be shared through a network of communities. While there is some transferability to private water data – an eventual goal of the IOW – **the present focus is on public water data.** Each component of the network described below (data producers, data hubs and a backbone organization) has different roles and responsibilities.

Data producers are the organizations collecting public water data. Examples of data producers include an irrigation district in its collection of diversion data, the USGS through its collection of stream gauge data, or an industrial water user monitoring pollution discharge at an outfall to meet NPDES permit requirements. The production of public data can be in response to a regulatory requirement (*e.g.*, EPA water quality permit) or as part of a public data program (*e.g.*, USGS stream gauge program). While *public* data producers may generate many other types of data as part of their operation, much of these data are for internal, operational purposes and not relevant to creating a water budget. For the purposes of opening water data, the focus is on data relevant to sustainably managing a water system: quantity, quality, and use. Data producers can also play an aggregation role; for instance, a state watershed monitoring program may collate data from multiple sources (water quality sampling stations and wastewater outfalls) to demonstrate compliance with watershed-wide regulations (*e.g.*, TMDL).

A key characteristic of data producers is that they have *authority* over how data are being produced. Because of this authority role, data producers shape the data being collected in terms of its resolution, frequency, collection method, reporting standards, etc. Alongside authority, data producers also have *intentionality* for the data being collected; in other words, there is purpose behind the data they are choosing to collect to fulfill their public roles and responsibilities. In some cases, this will be fulfilling a regulatory role--the data characteristics must be sufficient to ensure compliance with rules and regulations (*e.g.,* a utility-producer demonstrating compliance with EPA requirements). In other cases, this will be fulfilling a scientific or monitoring mission (*e.g.*, USGS, NOAA-NWS) in which case the produced data characteristics must meet evolving needs of the scientific community or agency mission.

A **data hub** is a formalized, structured source of open public water data. Some data hubs may upload the data onto their own storage servers while other data hubs may point to open data managed and stored by producers. Data hubs could be organized geographically (*e.g.*, by state or basin) or topically (*e.g.*, groundwater monitoring, water quality monitoring). There is a spectrum between data producers and data hubs. Data producers could eventually serve as a data hub if and when public data are made open, shared, and integrated. For example, the USGS stream gauge program serves as a data producer and as a data hub. Additionally, some states are moving toward making public data open, such as Colorado's Information Marketplace or Water Data for Texas. Data hubs that are also data producers have the potential to quickly make significant amounts of water data open because they possess the authority to make the data open.

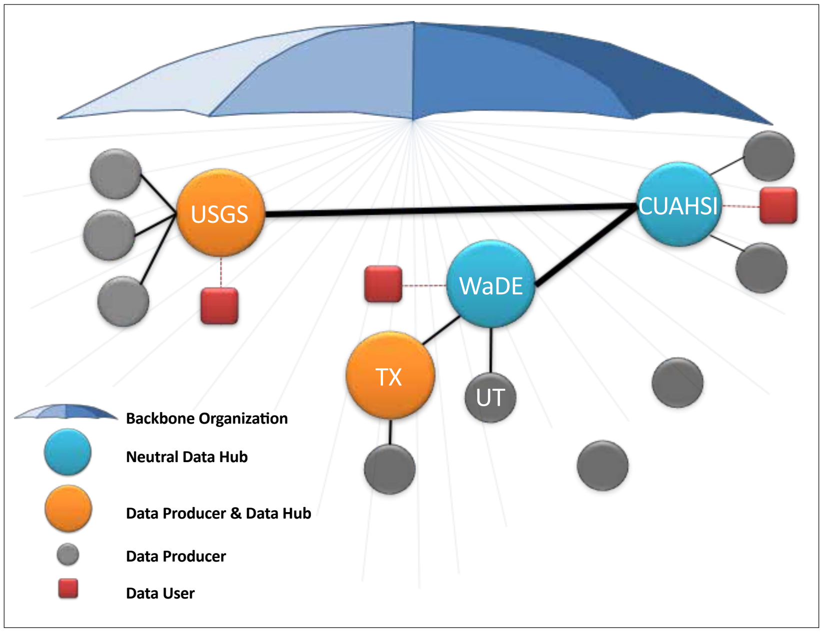
However, because different data users have different intentions and authorities, data hubs which are not producers can play the role of a trusted third party responsible for ensuring data integrity is maintained and privacy/security protocols are followed. An important distinction between data *producer-as-hub* versus a *neutral hub* is that in the latter case, the data hub will not have authority or control over data production.

Data users are entities that use open water data to create value. A data producer can be a data user, as can private organizations, academia, and government agencies.

The **backbone organization** is (currently envisioned as) an umbrella governance structure connecting data hubs to one another, as well as to data producers and users.

The **Internet of Water** is the system of interconnected data hubs through the leadership of the backbone organization (Figure A).

Figure A: Conceptual Diagram of the Internet of Water



Principle 3.1: Control and responsibility over data is best maintained by data producers.

In a federated system, data producers will maintain control over how water data are collected, stored, and curated. The purpose behind the data being collected, as well as current data standards and collection protocols, is to meet the mission and/or regulatory requirements of data producers (*i.e.*, intentionality) and thus should continue to do so. Because of their authority over data characteristics, data producers will also have the role of making data public (see Principle 2.1) in terms of requirements and logistics.

Culturally, there remains discomfort or distrust in sharing data. Participants noted a range of sources for this discomfort, from a culture of water managers and utilities to remain "below the radar," to negative experiences with data sharing that resulted in what was deemed to be inappropriate secondary analyses or misleading characterization of the data. By having data producers maintaining authority and responsibility for the data, with no regulatory authority or oversight from neutral data hubs or the backbone organization, confidence and trust can be maintained. Indeed, the neutrality of neutral data hubs provides greater potential for data sharing by industries or individuals than might be possible otherwise, as well as increases the potential for private data to eventually be provided and incorporated into the Internet of Water.

Principle 3.2: A federated system of public water data hubs provides scalability and financial stability to better meet the diverse needs of data users.

Structurally, a federated system is a good fit for the highly distributed nature of water management and use across the nation. In particular, it allows initial data sharing and integration to occur in response to user interests and priorities. This flexibility is consistent with a user-centered approach to development (see Principle 1.1). Data hubs often form to meet a specific need, having identified a value proposition that mobilizes funding and stakeholder participation. The primary role of a data hub is to share and integrate open data. As such, the size and shapes of hubs can vary widely. A hub may be small, such as a single watershed within a state; yet small hubs can have a large impact, particularly if they can quickly show the value proposition of sharing water data to address a problem (see Action 1).

Larger data hubs may span multiple states over a long period of time and require incorporating, standardizing, and integrating large volumes of disparate types of data (see Principle 2.2). Federated data hubs designed around creating value to meet specific needs are more likely to obtain support and interest from stakeholders (see Box 3).

In addition, a federated system decentralizes and diversifies the funding sources for the data integration efforts, which the group felt was likely to increase the economic sustainability of the IOW. Individual hubs may initially struggle to secure funding on a year-to-year basis until the value of water data has clearly been shown. However, more generally there will be greater stability for the Internet of Water as the diversity of funding serves a similar function to portfolios of diverse stocks in a fund.

Principle 3.3: A backbone organization should link data hubs and facilitate governance of the system, but not govern the production or use of data.

Current data hubs are disconnected from one another, producing insights specific to their community but not enabling broader sharing of insights. In the same way that governance and coordinating entities were necessary to create the Internet, the backbone organization will support and connect federated data hubs and aggregators by providing at least four key functions: advocacy and marketing; technical support; advising; and coordination (see Recommendation 3.2). Some of these functions need to be provided by an outside organization because data hubs would potentially compromise the trust already built within and between data producers if they were to step into an advocacy or marketing role. The background organization will not hold any data, but rather is envisioned to form a community of practice sharing a culture of openness around agreed upon principles.

BOX 3: DATA HUBS MEETING USER NEEDS

The Western States Water Council's Water Data Exchange (WaDE) is providing a platform for state agencies to shared derived datasets related to water planning; *i.e.*, water supply, availability, use and water rights. Bay Delta Live is a collaborative community focused on understanding the complex and dynamic ecosystem of the Sacramento-San Joaquin Bay Delta. Data hubs do not have to be spatially bound but could be defined by a unifying purpose. For example, the Open Water Data Initiative (OWDI) is designed to integrate fragmented data systems, starting with data systems in the federal government, while the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) is designed to facilitate publication of water data and models within the academic, research community. Communities may also form around the data. In addition to OWDI, the federal Advisory Committee on Water Information created the National Water Quality Monitoring Council (NWQMC). The NWQMC has developed a data portal to provide water quality data collected by federal, state, tribal and local agencies, and the National Groundwater Monitoring Network, which has in turn developed a portal to share groundwater data from multiple data producers. Both portals provide access to data held by multiple stakeholders in a common format.

PART 2: IMPLEMENTATION AND RECOMMENDATIONS

The findings and principles articulate the vision and guidelines for the development of a national water data information and policy framework. In this section, we will focus on the necessary and sequential steps to start implementing the findings over the next three to five years. These recommendations are interdependent and aimed at cultivating an active data sharing community around water data. That said, there may be an opportunity to phase in these recommendations over the coming years, which we indicate below.

ACTION 1: ARTICULATE A VISION

An essential first step is building the case for why open water data are needed, as well as the value in sharing and integrating water data. Clearly articulating the value proposition is fundamental to building a network of participants in the IOW (*e.g.*, increasing the number of data hubs) as well as building the necessary political and financial support to construct and maintain the elements of the IOW.

Recommendation 1.1: Articulate a vision of sustainable water resource management and stewardship enabled by open, shared, and integrated public water data.

To date, there has not been a compelling case made for investing in data sharing and integration. While water managers typically share a common sense of the value of an integrated water data system, the absence of quantified benefits is a significant barrier to building political support and investment. Early investment in quantitative studies of the value of open water data, including both cost savings (*e.g.*, money and time saved, opportunity costs) and its impact to more sustainably manage water systems (*e.g.*, increased environmental flows, water savings, and preserved habitats), should be an early priority to build the case and momentum for the IOW.

There is an opportunity for governments, academia, the private sector, and foundations to invest and participate in quantifying the benefits of open water data. The results of the analyses should be openly available to encourage further investment and development of the IOW.

Recommendation 1.2: Initiate an Internet of Water through regional pilots that solve near-term water management problems for key stakeholders through shared and integrated water data.

To ground the Internet of Water in a user-based approach, the public and private sector should develop water data sharing and integration efforts around specific water management problems, and around anchor tenants who would initially champion the development of the IOW. **Anchor tenants are key sectors, organizations, and institutional users who would initially benefit from open water data and integrated data sharing platforms.** These users would both realize significant near-term benefits from data integration efforts, and would also be visible examples of the benefits to other stakeholders. There was consensus that decision-makers and water managers were promising initial anchor tenants, with the general public (*e.g.*, appreciation for the value of water, understanding where water comes from) being longer term users of such integrated systems (see Box 4). Regional pilot studies would be quantitative analyses and case studies that make use of integrated and shared public data.

BOX 4: EVOLUTION OF ANCHOR TENANTS

The development of Global Positioning System (GPS) is a powerful example of how data sharing efforts are supported by an early anchor ***tenant*** and later expand to meet a wider range of use cases. GPS was initially developed for the military and then for airlines. It was not until private businesses created derivative products, such as Google Maps, that the general public became an anchor tenant for, and driving demand for improvements of, GPS technologies. The adoption of open and integrated data by decision-makers, the showcasing of success stories and value propositions will drive the demand for better data and better sharing. As the quality of data and the value of data sharing become quantified, the demand may shift to private companies and individual customers.

WATER MANAGERS: Access to water data and information creates value by clearly defining, and reducing uncertainty with regards to the amount, quality and use of water in a system. Water manager's benefit from reducing uncertainty. Improved data and improved forecasting via integration can further reduce temporal uncertainty in the system, enabling increased precision in water management.

WATER UTILITIES: Water utilities will see benefits for the rate-payers (lower costs), the utility (avoid new infrastructure), and the environment (more water is left in the system). Policymakers are beginning to demand utilities provide data to enact innovative rate structures or fund utility projects, while on the other side of the spectrum, individual customers want to see their water usage at a higher resolution than monthly.

AGRICULTURE: Agriculture could benefit from having a "water truth" in the basin where everyone agrees on how much water is in the system and its quality. Having this information could enable precision agriculture at a scale larger than the field. Agricultural and food industries rely on a supply chain located in geographically disparate regions; interoperable data would enable understanding immediate and long-term hydrologic risk exposure.

ENVIRONMENT AND SUSTAINABILITY: Collective actions are needed to sustainably manage ecosystems. Each stakeholder in a watershed, or an aquifer, holds different pieces of the puzzle. Data integration and sharing are necessary to know how much water there is, it's quality, and how it is being used. This information will allow better decisions to be made proactively, rather than reactively, to manage water needs for environmental and sustainability purposes.

BUSINESS AND INFRASTRUCTURE: In some regions, businesses are demanding water data to improve their situational awareness prior to relocating or expanding. Water security is becoming an important consideration of lending institutions and insurance companies, although both lack relevant data across the diversity of portfolios to allow understanding and quantifying systemic risk exposure.

ACTION 2: ENABLE OPEN WATER DATA

A primary focus of the principles is to make public water data open, transparent, and available for decision-making. The following recommendations outline key next steps to advance open water data.

Recommendation 2.1: Develop water data catalogs that identify all existing public water data maintained by states.

Basic data discoverability is an essential pillar to the Internet of Water. An initial step to increasing the *discoverability* of water data is to develop a public data catalog. Ideally, this catalog would include a list of *all* public water data, its metadata, and a link to the open data and the contact information of the data producer. States are logical candidates for hosting such a data catalog since states have authority over water rights, are often the permitting authority for water quality parameters, know the universe of data available, and often collate data from local entities such as utilities and irrigation districts for state water management plans. Even if data are not accessible in a machine-readable format, simply knowing those data exist creates enormous value, and thus **this recommendation is a critical, necessary step that should be done regardless of whether the subsequent recommendations are acted on.**

A catalog also facilitates gap analyses of water data both within a state and between states. Such analyses could go state-by-state to identify all water-related data available in a data registry, including data registry performance in meeting the principles of containing metadata, a working link to the data, and contact information of the data producer. **The goal of the gap analysis is to identify where the data do not exist to create a water budget (quantity, quality and use) for each state.** Gap analyses would also help promote different data sharing strategies, set benchmarks, and inform scorecards that can incentivize states to open water data. For example, a goal could be to improve the data coverage needed for a water budget from 20 to 80 percent, and metadata coverage from 10 to 100 percent. These analyses could be conducted by state governments, social sector partners or the new backbone organization (see Recommendation 3.2). To maximize impact, the results should be openly available and findings should be communicated to state officials as part of the advocacy and marketing strategy of the backbone organization.

Recommendation 2.2: Develop tools for opening existing, public water data and enable the use of those tools by producers and users.

Existing data need to be structured so that the data can be served through an API—Application Programming Interfaces. APIs are a set of rules to create consistency about how data can be accessed from a website; that is, if data are to be made available online, APIs establish how those data will be formatted. Through this consistency of reporting, APIs provide fast, continuous access to data, which is essential to hubs and users. While federal, state, and local governments are starting to create APIs to facilitate access to water data, their use is far from widespread. To increase the use of APIs, data producers should develop APIs for their particular type of public water data (*e.g.*, water utilities should develop APIs for public data produced and used by water utilities).

Many utilities, water districts, local, and state governments may be unfamiliar with APIs, and lack the internal capacity to develop these web-services. Smaller organizations, such as small water utilities or groundwater districts with constrained human or financial resources, are essential sources of data for water sustainability yet will likely be unable to rapidly produce their data consistent with APIs. In these instances, resources will be needed to create API templates, along with instructions for conducting such data transitions, by water sector and state (assuming each utility, district, industry, etc. are required to produce and report data in the comparable format). Further, there will need to be an education component and/or mechanisms to empower the actual transfer of data from current formats into APIs. Quite simply, once APIs are developed, many data producers will need to be resourced to transition their data into APIs.

The implementation of APIs will increase the use of water data by sophisticated public and private data actors. However, the general public will likely be unable to make use of these new forms of data in a short time. **As such, basic portals will need to be developed to allow the broader public to make use of data made available via this effort by data producers.** More sophisticated data users, including private sector users, will be primary drivers of early innovation in putting such API-enabled data to use, and are most likely to create value from the data in the near term. Nevertheless, basic, entry-point portals should be publicly available as well, although with limited capabilities (*i.e.*, no widgets or data visualizations).

Recommendation 2.3: Bind regulation, management practices, permitting, and funding to the provision of open data.

A particularly efficient mechanism of opening up public water data is to make regulatory compliance or funding contingent on the provision of open water data. An immediate opportunity would be to make water data related to permits and/or grant applications open data and potentially required to be made open as part of permit or grant application process (see Box 5). In addition, applicants for a permit or a grant to a regulatory agency typically must provide some type of data in support of that application, and often there is a large amount of data associated with the permit application. Regulatory agencies could require that rather than submitting those data, the grant or permit application process requires that the data be made open as part of the application process.

When checking for regulatory compliance, or when evaluating a permit or grant application, decisions should be based on open data rather than communicated data (*e.g.*, data emailed directly from producer to regulator). This slight change in mechanism provides both a requirement and a check for appropriate open data: if a link to data is broken, a permit application cannot be evaluated.

It is important to note that this alternative requirement would be an additional burden for public organizations and local agencies unless there has been a prior enabling of such data management. The workflow for a government agency to produce regularly refreshed, well documented, machine-readable data requires skilled resources. Some attention to capacity building should be addressed by the backbone organization (see Recommendation 3.2). Thus, Recommendation 2.3 is highly contingent upon Recommendation 2.2 and 3.2.

BOX 5: NATIONAL POLLUTION DISCHARGE ELIMINATION SYSTEM

There is a tremendous amount of data that are required to be collected and reported as part of regulatory permits such as National Pollution Discharge Elimination Systems (NPDES), groundwater extraction, and water diversions. In many cases, these data are reported in formats that are not open or easily accessed. For example, NPDES permit holders will often collect significant amounts of water quality data, but those data are either not directly reported or are reported through uploading or emailing spreadsheets to regulatory agencies. Instead, if such data were made open, large volumes of data would immediately be generated and could be used to understand the spatial distribution of water quality throughout a watershed.

ACTION 3: CREATE AN INTERNET OF WATER

An Internet of Water is dependent on there being open water data, water data hubs, and some broader backbone organization that will serve several functions of weaving together those hubs along with setting the overall trajectory.

Recommendation 3.1: Existing water data hubs should be stabilized and further resourced.

Across the U.S., there are existing and emerging water data hubs. Notable examples include WaDE, CUAHSI, the National Groundwater Monitoring Network, and the Water Quality Portal. In addition, a number of states are making their water data public (*e.g.,* Water Data for Texas, Colorado Information Marketplace), which will function in ways comparable to water hubs. These existing hubs have developed strong relationships with data producers and users, along with vital infrastructure to support data sharing. **The Internet of Water should initially grow through investment in these hubs, allowing them to expand within their existing missions and mandates, and increase interconnections between them.**

These hubs have diverse funding sources, and as a result, increased funding will rely on a number of actors. Consistent with the recommendations of this report, these hubs respond to distinct user-groups with particular topical or geographic interests, and their funding sources reflect these constituencies. For example, state data resources tend to be funded through state governments, whereas organizations such as CUAHSI and WaDE have a mix of funding from federal agencies, grants, and philanthropic funds. Across the board, the existing water data hubs need to be stabilized in terms of funding for a known time horizon; such certainty is necessary to demonstrate to data producers and users the benefits of this new approach to water data. In addition, there is an economy of scale to all data hubs; further resources could generate substantial benefits because of the fixed costs associated with the basic infrastructure (personnel, intellectual, physical) of creating and operating any data hub. Over time, it will be important and necessary to invest in new data hubs that respond to emerging needs, but the current priority should be to stabilize and further existing water data hubs.

Recommendation 3.2: A backbone organization should be formed to structure and enable a system of federated data.

A national backbone organization would play a vital leadership and organizational role in connecting and creating a unified vision for open water data between hubs. The core mission of the backbone organization is in line with the goals of making public water data open, accessible, and shareable for the purposes of addressing the three fundamental components of a water budget. Ongoing state or regional efforts that are taking leadership roles should not be delayed while waiting for a backbone organization to be formed. Rather, the backbone organization is meant to support open data efforts at whatever status those efforts exist.

The group identified four key capacities, with some overlap between capacities, which a national backbone organization would need to provide to help create a successful and growing data sharing environment within and between data hubs.

1. **Advocacy and Marketing:** The lack of a clear, unified voice to articulate the value of open, integrated water data systems was identified as one of the most significant barriers to progress. A key component of advocacy (and/or marketing) would be focused on creating a culture of data openness and transparency and working with data producers to open their data and engage with relevant hubs. Water is historically undervalued as a resource, water data even more so. This is a significant barrier to building political support and investment in open water data and data hubs. Part of advocacy would be to work with data aggregators and hubs to quantify the impact of the data sharing (see Recommendation 1.1), making that information available to help organizations make the case for investing in open water data efforts for a more sustainable future. Marketing involves promoting the value of open and integrated water data to a wide audience, creating a unifying vision for open data sharing. It is also about managing tensions between data hubs, producers and users by fostering opportunities to strengthen relationships between different elements of the IOW.

2. **Technical Support:** An important role, particularly in the earliest stages of the IOW, would be working with data producers to open their data, including metadata development (see Principle 2.3), security requirements (see Principle 2.4)

and tool development and support (see Recommendation 2.2). Additionally, data most often used by several hubs may become the first set of focus activities for developing common technical standards. States with limited budgets and capacities noted that if an entity like the backbone organization could provide templates for data registries, schemas, etc., then it would be more feasible for states (or local water districts) with limited budgets and resources to push data out or maintain a data portal someone else has built.

3. **Advising:** Advising here refers to non-technical support, such as establishing performance metrics and self-diagnostic tools to ensure data sharing efforts are progressing and deficiencies or gaps are identified (see Recommendations 2.1 and 2.3). The role is not punitive, but rather informative, allowing different agencies to benchmark their progress in sharing water data and identifying areas of improvement. These types of tools can demonstrate current leading practices, as well as provide options and opportunities for improvement. For example, WaDE may lead the innovation in how to effectively share data related to water use, starting first with derivative products and proceeding to raw data. The federal government may be a leader in sharing ambient monitoring data for water quantity and quality. CUAHSI may serve as the repository for water research data and modeling. The backbone organization would synthesize the lessons learned from these leaders and thus allow cross-fertilization between hubs, as well as enabling new hubs to be most efficient for their identified purposes, without mandating the approaches used by hubs. Additionally, the advising role could identify what data are necessary to create a water budget for sustainability and compare those data with the gap analysis to identify where the data are currently insufficient and prioritize open data sharing efforts to meet those gaps. The backbone organization will create a centralized registry pointing to the data registries of data hubs and aggregators.

4. **Coordination:** Coordination is distinct from marketing in that the focus is on strategic planning, and understanding the bigger questions or broader problems that data sharing communities could address collectively. The coordination role will look toward ways to tie the value of open data to an objective (*i.e.,* economic, infrastructure, inventories, etc.) and connect with potential funds (*i.e.*, grants) as they become available. The coordination role would also provide a place for data users to directly communicate and provide feedback or demand for improvements in data quality and standards in future iterations so data can evolve to solve new problems or be complimented by new technologies. The backbone organization will coordinate ongoing data sharing efforts to reduce redundancy and create efficiencies between projects.

Recommendation 3.3: The backbone organization should be a non-profit organization but with a cooperative agreement with a federal, non-regulatory agency.

In addition to there being widespread support around the functions of the backbone organization, there was also consensus that the organization be a non-regulatory and non-profit entity. The success of the backbone organization will be dependent on its capacity to act as a trusted broker between data hubs, data producers, and data users spanning multiple water sectors and interests. It was agreed the backbone organization should not be part of a federal agency due to the uncertainty associated with federal funding, and to engender trust between data producers and the backbone organization; much of the group felt that such trust would not be possible if the organization were part of a federal agency. A vital additional benefit of a not-for-profit structure is that it would allow the organization to freely advocate governments to support open water data resources, a key gap in the current landscape.

While important to not be part of a federal agency, the backbone organization may want to establish a cooperative agreement with a federal agency that allows it to directly support and extend the capacity of federal agencies to meet their mandates. The federal government plays a vital leadership role in collecting, integrating, and sharing ambient water data across the nation. They have also led the development of several technical standards and the geospatial fabric for the Nation. In addition, the mission of the Internet of Water is highly aligned with that of the federal government—coordination between states and local governments. For example, the mandate by the SECURE Water of 2009 for the USGS to develop a national water census, which is highly consistent with the water budget approach recommended here.

There are a number of fiscal structures that could underlie the backbone organization, most notably that of the University Corporation for Atmospheric Research (UCAR). As the backbone organization matures, it could develop a membership model through which it is supported by annual donations or fees from participating organizations, including state and local governments, as well as private industry. In addition, the organization could explore fee-based services, such as technical consulting, or value-add data products, to provide additional financial stability.

Recommendation 3.4: Develop proof-of-concept on integrating data from multiple hubs to advance a water budget.

While a federated data system provides flexibility to address local problems and obtain funding, there is additional value to further integration of data from hubs. For example, WaDE collects water use data but does not collect data related to natural systems. In contrast, data hubs such as the USGS, EPA and CUAHSI collect data on the quantity and quality of water within the natural system. **A water budget requires both of these types of data to be integrated within a watershed**. A proof-in-concept that develops a shared platform of the relevant data needed to create a water budget at the location of power plants within one basin of interest could provide a powerful example of integrating open water data between hubs for sustainable water management.

DIALOGUE PARTICIPANTS

Participants in the **Aspen Institute Dialogue Series on Sharing and Integrating Water Data for Sustainability** were invited as experts in their fields. As with all policy dialogues in the Aspen Institute Energy and Environment Program, the format followed the Institute's time-honored approach to intentional, values-based dialogue and adhered to a strict not-for-attribution rule throughout the duration of the Dialogue. Individuals who participated in the dialogue are listed for identification purposes only –their organizations do not necessarily endorse the report's narrative or findings.

Jerad Bales, Executive Director, CUAHSI

Ryan Barr, Director, E&J Gallo Winery

Kelly Bennett, Partner and Director, Research & Analytics,
Ponderosa Advisors/ Water Sage

Budhendra Bhaduri, Director, Urban Dynamics Institute, ORNL

Nathaniel Booth, Chief, Office of Water Information, USGS

Robert Bruant, Subsurface Manager, Pioneer Natural Resources Company

Albert Cho, Vice President, Strategy and Business Development, Xylem Inc.

Juliet Christian-Smith, Senior Climate Scientist, Union of Concerned Scientist

Peter Colohan, Director, Service Innovation, Office of Water Prediction, NOAA

Steve Davis, Digital Water Business Development, GE Power,
Water & Process Technologies

Ann Elise Debelina, Digital Marketing Leader, GE Water & Process Technologies

James Eklund, Director, Colorado Water Conservation Board

Claudia Emerson, Director, Institute for Ethics and Policy for Innovation,
McMaster University

Jack Fellows, Director, Climate Change Science Institute, ORNL

Greg Gearheart, Deputy Director, Office of Information Management and Analysis,
CA Water Resources Control Board

Biju George, Chief Operating Officer, DC Water

Maurice Hall, Associate Vice President, Ecosystems-Water, EDF

Sam Marie Hermitte, Program Specialist, Texas Water Development Board

Jim Jordahl, Director, Programs and Operations, Iowa Agriculture Water Alliance

Ted Kowalski, Senior Program Officer, Environment, Walton Family Foundation

Jeff Lape, Deputy Director, Office of Science and Technology, U.S. EPA Office of Water

Sara Larsen, Manager, Water Data Exchange Program (WaDE), Western States Water Council

Brewster McCracken, CEO, Pecan Street Inc.

Pam Muse, Supervisor, Planning & Data Management, Arizona Department of Water Resources

Mike Myatt, Program Officer, Water Foundation

Shirley Neff, Senior Advisor, EIA

Jeffrey Odefey, Director, Clean Water Supply, American Rivers

Greg O'Hearn, Information Systems Manager, Milwaukee Metropolitan Sewerage District

Jill Ozarski, Program Officer, Environment, Walton Family Foundation

Emily Read, Senior Advisor, Office of Water Information, USGS

Sarah Richards, Water Program Officer, The Cynthia & George Mitchell Foundation

Will Sarni, Principal, Water Foundry, LLC

Lea Shanley, Co-Executive Director, South Big Data Innovation Hub, Renaissance Computing Institute

Michael Sullivan, Global Cognitive Solutions Leader, Water and Environment, IBM

David Totman, Industry Manager, Global Water Practice, Esri

Deven Upadhyay, Manager, Water Resource Management, Metropolitan Water District of Southern California

Lisa Williams, Water Resource Manager, Arizona Department of Water Resources

Dwane Young, Senior Data Advisor, U.S. EPA Office of Water

MODERATORS:

Martin Doyle, Director, Water Policy Program, Nicholas Institute for Environmental Policy Solutions at Duke University

Kathy King, Associate Principal, Redstone Strategy Group, LLC

David Monsma, Executive Director, Energy and Environment Program, The Aspen Institute

RAPPORTEUR:

Lauren Patterson, Policy Associate, Water Policy Program, Nicholas Institute for Environmental Policy Solutions at Duke University

THE ASPEN INSTITUTE ENERGY AND ENVIRONMENT PROGRAM:

Nikki DeVignes, Program Manager

Greg Gershuny, Managing Director

DEFINITIONS

Accessible: data and associated metadata being machine-readable and available, eventually, via web services.

Anchor Tenants: sectors, organizations, and specific institutions who would significantly benefit from the use of shared and integrated water data.

Application Program Interface (API): a set of routines, protocols and tools specifying how software components interact (*e.g.,* set rules for scraping data).

Backbone Organization: an umbrella governance structure connecting data hubs to one another, as well as to data producers and users.

Controlled Access: only a limited set of users is allowed to access the data in any form; users may be required to meet certain characteristics and agree to terms of use.

Data Confidentiality: similar to data privacy in that it is concerned with limiting access to information, but focus is to limit the risks of particular harms.

Data Hub: a formalized, structured, source of open water data.

Data Privacy: interests that an individual or institution may have in restricting access to particular kinds of information about them.

Data Producer: an entity that collects data.

Data Quality: attributes such as completeness, uniqueness, timeliness, validity, accuracy, resolution, precision, and consistency.

Data Security: protective measures established to prevent unauthorized access to data and protecting data from corruption.

Data Sharing: specific arrangement by which two parties agree to transfer a defined set of data, subject to specific conditions of use, re-use, and distribution.

Data Standards: documented agreements on the representation, format, definition, structuring, tagging, transmission, manipulation, use, or management of data.

Data User: an entity, private or public, involved in accessing and investigating data.

Discoverable: easily found via a search engine, within an application, or on a publicly accessible website.

Federated Data System: a system in which data producers maintain data within their own system but allow access to data by others in the network through a shared catalog and agreed upon standards.

Harmonization: data are provided in a user ready format with all units and methods reconciled.

Internet of Water: a system of interconnected data hubs through the leadership of the backbone organization; the network by which open public water data would be shared

Interoperable: formatted such that computer systems can 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.

Machine-readable: formatted in a standard computer language that can be read automatically by a web browser or computer system.

NHDplus: a geospatial, hydrologic framework dataset built by EPA and USGS, first released in 2006.

Neutral Data Hubs: data hubs that have no authority or oversight of data producers.

Open Data: data that can be freely used, re-used, and redistributed by anyone, subject, at most to the requirement to attribute the source of the data.

Public Data: data that are collected by public agencies—federal, state, or local—and those data collected and/or reported for government-required regulatory purposes.

Standardization: data includes additional columns, such as units or methods.

Tiered Access: differential access to data by multiple user groups who may be required to meet certain criteria and agree to specified terms of use.

ACRONYMS

API	Application Program Interface
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science
EPA	Environmental Protection Agency
Esri	Environmental Systems Research Institute
GPS	Global Positioning System
HUC	Hydrologic Unit Code
IOW	Internet of Water
NHDplus	National Hydrology Dataset
NOAA	National Oceanic and Atmospheric Administration
NPDES	National Pollution Discharge Elimination System
NWQMC	National Water Quality Monitoring Council
NWS	National Weather Service
OGC	Open Geographic Consortium
OWDI	Open Water Data Initiative
PII	Personally Identifiable Information
TMDL	Total Maximum Daily Load
UCAR	University Corporation for Atmospheric Research
UIC	Underground Injection Control
USDA	United States Department of Agriculture
USGS	United States Geological Survey
WaDE	Water Data Exchange
WQX	Water Quality Exchange

APPENDIX 1

Synthesized results from eight participants responding to a survey regarding which data elements they use or manage and which they felt were important to include as core data elements.

CATEGORY	DATA ELEMENT	DESCRIPTION
Quantity	Streamflow	All geo-located time series data
	Reservoirs	Volume
	Groundwater	Volume, yield, level
	Evapotranspiration	Potential, Actual
	Climate	Precipitation, Temperature, Snowpack
	Soil Moisture	Includes soil characteristics
	Inter-basin Transfers	Volume and locations
Quality	Discharge	NPDES permits and data
	Water Treatment	Water utility intake data
	Groundwater	Well measurements
	Surface water	Lakes/streams
	Constituents	Temperature, DO, pH, TSS, TDS, N, P, Velocity, Depth
	Regulatory Thresholds	Federal, State, Local
Use	Withdrawals	By sources (GW/SW) and Sector (Municipal, Industrial, Ag, Energy, etc.)
	Consumption	By sources (GW/SW) and Sector (Municipal, Industrial, Ag, Energy, etc.)
	Return Flows	Includes point source discharges
	Reclaimed Water	Use, Stored, Discharged
	Water Transfers	Inter-basin as well as intra-basin
	Irrigated Acres	By crop type
	Water Right Appropriations	Historic and current
	GW / SW Laws & Management	

APPENDIX 2

The Dialogue Series included discussion about "early wins" through water information pioneers that take immediate steps to share and integrate water information and demonstrate how the overall vision and approach of the Dialogue could be achieved. Table 1 provides some illustrative examples to advance the ideals of water sustainability through substantially improved, shared and integrated water information.

Table 1. Potential Pilot Projects to Share and Integrate Water Information for Sustainability

PILOT PROJECT	POTENTIAL ANCHOR TENANTS	DESCRIPTION
Public Access to Real Time Drinking Water Information	-Water Utility/Provider	Customers would access near real time information about drinking water quality (leaving the treatment plant and in selected monitoring locations in the distribution system) on the utility's website.
Public Access to Information on Lead Service Lines	-Water Utility/Provider	A water utility compiles and posts information about the known extent of lead service lines throughout their service area. See DC Service Map (www.dcwater.com/servicemap).
Public Access to Wastewater Effluent Discharge and Receiving Stream Information	-Wastewater Utility -USGS -State -Watershed Groups	A wastewater utility posts ongoing effluent discharge monitoring data on its website along with appropriate ambient water quality information, provided by another entity (*e.g.,* watershed group, state, USGS). This helps the public understand the information and facilitates the watershed-based coordination.
Solving Water Quantity on a Regional/Watershed Scale	-Industry -Water Utility/Provider -States	Several parties convene to share information on water availability and use on a watershed scale to address near and long-term water quantity and use issues, such as facilitating water reuse.
Solving Water Quality Challenges on a Watershed Scale	-States -Utilities -Industry -Watershed Groups	Watershed stakeholders convene and develop an integrated watershed monitoring plan and data system that captures past, current and future water quality information (public and private) that then serves as the information foundation to develop watershed restoration efforts. Under some conditions, this could be facilitated under a voluntary framework or under a state issued watershed-based permit.