Building a Texas Water Data Hub as a model for National Water Data Infrastructure

by
Kelly Swanson

A capstone research project for Johns Hopkins University in conformity with the requirements for a Masters degree in Environmental Science and Policy

Baltimore, Maryland
December 2022

Executive Summary

As cities grow, agriculture expands, and energy needs continue, water is a resource that must be carefully tracked. In the past, different areas of water use and quality have not been monitored. Problems with water data exist from the ability to collect information to sharing it out effectively. To address this problem in Texas, this research focuses on Texas' effort to create a water data hub that can provide findable, accessible, interoperable, and reusable (FAIR) data (*Making Public Data FAIR*, 2018). Texas is one of several other states, counties, cities and federal organizations working to publish water data in the form of a hub platform. This effort is made to answer the problem of clearly analyzing data to understand how water is managed. In order to provide water resource managers with the best decision making information, water data needs to be collected, organized, and provided in useable ways. This research provides the steps taken by the Texas Water Data Hub team to find, understand, organize, and provide water data.

The Texas Water Data Hub team used several overlapping methods to gather data and create a water data hub. In this research, the steps will be presented one at a time, but it is important to note that this effort requires flexibility as many different functions of the hub must be addressed simultaneously. First, understanding the background of water data was addressed. Knowing past industry norms and vocabulary played a role in making new determinations. Next, collecting current data in an inventory and applying categories was completed. Working with Subject Matter Experts then helped to update standards and build relationships between datasets. Finally, working with stakeholders like data producers and users to understand and address data needs helped to make the alpha and beta design of the hub.

This research provides a detailed explanation of the steps the Texas Water Data Hub team took to address the problem of un-FAIR water data. By providing this information, other organizations can learn and take the necessary steps to organize water data as well. This research can also assist in scaling up this work to a federal level in order to provide a standardized across the nation.

The ability to access and analyze water data will improve the ability to make decisions regarding it's use. This is a large task in that water is used and operates in distinct ways, making it difficult to organize ridged categories. However, if standards and collective efforts can be made, decisions regarding this resource will be based on structured information rather than hopeful predictions.

Abstract

Findable, Accessible, Interoperable, Reusable (FAIR) water data is a buzz word in the industry for good reason (*Making Public Data FAIR*, 2018). Without these objectives, poor water data across the United States will continue to cripple the ability of decision makers to manage and develop sustainable practices (*Building Data Infrastructure*, 2022**)**. In an effort to implement these standards, this research was designed to first understand the past and current water data infrastructure throughout Texas and the United States and then create a findable, accessable, interoperable, and reusable (FAIR) water data hub (*Making Public Data FAIR*, 2018). An important part of this effort was to include stakeholders and decision makers from the water data industry. This research provides an overview of initial data collection and follows with detailed updates to water categorization and standards, stakeholder engagement and best practices, the creation of the Texas Water Data Hub and finally, recommendations to expand this state effort to a national level. The discussion speaks to the complexity of organizing water data due to the overlapping needs of such a project. The conclusion points out the additional challenges to scaling up these procedures to a national level. All of these efforts are part of building FAIR water data and is essential in our increasing need and care of water.

Acknowledgments

I would first like to thank the Texas Natural Resources Information Systems (TNRIS), a department of the Texas Water Development Board, for the time I was able to spend with members of the Texas Water Data Hub team. The work they do is exciting and so important to the future of water. Meeting several times thought the week to work on inventory, data analysis, design, and overall hub roll-out allowed me to appreciate the complexity of water data, but also provided motivation to continue in this effort. I learned a lot through group meetings and one-on-one interactions regarding organization and vision. Our work as a team was focused and effective while positive and motivational. A special thanks to Taylor Christian, my mentor, who included me in this work and guided my ability to contribute. Taylor allowed me to ask questions and offer suggestions while also sharing her expertise.

I am also grateful for Dr. Schwebel who patiently walked me through the research paper writing process. He always made himself available to meet with me and answer my questions.

Financial and Equipment Support

The Texas Water Development Board: Research and development of the Texas Water Data Hub

Cynthia and George Mitchell Foundation: Grant #R-2107-57642

Lyda Hill Foundation: Contract #2200012605 for grant

Windward Fund: Windward-WFIN-Texas Water Development Board-Subgrant-015603-2021-11-01

Contents

List of Tables

List of Figures

1. Introduction

Water is one of our most important resources. As such, it would be reasonable to assume that water is carefully looked after. Unfortunately, in the United States, this resource is often taken for granted. Water has been contaminated, left unmonitored and used in a variety of frivolous ways. As the condition of climate change becomes more apparent to our population, many of our resources have been given more consideration (Reimer, n.d.,p. 1). After witnessing communities abandoned when wells dry up, the need to turn to emergency supplies due to toxic drinking water, and reoccurring national drought, water is getting some new attention. More people in the public, political, and special interests arenas are beginning to understand the need to collect and organize water data as an important step in understanding this resource and using it wisely (Maidment, 2008).

David R. Maidment described the creation of a WATERS Network observatory in his paper “Brining Water Data Together” (Maidment, 2008). It was proposed that this framework connect and organize water data around the country rather than leaving management groups to their own methods. Furthermore, finding and using data presented in various formats makes it difficult to take advantage of the information that is recorded (Maidment, 2008).

The lack of standardized water data can be contrasted with records collected by the Energy Information Administration (EIA). While it is possible to research and predict energy usage, it is much more difficult to do so with water data. Water data would need to be collected from individual water utilities, of which there are often dozens per city in the US. To streamline this objective, *Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database* calls for a national water database in order to promote sustainable and best management strategies (Chini & Stillwell, 2017). Such a database would bring water data closer to the same standardization use for energy data.

When water data can be found, which is a major challenge, it is often difficult to interpret, shares broken metadata making it difficult to reuse, or doesn’t cover a wide enough rage of information to be helpful. Larger agencies like the U.S. Geological Survey (USGS) and the U.S. Environmental Protection Agency (EPA) host several water data collection sites like the Safe Drinking Water Information System, the National Water Quality Assessment program, the Water Quality Portal (WQP), the National Water Information System (NWIS), and have collected information like the Community Water Systems surveys published in 1995, 2000, and 2006 (Josset et al., 2019). While these sites make an attempt to compile and publish water data, it is rarely findable,

accessible, interoperable, and reusable (FAIR) (*Making Public Data FAIR*, 2018). See more programs and challenges made by the USGS in the **Literature Review** below.

To confront these long standing issues, New Mexico Representative Melanie Stansbury along with several cosponsors, introduced the Water Data Act to the House of Representatives in May 2022. This bill addresses the modernization of water management through a national water data system and aims to create a national Water Data Council and grant program supporting innovation in the industry. (Press Release, 2022).

This research examined existing water datasets, water policies and overall accessibility to water information at the state level in Texas, which can then be used at the national level for the United States. The goals of this research were threefold: 1) to create updated and enhanced accessibility and quality standards based on historical research for water datasets and metadata; 2) to update the definitions and relationships that describe water types in order to organize them ontologically; and 3) to highlight best methodologies for stakeholder engagement. To reach these goals, this research provides three processes to engage: 1) organize and publish the Texas Water Data Hub or state water data center, which can later be analyzed at the national level; 2) apply ontology programs to build data definitions and relationships; and 3) hold conferences, workshops and meetings to engage state agencies, private and public stakeholders.

1.1 Historical water data efforts

Although currently upgrading its prediction models, the USGS compiled water use reports for the Nation every 5 years since 1950. Available data was reviewed in order to make estimates for the Nation as a whole (*National Water Census: Water Use | U.S. Geological Survey*, 2019). Additionally, the EPA historically supported the Legacy Data Center (LDC) from early 1900's until 1998. The LDC stored water quality measurements for surface and groundwater (*Water Quality Data Download | US EPA*, 2022).

Similarly, Texas began compiling data and planning for the future of its water in 1957 after a severe drought. That year the Texas Water Development Board was established. Later in 1972, the Texas Natural Resources Information System (TNRIS) was developed as an extension of the Board and tasked as a central repository for water-related data (*About TWDB | Texas Water Development Board*). Across the state, 16 regional water planning groups and about 450 volunteers work together to update the 50 year water plan every 5 years (Rosen & Roberts, 2018).

1.2 Current water data access and standards

The most current federal level outlets for water data belong to the USGS and the EPA. Water Data for the Nation and the National Water Information Systems are hosted by USGS and include a section for “Current Conditions” and a download site and map with links to collected water data. Data categories hosted are: Surface Water, Groundwater, Water Quality, and Water Use *(USGS Water Data for USA).* Two problems associated with this data collection are the difficulty in downloading large files and the lack of information regarding water use (Josset et al., 2019). However, the USGS has been sited as the “gold standard” for data access when pertaining to surface water (Rosen & Roberts, 2018). For example, daily data from 29,152 sites is organized and presented in a FAIR process *(USGS Water Data for USA)*. As the USGS was consulted in the Texas Water Data Hub planning, additional USGS platforms are summarized in the **Literature Review** below to describe learnings from Subject Matter Experts.

The Water Quality Portal, supported by the EPA, provides optional categories for download, but after a lengthy query, the user is often unable to access data (*EPA, n.d.*). In general, the WQP is limited by its dependence on state-level data collection. The platform serves mainly to organize data found by states as they monitor and enforce national regulations such as the Clean Water Act (Josset et al., 2019). This reiterates the importance of state-level data hubs.

Agencies throughout Texas have been making efforts to better manage water data. Some of the work done by these organizations like The City of Austin, the Lower Colorado River Authority (LCRA), and the Texas Advanced Computer Center (TACC) is described in the **Literature Review** below. The “Connecting Texas Water Data Workshop: Building an Internet for Water” held in 2018 provided a good case study on current water data access and standards. The workshop hosted 90 experts from across Texas working in “government, water agencies, utilities, academia, business, industry, research institutes, and water associations and advocacy organizations” (Rosen & Roberts, 2018, p. 3). The workshop reviewed current concerns and sought to identify gaps in public data. Categories of data such as stream flow, soil moisture, water use, water rights, and water quality were among those were found lacking and data types with reusable metadata or available to the public in raw form were needed. Ultimately, the workshop solidified the need for FAIR data in Texas. For a full report of the workshop, along with survey questions and forms, see **Appendix A**.

As a result of the workshop and growing political and private support of updating water data, the Texas Water Developement Board organized the ideas around creating a Texas Water Data Hub (Hermitte, 2021). Alongside the project, this research addresses

the problem of how to create a centralized water data hub by understanding past and current water data, updating standard categories, and including influential stakeholders.

Below a **Literature Review** has been conducted to investigate the work of Subject Matter Experts consulted while building the Texas Water Data Hub. These 9 organizations were influential in determining new water data categorize (see **Table 3.2**) and gave advice for the vision of creating a hub. Additionally, the technical aspect of building a hub platform was influenced by the examples shared in the below **Literature Review**. See **Appendix C** for additional information on Subject Matter Experts.

2. Literature Review

The Texas Water Development Board consulted with 9 different organizations in order to collect preliminary ideas and practices best used to build a water data hub. These organizations include: California, New Mexico, Colorado, United States Geologic Survey (USGS), Internet of Water, City of Austin, Lower Colorado River Authority (LCRA), Texas Advanced Computer Center (TACC) and The First Street Foundation - Flood Factor. Literature published by these organizations was reviewed below in order to facilitate the efforts of the Texas Water Development Board to build a water data hub. Technical models and design, data storage, computing capacity and functions, were all studied and considered in this review.

California

The California Department of Water Resources was established in 1956 by the California State Legislature. The agency is responsibly to protect, conserve, develop and manage California's water supply. In 2020, the California Natural Resources Agency, a division of the Department of Water Resources, created an initiative called the CNRA Monitoring and Stewardship Unit (MSU) to centralize the way bond-funded projects are tracked (*MONITORING AND EVALUATION OF CALIFORNIA NATURAL RESOURCES AGENCY INVESTMENTS*, 2020, p. 1). The proposed outcome included the development of a "relational database system…that will enable spatially explicit analysis at multiple scales" (*MONITORING AND EVALUATION OF CALIFORNIA NATURAL RESOURCES AGENCY INVESTMENTS*, 2020, p. 22). The use of this system would then increase science based decision making related to water management and planning.

The CNRA Monitoring and Stewardship Unit named the California Natural Resource Agency Open Data Platform (ODP) as a hub able to store data from across the state and provide a single location making information findable, accessible, interoperable,

and reusable (FAIR) (*Making Public Data FAIR*, 2018). The department also noted the importance of keywords, search ability, the reduction of data errors, and improved analytical uses, again making this data FAIR (*MONITORING AND EVALUATION OF CALIFORNIA NATURAL RESOURCES AGENCY INVESTMENTS*, 2020, p. 22).

The California Natural Resources Agency Open Data Platform (ODP) now supports 19,828 datasets, 9 topics including Oceans, Water, Wildlife, Land Management, Energy, Conservation, Climate, and Natural Hazards, shares videos to train data producers and users, and works with 9 other state organizations to keep FAIR data (*The California Department of Water Resources*, ODP).

After outlining the use of the Open Data Platform, the MSU focused on stakeholder outreach and participation. Stakeholders included program managers, academics, non-profit organizations, science professionals, and technical staff. The MSU decided to work with stakeholders through a series of workshops and various group meetings. 7 questions were determined to guide stakeholder efforts:

1. What data are necessary for assessing project performance?
2. What data do programs have the capacity and expertise to reasonably collect?
3. Should all bond-funded projects be tracked? Or should a subset (i.e. 10%) be tracked based on risk-based selection criteria?
4. Who should be responsible for conducting the monitoring at a specific site (grantee, third party or program staff)?
5. Who should be responsible for entering data into the system?
6. What datasets already exist that can help answer resources management questions?
7. What functions and analytical capabilities should the system serve?

(*MONITORING AND EVALUATION OF CALIFORNIA NATURAL RESOURCES AGENCY INVESTMENTS*, 2020, p. 24)

New Mexico

In response to the 2019 Water Data Act NM House Bill 651, New Mexico created the Water Data Initiative (WDI). Their current effort includes 133 datasets from 49 sources on the New Mexico Water Data hub. Hub categories include: Climate, Ecosystems & Wildlife, Energy, Infrastructure, Natural hazards, Water Use, Water Quantity, Water Quality and Water planning (*About Us*, Data).

To continue building the hub, administrators have planned a "federated data model" (*About Us*, How we're doing it) or a Type A hub (*What Is an Internet of Water Data Hub?*, 2022). This type of hub allows producers to keep ownership of data and be responsible for standardization. The New Mexico hub administrators are working to format data to Application Programming Interfaces (APIs) in order to provide interoperable and accessible service (*About Us*, How we're doing it).

Colorado

The Colorado Decision Support System (CDSS) Data & Tools and the Colorado Information MarketPlace (CIM) are two ways to access Colorado water data. Provided by Colorado's Division of Water Resources under the Department of Natural Resources, these tools are stored on a platform called the HydroBase (*Data & Information | Division of Water Resources*, n.d.). Online Tools, DWR Web Services, Map Viewer, and Attach Locally are also topics hosted on the HydroBase to provide information and tools for "water rights, structures, diversions, irrigated lands, streamflow, and other data" (*HydroBase*, n.d.).

The CDSS Data & Tools centralize datasets including Administrative Calls, Climate Stations, Dam Safety, Diversion Records, Groundwater (Water Levels/Logs), Stations (Streamflows/Ditch Diversions), Water Rights, and Well Permits (*Colorado's Decision Support System*, n.d.). These datasets lead to a variety of historical scanned records to daily updated water collection information. The CIM shows 349 datasets related to water data.

Colorado worked with the Open Water Foundation in order to build the HydoBase and many of the datasets stored there. The Open Water Foundation works with organization and government agencies to develop, maintain, and improve open source water data tools (*Open Water Foundation*, n.d.). The foundation also reaches out to students at the Colorado Water Institute (CWI) and other universities in order to continually enhance CDSS features (*Open Water Foundation*, n.d.).

The United States Geological Survey (USGS)

The United States Geological Survey maintains the nations largest water data system network (Blodgett et al., 2016, Introduction). In an effort to contribute to the Open Water Data Initiative (OWDI), the USGS offers several water systems including the National Water Information System, the National Groundwater Monitoring Network, the Water Quality Portal, the Water Use and Availability Data Systems, the Watershed Modeling Data Systems,

The National Water Information System (NWIS) began as a collection of regional databases, thus depending on state collections. Efforts since around 2014 have been made to combine datasets into a single platform have been attempted. One challenge is the difference between state boundaries and hydrologic boundaries. This issue leads to duplicate data with conflicting information. However, the NWIS has since worked to implement standards and feedback to standardize the process of using and producing such data (Blodgett et al., 2016, NWIS).

Although some groundwater information is found on the NWIS, in 2007 a National Ground-Water Monitoring Network (NGWMN) was designed to house any available groundwater data across the nation (Blodgett et al., 2016, NGWMN). This information is now shared on a map-base graphical user interface, as a web service data portal, a "Open Geospatial Consortium (OGC) - Sensor Observation Service (OGC-SOS)" providing a web feature service. These services work to provide interoperability for water data users and providers (Blodgett et al., 2016, NGWMN). Due to this design the NGWMN is referred to as a 'hub' with four parts: the web application, the data mediator, and basic site information. Then the web services from providers work as 'spokes' out from the 'hub,' which allow users to request information from the application registry and received it in a standard format via the data mediator (Blodgett et al., 2016, NGWMN).

This model does present some challenges. As the data is stored on owner's sites, problems may arise if the internet connection is compromised or if the owner changes the data format. These issued have been addressed by creating a cache that is updated frequently and can provide information if there is a problem in the normal work flow order. The NGWMN is currently working to update web languages and offer additional features including lithologic and geologic vocabulary (Blodgett et al., 2016, NGWMN)

The USGS in partnership with the Environmental Protection Agency (EPA) agreed to work together to aggregate water-quality data in 2004. This new working group is called the National Water Quality Monitoring Council.

In 2004, leaders from the EPA and USGS signed a memorandum of understanding to create a tool to integrate and serve water-quality data from both agencies under the aegis of the National Water Quality Monitoring Council. The Water Quality Portal was created with standardized vocabulary between water samples and the EPA supported Water Quality Exchange and the USGS NWIS systems. This allows for users to find multiple options for one query, such as 'river' and 'creek' in response to a search for 'stream' (Blodgett et al., 2016, WQP). In 2014, the US Department of Agriculture's data

system was added to the WQP and presented multiple challenges as platform use rose significantly. While updates are being made, the WQP essentially operate as a cache that is regularly updated (Blodgett et al., 2016, WQP).

In the past, water use and future use predictions by county have been reported every 5 years through the USGS and stored on the NWIS web interface (Blodgett et al., 2016, Water Use and Availability Data Systems). However, “Site-specific water-use estimation” is now more preferable (Blodgett et al., 2016,Water Use and Availability Data Systems). In order to produce this type of information, the USGS is working on the National Water Census platform which will first focus on precipitation and evapotranspiration data. This type of data can be analyzed through the water budget to create use and predictions models. At this time, site- specific data has not been available and search options are minimal. The USGS is working toward these more FAIR options.

It is also important to consider soils, land use, and other spatial data when modeling water-resources (Blodgett et al., 2016, Watershed Modeling Data Systems). The USGS offers the Geo Data Portal (GDP) and the SPAtially Referenced Regressions On Watersheds (SPARROW) Decision Support System to provide data for building and running new and pre-existing models. These systems have a catalog of data sets and data and metadata standards which allow interoperability between other networks as well (Blodgett et al., 2016, Watershed Modeling Data Systems). One limitation on these systems is the complexity of file sizes. Large spacial areas of data are difficult to use for smaller references. Another limitation is the large amount of computer power to run some models. However, these systems also preform by running USGS data with a large archive of compatible data, which allows for data-integrated actions (Blodgett et al., 2016, Watershed Modeling Data Systems).

Internet of Water

The Internet of Water promotes 4 water data hub types built on the principles of data that is inventoried, standardized, and detailed. Geoconnex is the tool used to relate data and metadata and proving organization throughout (White, 2021). The 4 water data hub types are Hub Type A: Distributed, Hub Type B: Blended, Producers Push Data to Hubs, Hub Type C: Blended, Hubs Pull Data from Producers, and Hub Type D: Centralized (*What Is an Internet of Water Data Hub?*, 2022). Hub Type A requires that data is standardized and stored locally by producers. A metadata catalog allows for search queries. Hub Type B requires that producers standardize data and then share ownership with hub managers. Data is stored centrally. Hub Type C requires that hub managers collect and publish data. Hub Type D requires hub managers to collect and

standardize data to be stored on the hub with catalogged metadata (*What Is an Internet of Water Data Hub?*, 2022).

The Internet of Water is also building a tool called Hubkit to collect data, create standardization, and publish data (*What Is an Internet of Water Data Hub?*, 2022). Combined with Geoconnex, water data on these types of hub platforms will be more searchable and accessible.

The City of Austin

The City of Austin provides public data through it's Open Data Portal. Data can be found by topic or through the Data Catalog. The catalog has 12 Categories including city operations and also Environment. The catalog also filters for View Types like charts, datasets, and maps, as well as Departments, Tags, and Federated Domains. These features make it simple to find water related resources. Without further details on how data is managed or stored, the website refers to it's provider "Tyler Technologies" as the software company powering this open data technology (*Open Data Portal*, About Us).

Lower Colorado River Authority (LCRA)

The Lower Colorado River Authority in cooperation with the Bureau of Reclamation operates hydrologic and meteorologic monitoring stations (HydroMet). This network of stations collects data remotely and transfers the data through satellite to produce real-time data (*Reclamation - Missouri Basin and Arkansas-Rio Grande-Texas Gulf - HydroMet*, 2021) . This data can use found on the LCRA HydroMet platform and includes 275+ remotely operated river and weather gauges (*LCRA Hydromet*, n.d.). The data collected can be used to forecast streamflow and runoff events to inform water managers.

Texas Advanced Computer Center (TACC)

The Texas Advanced Computer Center (TACC) offers a range of data services. TACC operates high-speed/density data systems with large computational systems that allow greater amounts of data to be analyzed (*Getting Started - Texas Advanced Computing Center*, n.d.). The center creates tools to make data accessible on open platforms and new interfaces. These options allow water data procedures and users to access this stored data (*Advanced Computing and Water Management - AAAS 2018 Podcast - Latest News - Texas Advanced Computing Center*, 2018).

The First Street Foundation - Flood Factor

The First Street Foundation - Flood Factor builds environmental models using data provided through different agencies, by cleaning it and formatting it to run various operations (*Models and Methodology*, n.d.). Within the data, digital object identifiers (DOI) have been placed to support search functions and build interoperable relationships between datasets. Another important tool the foundation implements is a data dictionary. The data dictionary categorizes data making it more identifiable and useable. These categories include metadata like the DOI, location, census information, short and long term averages and future predictions, and influential climate alterations (First Street Factor, Documentation).

Having reviewed the work of Subject Matter Experts, the Texas Water Data Hub team could then implement examples and create areas of state interest. The following section will detail the methodology followed to meet the three research goals: 1) to create updated and enhanced accessibility and quality standards based on historical research for water datasets and metadata; 2) to update the definitions and relationships that describe water types in order to organize them ontologically; and 3) to highlight best methodologies for stakeholder engagement. These goals lead to solving the problem of how to build a Texas Water Data Hub which includes understanding past and current water data, updating water data categories, and involve stakeholder outreach and perspective.

3. Methodology

In order to build a Texas Water Data Hub, historic and current data standards needed to be assessed as a starting point. Much of this information could be assessed quantitatively based on the number of involved organizations, use of traditional water categories, or details of metadata. The needs of data providers and users were also analyzed qualitatively through interviews and meetings. This information provided a foundation for new organizational ideas and solutions regarding the fractured water data landscape. Using professional judgment to interpret the collected information, new water data categories could be assigned, relationships established and a new Texas Water Data Hub could be created.

3.1 Water data categories and ontology

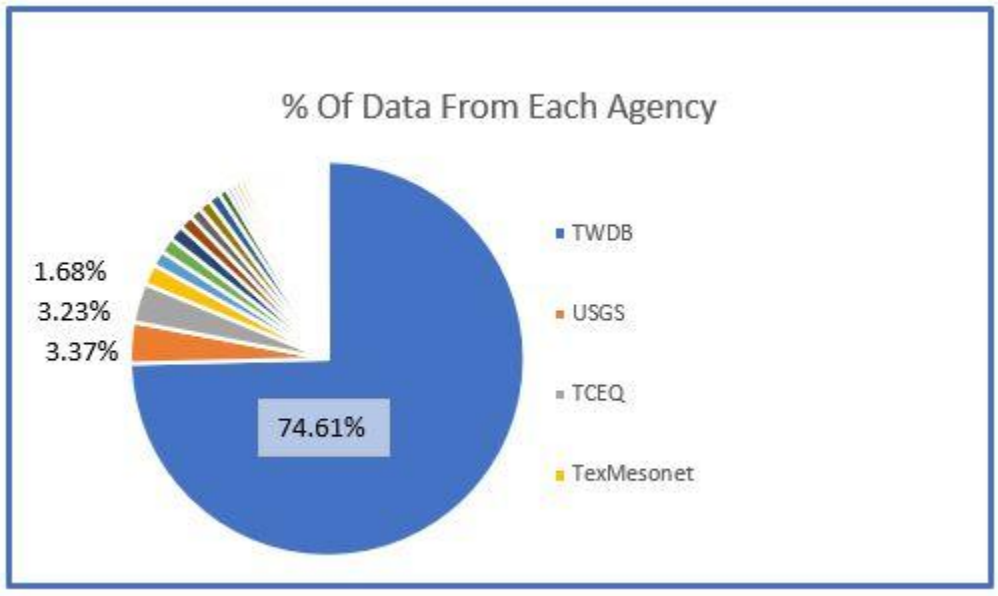
Water data categories refers to the type of water data information collected. One might call stream flow station data “Surface water” or the conditions of a bridge crossing from

mainland to an island “Infrastructure.” The accuracy of these categories is important because it allows those providing data or using it to be organized, and make information “findable.” After establishing the category a set of data pertains to, relationships to other data sets can be established through an ontology. An example might be the relationship between stream flow and lake level. Both could be categorized as “Surface water” and share an ontological relationship through detailed properties found in the data.

3.1a: Water data categories

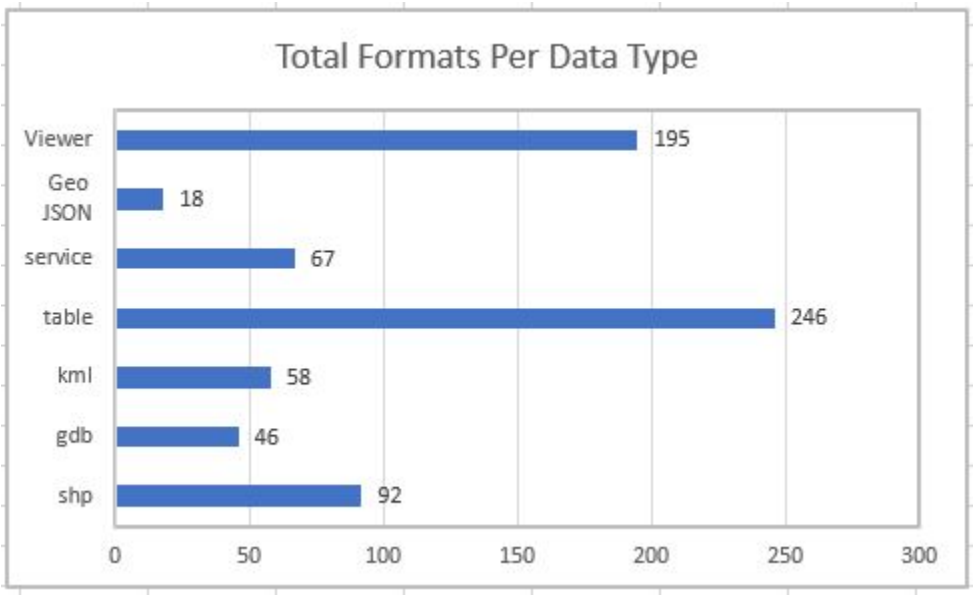
To analyze data categories, it was necessary to first research different providers and platforms where water data is currently held. Over several months, water data was inventoried from 52 agencies that held information pertinent to Texas users. For a full list of providers and platforms see **Appendix B**.

Figure 1: Pie chart of agencies providing public Texas water data originally indexed from 2019-2021



Much of the data collected from the 52 agencies came in varying formats. From the initial inventory, data contained none, one, or more of the following formats:

Figure 2: Different data formats recorded on initial inventory



10 water data categories were created based on information from the 52 agencies contributing water data in Texas. These categories were: Ground water, Surface water, Natural Hazards, Regulatory and Compliance, Boundaries and Infrastructure, Environmental and Climate, Coastal, Water Use, Water Quality, Planning.

Table 3.1

Historic Water Data Categories	
Groundwater	Boundaries and Infrastructure
Surface Water	Environmental and Climate
Planning	Coastal
Regulatory and Compliance	Water Use
Natural Hazards	Water Quality

Several steps were taken to update categories. Texas Water Data Hub planning members first looked for examples from other state hubs and water agencies including: California, New Mexico, Colorado, United States Geologic Survey (USGS), Internet of Water, City of Austin, Lower Colorado River Authority (LCRA), Texas Advanced Computer Center (TACC) and The First Street Foundation - Flood Factor (information provided from the Subject Matter Expert Research Plan, Learnings, and report, see **Appendix C**). **Table 3.2** lists findings from additional examples. These Subject Matter

Experts were consulted and their data platforms have been described in the **Literature Review.**

Table 3.2

Additional Categories Learn by Examples	
Water Quantity	Safety
Administrative Calls and Analysis	Energy
Ecosystems and Wildlife	Public Supply
Thermoelectric Power	Irrigation
Mining	Livestock
Industrial	Aquaculture

(*Austin Water*), (*California Natural Resources Agency*), (*Colorado's Decision Support System*), (*Hubs*), (*Lower Colorado River Authority*), (*New Mexico Water Data*),
(*Texas Advanced Computing Center: The University of Texas at Austin*), (*USGS Water-Use Data Downloads*, 2018), (*What's Your Flood Factor?*)

Final Categories selected for the Texas Water Data Hub can be found in section **4. Results, Table 4.1**.

3.1b: Ontology

In order to allow data to be more findable and interoperable, an ontological framework was considered for the Texas Water Data Hub. Two programs, Protege and Karma were considered in this research. Additionally, "Ontolopoloosa" was held in order to educate decision makers invested in the Texas Water Data Hub (see **Appendix D** for sample meeting notes).

Protege allows programers to build relationships between data in order to apply filters and upscale search options (The Board of Trustees of the Leland Stanford Junior University). In an effort to discover the validity of using Protege to work with Texas water data, two Protege example ontologies were taken. "The People_Example Ontology" was created to view and handle a simple ontology process and be an introduction to the framework. Several web languages like OWL, SWRL, SPRQL, and SHACL are also presented (DeBellis, 2021). The "New Protege Pizza Ontology" provided a more in-depth use of Protege and allowed users to go through a more realistic process of creating an oncology. This example also includes practice with web

semantics SWRL, SPARQL, and SHACL as well as IRI concepts and namespaces (DeBellis, 2021).

Several meetings were held to educate and practice ontological frameworks. These meetings were referred to as “Ontolopalooza” and included members from the Texas Water Developement Board (TWDB), the Center for Space Research (CSR), the Texas Community Health And Resource Management team, the Texas Integrated Flooding Framework (TWDB), and research scientist Dr. Deborah Khider from the University of Southern California. Dr. Deborah Khider instructed these meetings as water data decision leaders worked together to implement ontological frameworks into what would become the Texas Water Data Hub and other important data sites.

The Karma Data Integration Tool was also researched. This tool can be used to map datasets to ontologies. In this case, Protege would be used to create the ontology, the information would be exported to a RDF syntax (TURTLE is recommended if using Karma) and then Karma would be used to map it (Goel et al.). This is a function also available in Protege.

The effort to make data findable and interoperable has been addressed through several technical models. The **Literature Review** shares some of these models created by Subject Matter Experts from various organizations. Building in an ontology to render data findable and interoperable is an important part of addressing the research problem of how to build a water data hub. “Ontolopalooza” and meetings held by stakeholders to learn more about building data relationships is a step in achieving the goal of building an ontology on the Texas Water Data Hub. Considering the large task of learning and implementing such programing, the hub team leaned on the advice of stakeholder research as discussed in section **4. Results, 4.1 Water data categories and ontology.**

3.2 Stakeholder engagement

Stakeholders are an important part of any project because of their added perspective, experience, investment, skillset, and other valuable assets. Involving stakeholders also helped in answering the problem of how to build a Texas Water Data Hub by gleanings from the understanding and needs of various professionals, academics, and organizations. Additionally, this section addresses the third research objective to highlight best methodologies for stakeholder engagement.

In order to build a water data hub for Texas, stakeholder investment would need to come from both sides of data; the provider of data and the user. In order to find

stakeholders, understand their data landscape and supportive needs, and bring them together to build FAIR water data infrastructure, this research highlights best methodologies for stakeholder engagement. Texas Water Data Hub decision makers worked to establish best practices by holding conferences, workshops and meetings to engage state agencies as well as private and public stakeholders.This effort took place in two phases: Stakeholder research and User/Design.

3.2a Stakeholder research

This phase identified key stakeholders by compiling data providers and users. Individuals and teams of experts were consulted during the research phase of stakeholder involvement in order to organize important goals to achieve throughout the process of building the hub.

From the possible statewide group of key stakeholders, 4 groups were asked to provide and use water data in an initial test for data inclusion on the Hub. The Texas Water Development Board Springs Team, the TXMesonet team, the Texas Commission on Environmental Quality and a USGS Wells Data team supplied water data that would serve as an example of the standard for data and metadata ideally included on the Hub. These standards would allow for interoperable and reusable data as new data sets were added. Metadata standards include: Organization, Resource type, File location, Collection method, Category, Primary and secondary tags, Supporting URL, Spatial coverage, Data range, and Dataset download size (The Texas Water Development Board, 2022, slide 20, **Appendix E** shares categories and metadata standards).

The Texas Water Data Initiative Advisory Committee(TWDI), a group of interested stakeholders selected from government agencies, academia, environmental groups, and others by the Cynthia and George Mitchell Foundation, also provided valuable insight during the research phase. With the mission of the group to “advance collaboration, sharing, and use of FAIR (findable, accessible, interoperable, and reusable) water data in Texas, and a vision where “anyone can easily find, access, and use water data to inform decisions that improve water outcomes for Texas,” the committee provided a group of water data providers and users that influenced further stakeholder engagement (The Texas Water Development Board, 2022, slides 23-30, **Appendix E**).

Subject matter experts that contributed to **Table 1.2 Additional Categories Learned by Examples** and whose work was shared in the **Literature Review**, were also interviewed in order to learn from their implementation experience. Interviews were held over the course of several weeks and through virtual meetings. Compiling learning

goals from the experience of these experts, the Texas Water Data Hub team focused on continuing efforts to develop and understand these 5 points:

1. The process used to create data hubs including timelines and resources
2. Drivers behind hub creation, such as legislation, user demands, etc.
3. Mistakes to avoid and best guiding practices
4. Approaches to legacy data, data standards and standardization
5. Future considerations for inclusion on the Hub

(Information provided from the SME Learnings report, **Appendix C**)

3.2b User needs and Design

In order to better understand data users and producers who would benefit and contribute to the Hub, this research included 11 user research interviews (Hermitte, 2021). See **3.3 Subjects** below for more details on the selection of interviewees. These interviews provided insight into user behaviors with water data, including access, workflow, and sharing (The Texas Water Development Board, n.d., slide 6). Qualitative user research included 4 fundamental questions which were influenced by those asked by Subject Matter Experts from the California Natural Resources Agency Monitoring and Stewardship Unit (MSU) included in the **Literature Review**:

1. How do people search for and find data and information they need?
2. How do people use (water) data?
3. What is important to people when evaluating a data source?
4. How do producers update and share data?

Participants were invited to engage in 3 activities including work shadowing, a data search and explanation of actions, and a card sorting game to rank data priorities. See **Appendix C** for slide deck and **Appendix F** for User Research Plan.

The above effort was made in order to address research objectives 1 and 3: the publication of the Texas Water Data Hub privy to past and current water data and stakeholder engagement. Through understanding the needs of data users, the hub can be organized and designed in an effective way.

3.3 Subjects

User research was conducted with 11 participants who meet the below criteria:

1. Be willing to sign a consent form, be audio/video recorded, and have screen shots taken
2. Open to virtually meeting from their place of work or home and sharing their environment with the interviewers

Participants included 5+ data users, 5+ data producers, 3+ work with water data specifically, 5+ from identified contributing agencies (TWDB, TCEQ, USGS, LCRA, BRA, SARA), 1+ legacy data users, 1+ legacy data producers.

A 3 minute Research Participant Survey was emailed to possible participants. 11 individuals were chosen from the returned surveys to match the categories specified above. See **Appendix F** for User Research Plan and Research Participant Survey link.

For the purpose of this report, a filed IRB form was not necessary as professional members of the Texas Water Development Board conducted all interviews with permissions granted through the state agency.

3.4 Equipment

Microsoft Office, Forms, Video meeting platform, Protege software

3.5 Study Procedures

Due to the nature of the hydrologic cycle, every aspect of water from precipitation to aquifer levels is connected. Water is complicated. The research involved in understanding how to build a Texas Water Data Hub was multifaceted. With the research goals of understanding past and current water data infrastructure, updating categories and applying an ontology to data, and stakeholder outreach, the efforts to publish the hub overlapped and were interconnected. 5 areas were addressed in planning: tech, data, data intern, design, and leadership. Each of these separate areas needed an allotted amount of time and energy to be carried out successfully. Thus a timeline of research and development was created and can be found in **Appendix G** for 2021-2022.

3.6 Data Analysis

The initial 718 entries comprising the initial inventory were contributed by the Internet of Water and the Meadows Center for Water and the Environment at Texas State University, the University of Texas - Center for Space Research, Texas Disaster

Information System (TDIS), and the Texas Water Developement Board. After duplicates, unsupported sites, or entries without available data were removed, 524 entries remained. Once new categories were established, entries were resorted to fit. Data was then analyzed for category, type of format, producer, and location.This analysis allowed the hub team a better understanding of providers, types of available data, and to interpret any data gaps.

4. Results

Understanding and creating a Texas Water Data Hub, which could then influence water data collections at a national level, was a process involving several series of analysis and decision making. Working through this process included results found along the way to drive new questions and answers. Some such results can be found in section **3. Methodology**. Major results are presented in the following **Results** section below.

4.1 Water data categories and ontology

A team of 7 water data experts from the Texas Water Development Board, Department of Water Science and Conservation a division of the Texas Natural Resources Information System (TNRIS), analyzed the different water categories suggested from **Table 3.1 Historical Water Data Categories** and **Table 3.2 Additional Categories Learn by Examples**. Discussions on the presentation of information in these categories also revolved around recent shifting trends in water models, stakeholder meetings and traditional industry perceptions. One such trend is the interaction between groundwater and surface water. While this interaction might someday be considered under one category, such as Water Quality, the Texas Water Data Hub team decided to include all three categories for the time being (*Surface-Water and Groundwater Interaction Science in Texas - Overview | U.S. Geological Survey*, 2018). Several categories were considered broad enough to house additional categories. For example, the “Ecosystems and Wildlife” category in **Table 3.2 Additional Categories Learned by Example** can be included in the broader category “Soil/Environment/Climate.” Ultimately, 9 categories were designed for publication on the Texas Water Data Hub.

Table 4.1

Final Categories Selected for Texas Water Data Hub	
Groundwater	Funding and Infrastructure
Surface Water	Soil, Environment, and Climate
Planning	Water Use
Natural Hazards	Water Quality
Boundaries	

These categories were then applied to the water data initial inventory and all duplicate data was filtered out. 107 keywords were also added and applied to the newly categorized data (See **Appendix H** for the full list). A final water data inventory was complied and prepared for use in ontology work. The updated categories and new keywords provide findable water data. Due to standardized categories and detailed keywords used as metadata, data is more searchable and relationships can be built in providing an ontological framework that can be mapped.

After practicing several ontological scenarios and meeting with experts who use Protege software, Texas Water Data Hub leaders decided that building an ontology within the water hub is an important part of creating FAIR data and will begin with a focus on standardized categories and metadata. Building relationships between datasets will be an ongoing process, phased in as data is added to the hub. Considering the results of stakeholder research influenced the decision to take a phased approach. See details in **4. Discussion.** See **Appendix E** for meeting details.

This result addressed the problem of updating standards and water data categories. This decision was also one of the necessary steps leading to the creation of the water data hub and involved understanding past and current water data standards and included stakeholder engagement.

4.2 Stakeholder insights

After meeting with 4 key stakeholder groups selected to pilot datasets as mentioned in section **3.2a Stakeholder Research,** metadata standards were stated and 3 different data sources were determined:

1. Simple datasets: Data that is relatively easy to include such as API data, data available in multiple formats, and data with existing published metadata.

2. Value added datasets: Datasets that inclusion in the hub would provide value to the user such as publishing a previously unavailable dataset to the public or combining relevant data together to form a unique dataset.

3. Partner datasets: Data from agencies outside of the Texas Water Development Board (TWDB). The hub team will work with the agency to aid in making this data available and more FAIR (Findable, Accessible, Interoperable and Reusable) and to develop a positive user experience from the data partners and users within the hub.

Determining the type of water data that will be included in hub will keep FAIR standards a priority. Moving forward, this information will allow the hub team to relate systemic organization to data providers and users when considering data for inclusion. These decisions were also made by consulting with the Subject Matter Experts whose work was shared in the **Literature Review.** The inclusion of simple datasets

From the 11 interviews described in **2.2b User and Design**, 19 key insights were determined.

Table 4.3

Key Insights from User and Design	
1. Data access depends on who you know	11. Standardized data provides quick access and analysis
2. Project and legacy data are siloed	12. No standardization leads to manual data clearning and formating
3. Practicioners lack a way to share work	13. Lecacy data is not a priority given lack of standarization
4. Data are accessed through many locations and methods	14. Support tools did not help without good data
5. Users expect immediate access	15. Documented workflows lack useful processes
6. Data updates and changes disrupt workflow	16. Defined roles and communication improve results
7. Consistency is only possible through formal agreements and automation	17. Data sharing requires trust between producers and users
8. Data producers lack resources to share data	18. Users want to find an download as fast as possible
9. Search bars don't work, record level searches do	19. Unneccessary information is not needed
10. Location based search is important	

These insights allowed the hub team to organize 5 design criteria statements ensuring the success of the Texas Water Data Hub:

1. Provide a central location for water data that reflects the entire Texas water landscape
2. Establish automatic and easy ways to share data and updates
3. Provide intuitive methods to efficiently search and download data

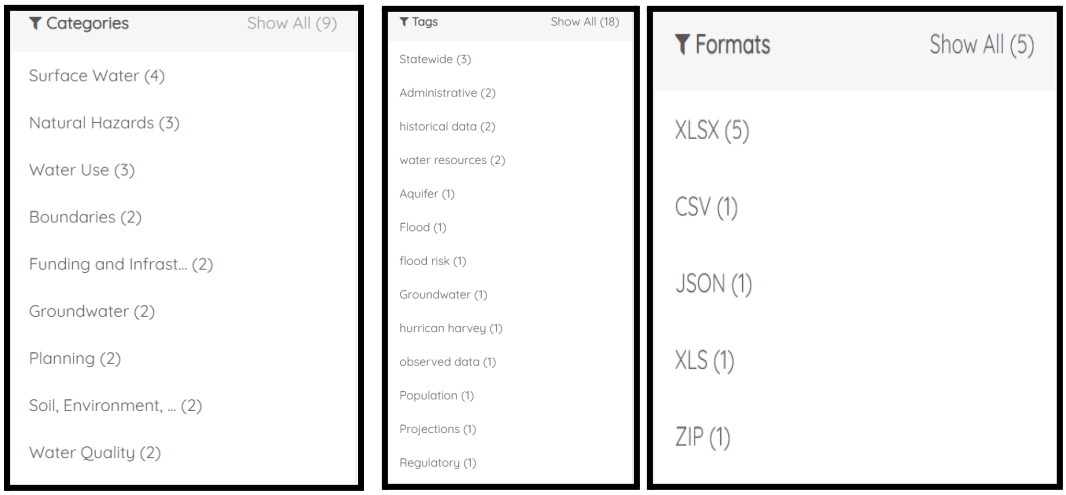
4. Emphasize clear communication and documentation to build trust and understanding
5. Assist statewide data interoperability efforts through standards and curated datasets

See **Appendix I** for slide presentation.

4.3 The Texas Water Data Hub

The Texas Water Data Hub launched Alpha format on July 1, 2022 for hub team leaders to review and share with key stakeholders for additional insight. 9 Categories, 18 key words, and 5 different formats resulted from the included datasets.

Figure 3: Alpha launch [Welcome - Texas Water Data Hub (txwaterdatahub.org)]



A soft Beta Launch was scheduled for December 5, 2022 and used an updated URL, txwaterdatahub.org. A promotional launch will follow on January 9, 2023, with public outreach.

The publication of the Texas Water Data Hub is the result of updated standards and categories, the beginning of ontological mapping, and the combined effort of stakeholders. The steps taken along the way answer the problem of how to build a water data hub and provides an example of how water data hubs can be completed at a national level.

5. Discussion

Confronting the task of building a state water data hub can be overwhelming. Data usage, and sharing, and mapping can be approach with different models and systems as shared in the **Literature Review**. While working on the many aspects of a water data hub, several processes coincide. This research broke these processes into the 3 steps of understanding past and present data collections and policy, updating and organization data, and engaging with stakeholders. Remaining flexible and able to work on several developing components is important. As new information is gathered additional results can influence the progress and lead to further updating and engagement with the water data community. Overlapping responsibilities and meaning is part of building an information hub of interconnected data.

Before beginning the creation of a water data hub, understanding the past and current landscape is helpful. Past decisions are often tied to policy that is dependent on location and availability of natural resources (Reimer, n.d., p. 2). The way stakeholders understand and talk about water data must be considered when making updates to ensure that buy-in and transitioning will occur. Consulting with Subject Matter Experts, such as those presented in the **Literature Review**, also leads to better design and understanding of what areas need to be updated.

While there is usually a large amount of water datasets to be found, each organization or collection can be so different that the data becomes difficult or impossible to use (Maidment, 2008). A good place to start in this process is gathering available datasets. Online research and stakeholder engagement is how datasets can be found. Reaching out to universities, water resource groups in the public and private sector, and environmentally focused non-profits is important stakeholder engagement that provides avenues into information that is not always FAIR.

As noted, stakeholders play a vital role in building a water data hub. Support in finding, updating, and sharing data is one such role and providing insights into needs and recommendations is another. After meeting with Subject Matter Experts (see **Appendix C**), the Texas Water Data Hub team gathered 6 key takeaways to keep in mind throughout the project:

1. Add value through problem solving
2. Set clear standards and governance
3. Build with the future in mind
4. Take a phased approach
5. Empower users

6. Make it a community effort

These ideas are important as each of the attendees in the meetings have aided in the development of new data hubs that meet FAIR standards. Taking a phased approach became an especially important guideline regarding building ontological relationships within data sets. Using a program such as Protege takes time to learn and understand and hours of buildout. To avoid exhausting hub resources in this specific area, the decision was made to slowly work on building in the ontology. At this time, focus will be directed on metadata with the inclusion of topics listed in **3.2a Stakeholder Research** such as Organization, Resource type, File location, Collection method, Category, Primary and secondary tags, Supporting URL, Spatial coverage, Data range, and Dataset download size (see **Appendix E**).

Since water data platforms at the federal level are driven by data collected at the state level, state wide efforts need to establish a FAIR water data landscape (Josset et al., 2019). If states can organize findable, accessible, interoperable, and reusable water data into statewide hubs, then the ability to take that FAIR data to a national scale would require fewer modifications. The work to standardize and share water data is a necessary process to renovating national water infrastructure.

6. Conclusion

It may be too simplistic to say that by building state water data hubs, a national hub would be an easy endeavor. Water and all of the ways that it is used and moves, is rarely an easy subject matter. In addition to finding and making water data accessible to the public, the national hub would need to address water data issues that cross state boundaries. Allowing each state to update and create water data hubs as they see fit will lead to interoperability errors. As several states have begun this process, the national effort must quickly follow. Foundational ideas have been documented through the work of many invested organizations which can now be scaled up in order to provide a strong water data infrastructure across the nation. Due to the variety of data and publishing practices, providing national standards and framework would serve to guide state efforts. It is recommended that national work draws from the knowledge of current industry leaders to update this aging system.

In order to recognize data gaps, it is important to consider small details and large picture issues. Water data around the world and in the United States has been described as “a new endangered species” (Vörösmarty et al., 2014). USGS reports have shown declining numbers of long-record streamgages in the United States starting in 1968 and falling by as many as 100 river gages each year (Lanfear & Hirsch, 1999,

1). Without the funding and ability to collect data in the first place, there is no data to organize, analyze or share.

It is also important to understand the interconnectedness of water, energy, food and demographics. These separate but related industries and issues must also be considered together. The overlap in these areas is considered a "nexus." For example, the "energy-water nexus" is the important use of water to cool thermoelectric power plants or in fuel refinement and energy needed for water treatment and transportation (*The Water-Energy Nexus: An Earth Science Perspective | U.S. Geological Survey*) (Chini & Stillwell, 2017). Currently, very little data is collected or analyzed on this level. The ability to analyze data across various sectors, including water, will also contribute to the wise use of water.

A final consideration is the advancement of new technologies in the water industry such as desalinization plants, water reuse, and even remote sensing. As the United States begins to focus on updating water data infrastructure, making sure to engage these new platforms will be critical in providing a way for water managers to make pivotal decisions for the future of this sector. Federal governance over such national water data could guide new policy, thus reaching out to state-level managers as well (Josset et al., 2019). Support at the national level to include data from these new technologies will build a functional model of modern infrastructure.

7. Summary

This research addressed the problem of how to build a Texas Water Data Hub and document the experience in order to provide an example of processes that can be scaled up to a national level. The 3 main objectives of this research included learning from past and current water data standards, updating categories and developing an ontology between datasets, and implementing stakeholder engagement.

Each step in the process of publishing the Texas Water Data Hub contributed to answering the problem of how to build such a platform. Understanding that there are many ways to technically design such a space came from consulting with Subject Matter Experts via meetings and interviews. The **Literature Review** also outlined many of the ways data can be published. Having these models and examples allowed the Texas team to determine which options would be most effective given the allotted resources.

Understanding the historical and current water data landscape also allowed the Texas team to update data and metadata standards following the FAIR (findable, accessible,

interoperable, and reusable) ideology (*Making Public Data FAIR*, 2018). Meeting this goal also brought into consideration the development of an ontology within datasets. This is an ongoing process and will continue to develop as new datasets are brought onto the hub.

Interviewing and meeting with data producers, users, and stakeholders drove the process of technical design and provided guiding values such as those found in section **5. Discussion**. Inclusion of these groups also contributed to buy-in and support that is needed to collect and format datasets. This process is also ongoing as data continues to be organized for use through the hub.

Building the Texas Water Data Hub as a space to share findable, accessible, interoperable, and reusable (FAIR) data included understanding historical and current water policy, updating data and metadata standards, and engaging with stakeholders. Each completed step led to the publications of the Texas Water Data Hub. This state model can be used as an example for other states and efforts made at a national level. Ultimately, updated water infrastructure will provide decision makers with information needed to take sustainable actions while protecting this important resource.

References

About TWDB | Texas Water Development Board. (n.d.). Texas Water Development Board. Retrieved October 30, 2022, from https://www.twdb.texas.gov/about/index.asp

About Us. (n.d.). New Mexico Water Data. Retrieved December 14, 2022, from https://newmexicowaterdata.org/about/#how

Advanced Computing and Water Management - AAAS 2018 Podcast - Latest News - Texas Advanced Computing Center. (2018, February 15). Texas Advanced Computing Center. Retrieved December 14, 2022, from https://www.tacc.utexas.edu/-/advanced-computing-and-water-management-aaas-2018-podcast

Austin Water. (n.d.). AustinTexas.gov. Retrieved November 15, 2022, from https://www.austintexas.gov/content/austin-water

Blodgett, D., Lucido, J., & Kreft, J. (2016, March 17). Progress on water data integration and distribution: a summary of select US Geological Survey data systems. *The Journal of Hydroinformatics*, *18*(2), 226-237. https://iwaponline.com/jh/article/18/2/226/24/Progress-on-water-data-integration-and

The Board of Trustees of the Leland Stanford Junior University. (n.d.). protégé. Retrieved November 10, 2022, from https://protege.stanford.edu/

Building Data Infrastructure. (2022, January 5). The Internet of Water Coalition. Retrieved October 16, 2022, from https://internetofwater.org/

The California Department of Water Resources. (n.d.). California Natural Resources Agency Open Data: Welcome. Retrieved December 11, 2022, from https://data.cnra.ca.gov/

California Natural Resources Agency. (n.d.). California Natural Resources Agency Open Data: Welcome. Retrieved November 15, 2022, from https://data.cnra.ca.gov/

Colorados Decision Support System. (n.d.). CDSS Data & Tools. Retrieved December 11, 2022, from https://dwr.state.co.us/Tools

Chini, C. M., & Stillwell, A. S. (2017, November 7). *Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database*. Journal of Water Resources Planning and Management. Retrieved October 26, 2022, from https://ascelibrary.org/doi/full/10.1061/(ASCE)WR.1943-5452.0000739

Colorado's Decision Support System. (n.d.). CDSS Data & Tools. Retrieved November 15, 2022, from https://dwr.state.co.us/Tools

Cox, M., Tinsley, W. E., Potts, M., Shurbet, M., Gilmore, R. B., Pitts, G., & Boswell, H. B. (1968, November). *1968 THE TEXAS WATER PLAN Summary* [THE TEXAS WATER DEVELOPMENT BOARD]. 1968 State Water Plan. https://www.twdb.texas.gov/waterplanning/swp/1968/index.asp

Data & Information | Division of Water Resources. (n.d.). Colorado Division of Water Resources. Retrieved December 11, 2022, from https://dwr.colorado.gov/services/data-information

DeBellis, M. (2021, September 30). *New Protégé Pizza Tutorial*. Michael DeBellis. Retrieved November 10, 2022, from https://www.michaeldebellis.com/post/new-protege-pizza-tutorial

DeBellis, M. (2021, October 22). *The People_Example Ontology*. Michael DeBellis. Retrieved November 10, 2022, from https://www.michaeldebellis.com/post/the-people_example-ontology

EPA. (n.d.). Water Quality Data Home. Retrieved November 9, 2022, from https://www.waterqualitydata.us/#countrycode=US&statecode=US%3A01&countycode=US%3A01%3A083&siteType=Well&startDateLo=09-22-1981&startDateHi=09-22-2020&sampleMedia=water&sampleMedia=Soil&characteristicType=Microbiological&characteristicType=Inorganics%2C%20

Getting Started - Texas Advanced Computing Center. (n.d.). Texas Advanced Computing Center. Retrieved December 14, 2022, from https://www.tacc.utexas.edu/use-tacc/getting-started/#data

Getting Started with First Street Data. (n.d.). First Street Foundation. Retrieved December 15, 2022, from https://firststreet.org/data-access/getting-started-with-first-street-data/

Goel, A., Knoblock, C. A., & Lerman, K. (n.d.). *Karma: A Data Integration Tool*. Center on Knowledge Graphs. Retrieved November 16, 2022, from https://usc-isi-i2.github.io/karma/

Hermitte, S. (2021, September 15). *Building the Texas Water Data Hub from the Ground Up*. The Cynthia and George Mitchell Foundation. Retrieved November 17, 2022, from https://www.cgmf.org/blog-entry/457/Building-the-Texas-Water-Data-Hub-from-the-Ground-Up.html

Hubs. (n.d.). Internet of Water. Retrieved November 15, 2022, from https://internetofwater.org/resources/hubs/

HydroBase. (n.d.). Colorado's Decision Support Systems. Retrieved December 11, 2022, from https://cdss.colorado.gov/software/hydrobase

Josset, L., Allaire, M., Hayek, C., Rising, J., Thomas, C., & Lall, U. (2019, March 1). *Earth's Future: The U.S. Water Data Gap—A Survey of State-Level Water Data Platforms to Inform the Development of a National Water Portal*. AGU Advancing Earth and Space Science. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018EF001063

Khider, D., Emile-Geay, J., McKay, N.P., Gil, Y., Garijo, D., Ratnakar, V., Alonso-Garcia, M., Bertrand, S., Bothe, O., Brewer, P., Bunn, A., Chevalier, M., Comas-Bru, L., Csank, A., Dassié, E., DeLong, K., Felis, T., Francus, P., Frappier, A., Gray, W., Goring, S., Jonkers,... and Zhou, Y. (2019), PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data. Paleoceanography and Paleoclimatology, 34: 1570-1596. https://doi.org/10.1029/2019PA003632

Lanfear, K. J., & Hirsch, R. M. (1999, December 19). USGS Study reveals a decline in long-record streamgages. *Eos, Transactions American Geophysical Union*, *80*(50), 605-607.

LCRA Hydromet. (n.d.). Hydromet - Map View - Stream Flow and Weather. Retrieved December 14, 2022, from https://hydromet.lcra.org/

Lower Colorado River Authority. (n.d.). Map View - Stream Flow and Weather. Retrieved November 15, 2022, from https://hydromet.lcra.org/

Maidment, D. R. (2008, March 1). *Brining Water Data Together*. Journal of Water Resources Planning and Management. Retrieved October 26, 2022, from https://ascelibrary.org/doi/full/10.1061/(ASCE)0733-9496(2008)134:2(95)

Making public data FAIR. (2018, November 26). Internet of Water. Retrieved November 9, 2022, from https://internetofwater.org/valuing-data/making-public-data-fair/

Manford, D., Dixon, R. M., & Dent, O. F. (1961, May). *A PLAN FOR MEETING THE 1980 WATER REQUIREMENTS OF TEX* [TEXAS BOARD OF WATER ENGINEERS]. https://www.twdb.texas.gov/waterplanning/swp/1961/index.asp

Models and Methodology. (n.d.). First Street Foundation. Retrieved December 14, 2022, from https://firststreet.org/methodology/

MONITORING AND EVALUATION OF CALIFORNIA NATURAL RESOURCES AGENCY INVESTMENTS. (2020, February 27). California Natural Resources Agency. Retrieved December 11, 2022, from https://resources.ca.gov/-/media/CNRA-Website/Files/Initiatives/MSU/White-Paper-and-Letter-Final.pdf

National Water Census: Water Use | U.S. Geological Survey. (2019, March 2). USGS.gov. Retrieved November 9, 2022, from https://www.usgs.gov/mission-areas/water-resources/science/national-water-census-water-use

New Mexico Water Data. (n.d.). Water Data for New Mexico. Retrieved November 10, 2022, from https://newmexicowaterdata.org/

Open Data Portal. (n.d.). Open Data | City of Austin Texas | Open Data | City of Austin Texas. Retrieved December 14, 2022, from https://data.austintexas.gov/

Open Water Foundation. (n.d.). Open Water Foundation – open data | open software | open decisions. Retrieved December 11, 2022, from https://openwaterfoundation.org/

Press Release. (2022, May 19). *Stansbury, Heinrich, Luján, Owens, Lee, Lawmakers Introduce Landmark Bipartisan Water Security and Drought Package*. Melanie Stansbury. Retrieved October 26, 2022, from https://stansbury.house.gov/media/press-releases/stansbury-heinrich-lujan-owens-lee-lawmakers-introduce-landmark-bipartisan

Reclamation - Missouri Basin and Arkansas-Rio Grande-Texas Gulf - HydroMet. (2021, December 10). Bureau of Reclamation. Retrieved December 14, 2022, from https://www.usbr.gov/gp/hydromet/index.html

Reimer, A. (n.d.). *U.S. Water Policy: Trends and Future Directions*. Department of Agricultural Economics, Sociology, and Education. Retrieved October 1, 2022, from https://aese.psu.edu/nardep/publications/working-papers/u-s-water-policy-trends-and-future-directions

Rosen, R. A. (2017). *2016 Texas Water Roadmap Forum* [Institute for Water Resources Science and Technology, Texas A & M University - San Antonio, San Antonio, TX]. https://libguides.tamusa.edu/ld.php?content_id=28446621

Rosen, R. A., Mace, R. E., Hermitte, S. M., & Wade, R. (2020, October 29). Internet of Texas Water: Use Cases for Flood, Drought, and Surface Water–Groundwater Interactions. *Texas Water Journal*, *11*(1), 133-151. https://twj-ojs-tdl.tdl.org/twj/index.php/twj/article/view/7109/6471

Rosen, R. A., & Roberts, S. V. (2018). *Connecting Texas Water Data Workshop: Building an Internet for Water*. Water Resources Science and Technology Book and E-Book Publications and Reports.

Seton, M. (2014, August 24). *Bridge Over Alton Water*. Mark Seton Photography. http://markseton.co.uk/2014/08/28/a-bridge-over-alton-water/

Surface-Water and Groundwater Interaction Science in Texas - Overview | U.S. Geological Survey. (2018, May 1). USGS.gov. Retrieved November 15, 2022, from https://www.usgs.gov/centers/oklahoma-texas-water-science-center/science/surface-water-and-groundwater-interaction

Texas Advanced Computing Center: The University of Texas at Austin. (n.d.). Texas Advanced Computing Center: HOME. Retrieved November 15, 2022, from https://www.tacc.utexas.edu/

The Texas Water Development Board. (n.d.). *4 Master Slide Deck* [Details research to support the Water Hub] [PPTX].

Texas Water Development Board. (n.d.). *Water Data For Texas*. Water Data For Texas. Retrieved November 10, 2022, from https://www.waterdatafortexas.org/reservoirs/statewide

The Texas Water Development Board. (2022, May 18). *TX Water Data Initiative Meeting* [TWDI Meeting and Hub Update]. PPTX.

Titles - S.4236 - 117th Congress (2021-2022): Water Data and Security Act of 2022. (2022, May 25). Congress.gov. Retrieved October 17, 2022, from https://www.congress.gov/bill/117th-congress/senate-bill/4236/titles

USGS Water Data for USA. (n.d.). USGS. https://waterdata.usgs.gov/nwis

USGS water-use data downloads. (2018, June 19). Water Resources. Retrieved November 15, 2022, from https://water.usgs.gov/watuse/data/index.html

The Water-Energy Nexus: An Earth Science Perspective | U.S. Geological Survey. (n.d.). USGS.gov. Retrieved November 24, 2022, from https://www.usgs.gov/news/featured-story/water-energy-nexus-earth-science-perspective

Water Law Overview - National Agricultural Law Center. (n.d.). The National Agricultural Law Center. Retrieved October 11, 2022, from https://nationalaglawcenter.org/overview/water-law/

Water Quality Data Download | US EPA. (2022, July 14). EPA. Retrieved November 9, 2022, from https://www.epa.gov/waterdata/water-quality-data-download

Water Resource Management Strategies. (n.d.). California Department of Water Resources. Retrieved December 10, 2022, from https://water.ca.gov/Programs/California-Water-Plan/Water-Resource-Management-Strategies

What is an Internet of Water Data Hub? (2022, September). Internet of Water. Retrieved December 13, 2022, from https://internetofwater.org/blog/what-is-an-internet-of-water-data-hub/

What's Your Flood Factor? (n.d.). First Street Foundation. Retrieved November 15, 2022, from https://firststreet.org/risk-factor/flood-factor/

White, C. (2021, November). *Geoconnex: A New Way to Connect Data*. Internet of Water. Retrieved December 13, 2022, from https://internetofwater.org/blog/geoconnex/

Vörösmarty, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Döll, P., Goodison, B., Hall, A., Jenne, R., Kitaev, L., Landwehr, J., Keeler, M., Leavesley, G., Schaake, J., Strzepek, K., Sundarvel, S. S., Takeuchi, K., & Webster, F. (2001). Global water data: A newly endangered species. *Eos*, 82(5), 54–54.

Appendix A

Texas A&M University-San Antonio

Digital Commons @ Texas A&M University-San Antonio

Water Resources Science and Technology Book and E-Book Publications and Reports

Water Resources Science and Technology

2018

Connecting Texas Water Data Workshop: Building an Internet for Water

Rudolph A. Rosen
Texas A&M University-San Antonio

Susan V. Roberts
Texas Center for Applied Technology

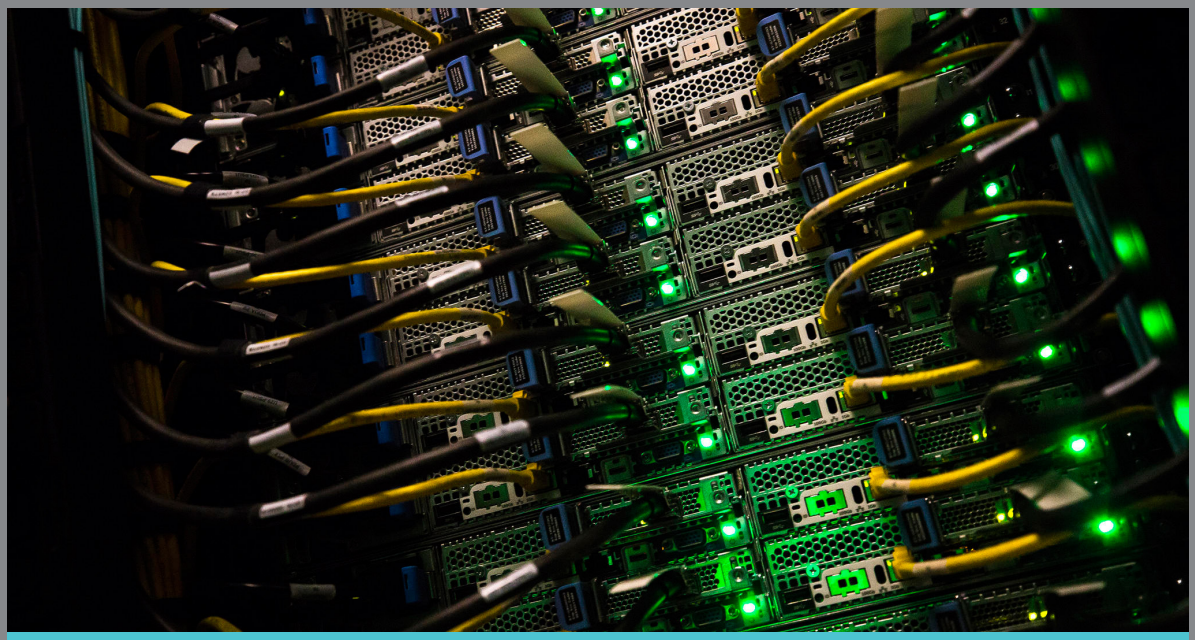
Follow this and additional works at: https://digitalcommons.tamusa.edu/water_books

 Part of the Water Resource Management Commons

Recommended Citation

Rosen, Rudolph A. and Roberts, Susan V., "Connecting Texas Water Data Workshop: Building an Internet for Water" (2018). *Water Resources Science and Technology Book and E-Book Publications and Reports*. 2.
https://digitalcommons.tamusa.edu/water_books/2

This Book is brought to you for free and open access by the Water Resources Science and Technology at Digital Commons @ Texas A&M University-San Antonio. It has been accepted for inclusion in Water Resources Science and Technology Book and E-Book Publications and Reports by an authorized administrator of Digital Commons @ Texas A&M University-San Antonio. For more information, please contact deirdre.mcdonald@tamusa.edu.



CONNECTING TEXAS WATER DATA WORKSHOP

Workshop Held
April 17, 2018

Location
Texas Advanced Computing Center
University of Texas

Sponsors and Organizers

Texas Water Development Board
The Cynthia and George Mitchell Foundation
Institute for Water Resource Science and Technology, Texas A&M University-San Antonio
Texas Advanced Computing Center, University of Texas
National Science Foundation Research Coordination Network for Climate, Energy, Environment and Engagement in Semiarid Regions

CONNECTING TEXAS WATER DATA WORKSHOP

INTRODUCTION

Texas' public and private companies, organizations, and agencies have collected water data for different purposes and at different scales for many years. These data are scattered across multiple platforms with different standards, often making important data sets inaccessible or incompatible. This leaves Texas' decision makers, industries, landowners, and communities with significant amounts of data of limited use to support real-time decision making, development of opportunities for water security, or for modeling an accurate picture of Texas' water future. To be useful in decision-making, water data must not only be open and transparent, but presented in a way that is relevant to the needs of decision makers.

On April 17, 2018, the Connecting Texas Water Data Workshop was held at the Advanced Computing Center on the J.J. Pickle Research Campus of the University of Texas in Austin. The workshop brought together almost 90 invited experts representative of Texas' government and water agencies, utilities, academia, business, industry, research institutes, and water associations and advocacy organizations.

Our goal was to engage workshop participants – all leading Texas water stakeholders – in the identification of critical data needs and in the design of a data system that facilitates access to and use of water data in Texas.

This report describes the workshop outcomes, presentations, discussions, and facilitated stakeholder sessions.

This report may be cited as: Rosen, Rudolph A. and Susan V. Roberts. 2018. Connecting Texas Water Data Workshop. Institute for Water Resources Science and Technology, Texas A&M University-San Antonio, San Antonio, TX 78224. (ISBN-13: 978-0-9986645-4-5) https://libguides.tamusa.edu/ld.php?content_id=42020932

Copies may be obtained at https://libguides.tamusa.edu/ld.php?content_id=42020932

On the Cover: Connections carrying data on the Texas Advanced Computing Center's Stampede 2, ranked the 12th most powerful supercomputer in the world. Photo by Martin do Nascimento/KUT Public Radio.

EXECUTIVE SUMMARY

CONNECTING TEXAS WATER DATA

The Connecting Texas Water Data Workshop brought together experts representative of Texas' water sectors to engage in the identification of critical water data needs and discuss the design of a data system that facilitates access to and use of water data in Texas. Participants worked in facilitated sessions to identify, describe, and list 1) who needs, 2) what data, 3) in what form, 4) to inform what decisions about water in Texas. They also worked to identify key data gaps in Texas water data, attributes of a comprehensive open access water data information system capable of informing water management decisions, and use cases or pilot projects illustrating the value of an open access, interoperable water data system.

Participants envisioned the ideal water data system for Texas as one with open access that includes an ability to obtain available water data, including raw data, metadata, and legacy data in a digitized form. The data system should be user friendly, robust, and provide real-time information using web services with source information and built-in visualization tools so that non-experts can use the system. Data and information should be free, and created and kept in consistent reporting formats so that data "talk to each other" as users search and gain access. The ideal form of data system is envisioned as consisting of several integrated data hubs specialized by water sector, with incentives for people to add new data and share existing data through the hubs. There should be adequate funding to sustain the data system over time.

Several steps to develop and promote an open water data system for Texas are recommended. Among these are developing use cases, establishing an advisory task force, designing the network structure for an open data system and hubs, identifying key users of the initial system, naming lead developers and hosts of the system, forming lines of support, and sharing information about open data experiences and best practices.

In Texas today, one needs to be an expert to find data that exist and to access those data and integrate them for practical use. Much of the data that do exist are not actionable. An open water data system for Texas is needed to support access to an accurate accounting of supply, quality, and use of water to better support decision makers in their efforts to enhance sustainable water use. Improved access, standardization, and integration of data will provide water managers and decision makers a better basis for data-driven decisions, enabling them to more confidently meet urban, agricultural, ecological, and industrial needs for water.

CONTENTS

THE WORKSHOP REPORT

Seeking better decisions about water in Texas

"The better the data, the better the science. And the better the science, the better the policy."

-- Kathleen Jackson
Texas Water Development Board

THE BEGINNING

In many parts of Texas the human population is growing rapidly, but water availability and use are affected by frequent droughts in some areas, flooding in others, and multiple human-caused events statewide. The consequences can limit economic growth, business, agriculture, and stable communities. Pressure is placed on public officials and water managers to ensure continued access to dependable safe water supplies, but too often the information needed to steward and manage water for multiple uses is either nonexistent, inaccessible, or unusable. Making better decisions about water will require more data, better data, data that can be universally used (interoperable), and access to all data.

Texas water data is housed at various state and federal agencies, water authorities and districts, local utilities, universities, and throughout the private sector. While the total constitutes considerable data, it exists in many forms, levels of resolution, degrees of temporal value, and states of accessibility and usability that range from open access and user friendly to complete inaccessibility and uselessness. Without access and usability, much of the data that potentially could be used to make better decisions about water is lost to any use.

Access to Texas' water data resources is essential if Texas is to succeed in addressing its growing calls for water conservation and increasing water demand for urban, agricultural, ecological, and industrial uses. Texas data can be made available through open data systems or hubs (see Appendix IV for glossary of terms) that enable networked access designed to be usable and relevant to the needs of data users and decision makers.

Workshop attendees were asked to offer suggestions covering a range of key attributes of an open, interoperable, interconnected, comprehensive, and user relevant data system and networked data hubs. To help organize and focus thinking, workshop participants were led through a series of exercises culminating in identification of possible use cases that may serve as models for open data systems.

This report summarizes the workshop sessions and provides extensive detail in the synthesis text and appendices. The report supports continuing dialogue among workshop participants and involvement of stakeholders who did not attend the workshop. The workshop was intended to be the beginning of an engagement process involving all water stakeholders that use or need water data, especially water decision makers.

The Connecting Texas Water Data Workshop provided an important opportunity for Texas water data experts to join together and offer input essential to improving the state of water data in Texas.

TOWARD AN INTERNET OF WATER

In much of the United States today it can be a complex and time consuming experience to learn something as simple as the safety and quality of water coming from your own tap, according to Dr. Martin Doyle of the Nicholas Institute for Environmental Policy Solutions at Duke University speaking at the workshop. Dr. Doyle's detailed comments can be found in Appendix I.

Many decisions are made today on the basis of instantly available data, but for water, which is the most important ingredient for life on Earth, access to data for most Americans is far from instant.

There is a fracturing of where water data come from and a wide range of organizations that generate and store data. According to Dr. Doyle, the US Geological Survey and associated water science centers in the various states that maintain the data and stream gauge network for the National Water Information System serve as the "gold standard" for nationwide surface water data and open access. This system presents a ready foundation and model for building a nationwide open network for public water data collected for multiple mission-specific sectors and interests such as energy, agriculture, community development, forestry, fisheries, endangered species, watersheds, and so on.

Dr. Doyle and collaborators are seeking a means to have data that come from these various sources made available and viewable on a real-time basis. This has been termed the "internet of water."

The internet of water was described during a water dialogue held by the Aspen Institute. Following the forum, a group of funders came forward to support the initiative. A dialogue series then pulled together water experts from utilities, state and federal government, oil and gas, philanthropy, academia, nongovernmental organizations, software companies, and other sectors. The result was a consensus formed by people with different perspectives around the following key findings:

- The value of open, shared, and integrated water data has not been widely quantified, documented or communicated.
- The most necessary step in using water data for sustainability is making public water data open by default, discoverable, and digitally accessible.
- Water data can be most effectively integrated through an internet of water.

Dr. Doyle offered three suggestions to create an internet of water:

1. Form a vision about how water data will be used, along with a declaration of usefulness and quantification of value.
2. Develop a series of regional pilots, or use cases, that solve real-time, real-world water management problems and demonstrate the value of water data.
3. Start an internet of water by using public water data already collected and curated.

TEXAS WORKSHOP OBJECTIVES

For Texas, basic information was collected at the workshop by focusing participant work on four key objectives:

1. To identify, describe, and list (a) who needs, (b) what data, (c) in what form, (d) to inform what decisions about water in Texas.
2. To define the desired future of water data management and access in Texas by listing data gaps, accessibility options, and key attributes of a comprehensive open access water data information system.
3. To initiate development of use cases for Texas water by identifying critical needs of Texas data providers and consumers.
4. To list ideas on next steps to further define, design, and build a water data system for Texas.

Texas water planning requires access to and use of large amounts of data from many sources, provided in ways decision makers can work with. Texas water plans look out 50 years and are updated every 5 years. The Texas regional water planning process involves more than 450 volunteers across the state representing big cities, small communities, agriculture, manufacturing, and all the other water users. The plans are data- and science-driven, and prepared cooperatively with 16 regional water planning groups. The volunteers in these groups come together to compile strategies to address future water needs and determine how much water we have today, what we need to do for tomorrow, and what strategies or projects we need to put in place to get us where we need to be in the future. We use the best data available and make it transparent and usable on multiple platforms. But in spite of all the work on water plans, we don't plan to plan, we plan to build.

With anywhere from 1,000 to 1,200 people moving to Texas every day, and not one of them bringing any water with them, we seek new supplies not just to ensure current residents have the water they need, but also to supply the needs of a growing population.

-- Kathleen Jackson
Texas Water Development Board

WHO NEEDS WHAT DATA, IN WHAT FORM, TO INFORM DECISIONS

Thousands of decisions about water are made daily in Texas. Many of these decisions use data, and many others would be made better by the decision makers having open and easy access to usable data. To help better understand the scope of who needs data to inform water decisions in Texas, and in what form the data are needed, participants were asked to make lists. They were asked, "Which data must be easily accessible and interoperable?"

Only by understanding how data are used by decision makers can future data systems be built to effectively inform decision making.

In answer to the question, "Who needs data?" the six workgroups provided over 60 different responses, ranging from "everyone" to specific water decision makers, such as the National Weather Service. The relative frequency of listing of users can be readily seen using a word cloud (Figure 1) where the size of each word indicates the frequency of mention in the reporting of the workgroups.

At the top of the list are farmers and researchers. Other groups having multiple mentions by the workgroups included planners, insurers, agencies, oil and gas industry, developers, consultants, and utilities. There are a wide variety of users of water data, ranging from users requiring highly synthesized data to users where only raw data will suffice.

Terms used by one work group to describe who needs data were sometimes different terms that point to the same users, such as the terms "General Public" and "Everyone." In other cases a description for who needs data used by one workgroup was sometimes inclusive of a description used by another workgroup, such as the broad term "Academics" and more restrictive term "Academic Researchers." In still other cases a specific category of data user was associated with a specific user group, such as "Agriculture" and then described as universally associated with all user groups by another workgroup. To help draw meaningful connections, Figure 2 displays how

Figure 1. Who needs data? Size of each word indicates the frequency of mention in the reporting of the workgroups.

many workgroups mentioned users associated with major categories of use, such as for "Agriculture," and which specific users and how many were mentioned for each category. The tie between all water users is indicated by the center circle, with different terms listed in the circle used by the six workgroups that point to "Everyone." Note that the general technical professions, "Resource Managers, Engineers, Planners, and Consultants," were mentioned as "who needs data" for virtually every use.

A complete listing of all responses by each workgroup is provided in Appendix II.

Participants in the workgroups were then asked, "What data do data users need?" (Figure 3). As with who needs data, there are many kinds of data needed. There were over 60 different answers, with some being subcategories of others. There also were several categories of needed data that were mentioned repeatedly by the workgroups. These included soil moisture, stream flow, water rights, water use, and water quality.

The next question focused on the form of data needed. While there were over 50 descriptions of the form of data needed, only two stood out. These were raw data and metadata. They were mentioned most, with many other terms used to describe various degrees of open data, accessible data, usable data, free data, and standardized data. (Figure 4)

Participants were then asked to describe the purposes for which data are most needed. There were about 50 different responses with very little overlap. A wide variety of purposes for which data are needed is not surprising given the wide-diversity of interests of participants and the situational, geographic, and temporal

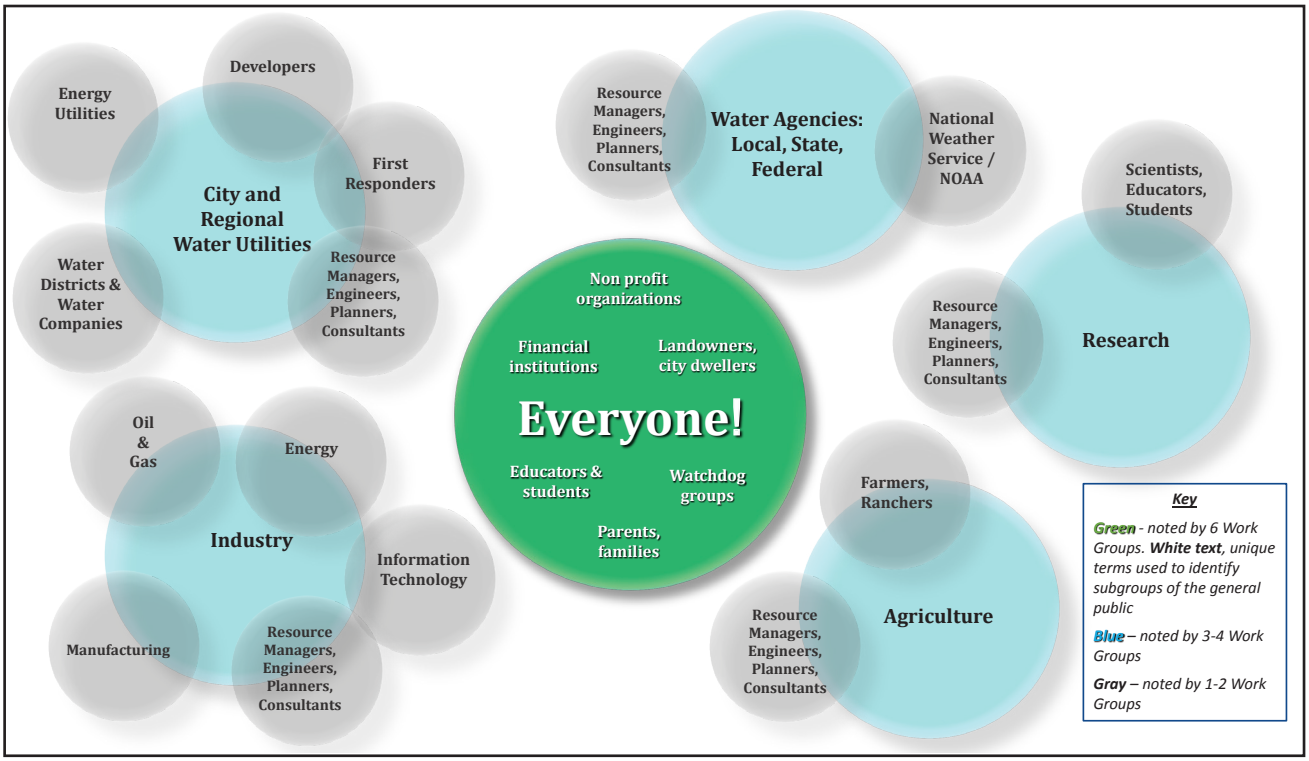


Figure 2. "Who needs data?" aggregated by users associated with major categories of use. (Large circle noted by 6 workgroups, medium by 3-4, and small by 1-2 workgroups.)

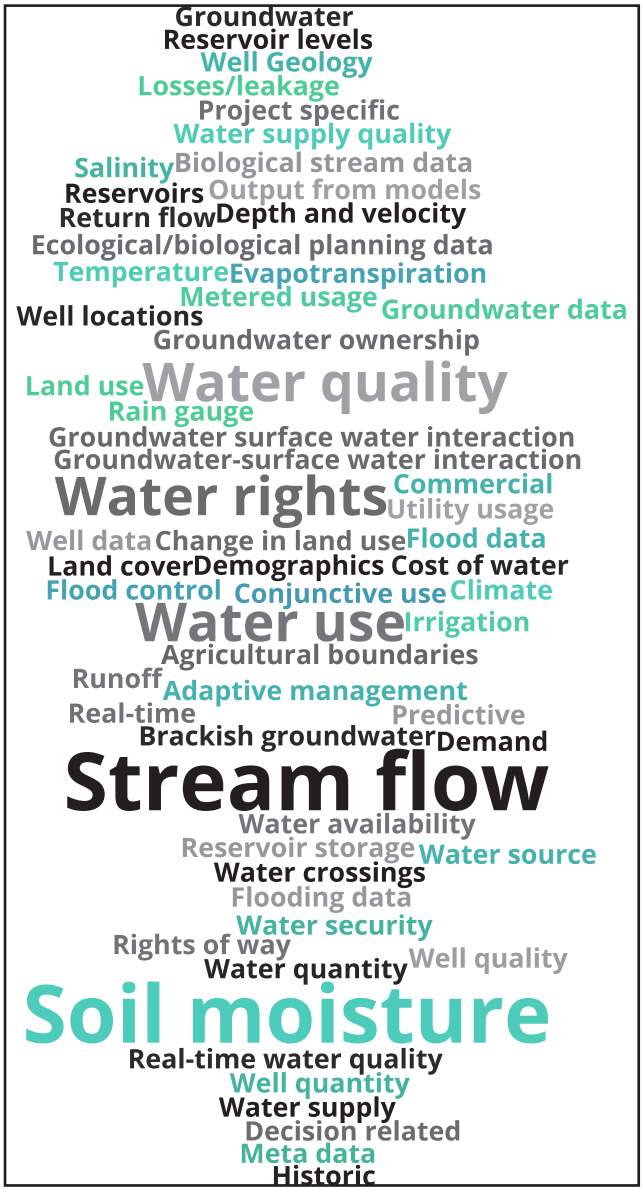


Figure 3. What data are needed.

variability of water-related decisions. Responses ranged from general purposes, such as understanding how much water a person uses or how clean one's water is, to highly technical purposes such as making flood risk determinations and updating water availability models. The full range of recommendations can be seen in the workgroups' results in Appendix II.

Narrowing the questions still further, participants in the workgroups were asked to describe gaps in water data that need to be filled. Not all groups listed gaps, but the data gaps that were noted provide insight into where more data are needed now and for the future. Examples from the list include more data on hydraulic fracturing water, citizen science data, climate forecasting data related to the groundwater-surface water interface, and real-time estuary inflow data.

Data gaps were generally distributed within three main groupings. These groups were (1) gaps in access to and integration of data, (2) gaps in data availability due to insufficient amounts of data or lack of any data at all, and (3) gaps in specific types of data. Data gaps are grouped by category and listed in Figure 5.

Appendix II provides detailed descriptions of data gaps by workgroup.

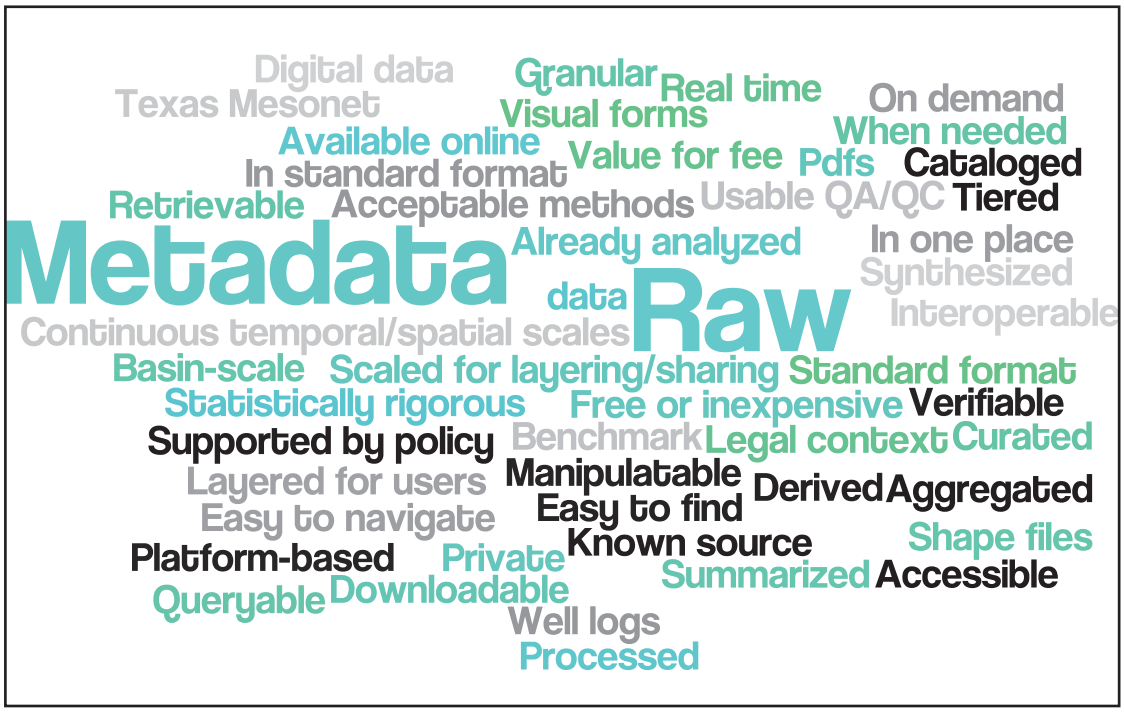


Figure 4. What form of data is most needed.

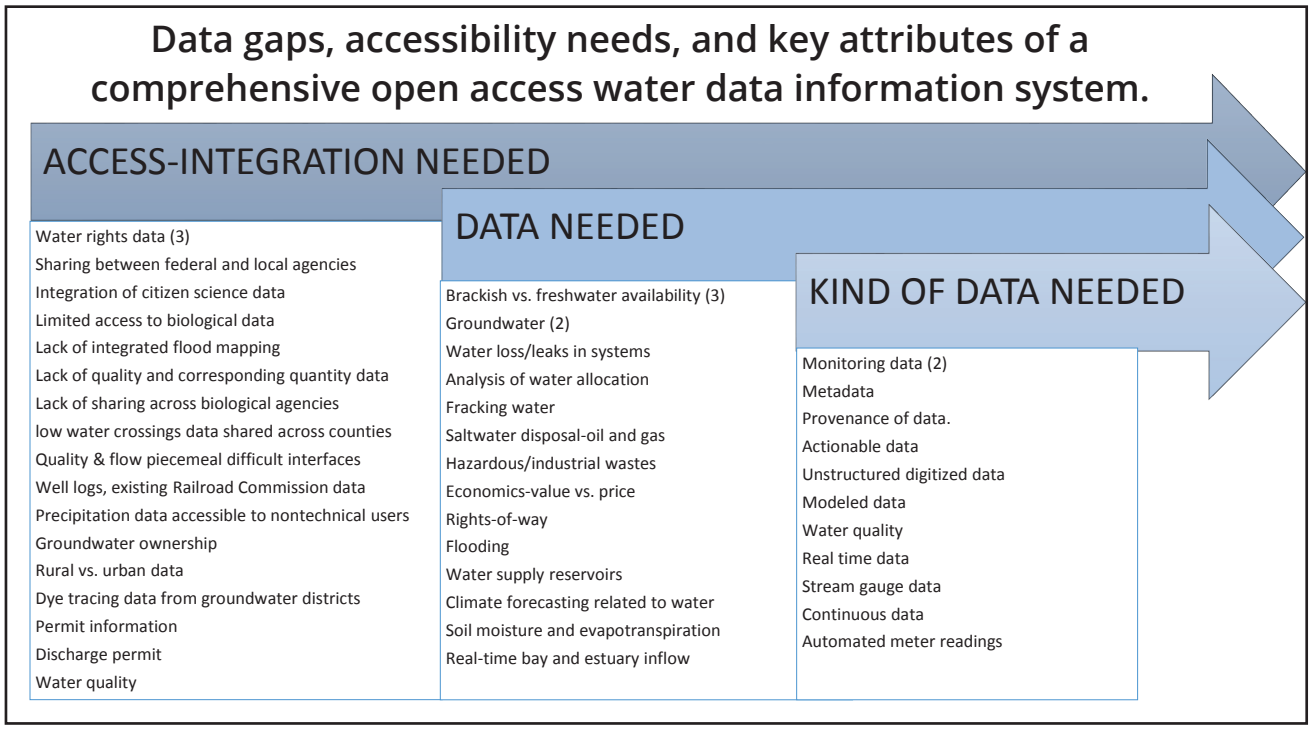


Figure 5. Data gaps by category.

THE IDEAL WATER DATA SYSTEM

The ideal data system was described as a series of integrated data hubs or nodes – with more added over time – specialized by water sector and application (i.e., ranging from expert to general public water stakeholder), with incentives for adding data into the hubs.

Following the workshop, participants were asked to respond to a survey and describe the ideal hosting option for open data hubs or systems. Respondents were almost evenly split in recommending (1) a Texas state agency, (2) a consortium of Texas state agencies and universities, and (3) a consortium of Texas state agencies, universities, and the private sector. A summary and complete responses to the survey are available in Appendix VII.

Overwhelmingly the most critical data needed to be included in an open data system are (1) raw data or data as close to raw data as possible, and (2) metadata.

Researchers and other highly technical users of data have the greatest need for such data. Several participants represented such interests at the workshop. However, such data may also be among the most difficult to access in general without an open system due to the likelihood of it being proprietary or difficult to access readily due to matters of interoperability or quantity.

Data needed by the full diversity of users must be easily accessible and interoperable to serve a wide variety of user needs. This includes needs for data at various geographic, spatial, and temporal scales, and in formats that conform to standards generally employed by the various users of data. Participants also identified qualities of data essential to ensuring data usefulness, such as data being findable, accessible, universally usable, and reusable. They suggested these qualities should exist in the ideal water data system.

One group used the acronym “FAIR” to underscore these qualities.

Water data should be **FAIR**:

F - Findable

A - Accessible

I - Interoperable

R - Reusable

RECOMMENDED USE CASES

To help organize and make a clear case for improved access to usable data to manage water supplies in the future, workshop attendees were asked to identify potential "use cases" that may serve as ready models for open data systems.

A use case is a short summary organizing in a concise and consistent format the data gaps, needs, uses, users, regulatory requirements, and workflow for a particular objective. Use cases serve as a tool for organizing and assessing stakeholder data needs, and communicating those needs to decision makers in water industries, utilities, and governments. They are developed to demonstrate the value of improved data for decision making.

Participants identified 35 potential use cases (Appendix III). Use cases varied greatly, without a single use case idea recommended by one group repeated by any other group. Several major categories of use case emerged, along with a general "water use case" category. Major categories were (1) groundwater, (2) water rights, and (3) event planning, which included two subcategories: (a) drought planning, and (b) flood planning (Figure 6). For example, in the four instances in which flooding was the general topic, the context was for (1) prediction and emergency response, (2) managing ephemeral streams, (3) impacts, and (4) crowd-sourcing observations in different water sources and for water quality.

Five of the workgroups each arrived at a consensus on a single use case for potential future development (Table 1). All five of the use cases recommended focus heavily on data needs for direct water use and management, including environmental management. These use cases involve technical water database management as well as socio-economic and policy data challenges. They are what are arguably among the most pressing data use challenges facing Texas decision makers.

We hope work on these agreed-upon use cases will proceed to illustrate the value of data in past decision making or to form a pilot for future decisions using data and data systems. We envision that these use cases will be responsive to stakeholder data needs, as well as useful for technical developers seeking to better understand the data needs of system users. Beyond the workshop, we hope to engage stakeholders in completing a set of use cases that help demonstrate the need for and use of data hubs for water and decision making.

Table 1. Top use cases recommended for Texas by consensus in five of the workgroups.
• Water utility reporting to the Texas Water Development Board
• Environmental flow transactions
• Flood water management in ephemeral streams
• Integrate and update the Texas Water Availability Models (WAM) and Groundwater Availability Models (GAM)
• Risk management of the probability of reservoir water supplies falling below criteria at 3, 6, 9, and 12 months

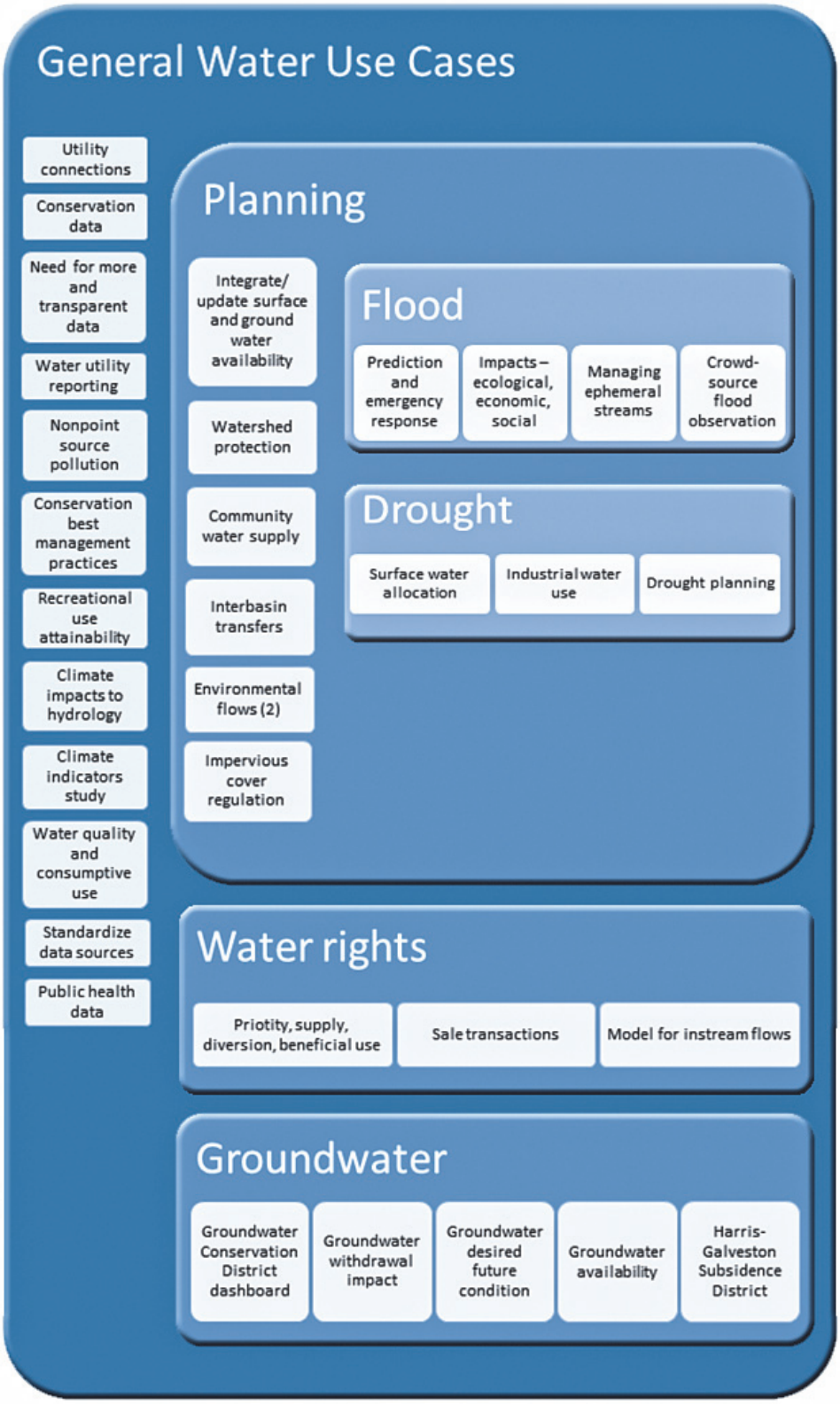


Figure 6. Use cases by categories and subcategories.

IMAGINE THE FUTURE

Work group participants consistently expressed an overarching belief that in Texas today, one needs to be an expert to find data that exist and to access those data and integrate them for practical use. They stated that much of the data that do exist are not actionable. This situation need not stand in Texas for water data. Participants created a better vision for the future and outlined a series of paths and actions to get there, including use cases as examples and pilots to achieve the desired outcomes.

Participants described a vision for the ideal water data system for Texas as one with open access that includes an ability to obtain available water data, including raw data, metadata, and legacy data in a digitized form. The data system will be user friendly, robust, and provide real-time information using web services with source information and built in visualization tools so that non-experts can use the system. Data and information will be free, and created and kept in consistent reporting formats so that data will "talk to each other" as users search and gain access. The ideal form of data system is envisioned as consisting of several integrated data hubs specialized by water sector, with incentives for people to add new data and share existing data through the hubs. There will be adequate funding to sustain the data system over time.

SPRINGBOARD TO THE FUTURE

Potential classes of use cases for future development

1. ***Events***, such as floods, droughts, and water supply ups and downs.
2. ***Markets***, can market forces be used directly or indirectly to drive new data and more access?
3. ***Unusual to the water sector, but important users,*** such as insurance companies, real estate developers, and banks.
4. ***Better decisions on costs or investments***, such as building new infrastructure and timing of reservoir releases.
5. ***Public engagement***, such as user-friendly dashboards and delivery of personal or neighborhood water usage information.
6. ***Uses already underway*** where improvements or additions to existing data will provide quick results.
7. ***Conflicts*** emerging or ongoing, including a use case associated with legal action contemplated or ongoing.
8. ***Locally-relevant successes*** showing where a small amount of data was used to change decisions affecting a local area or group.

Work on use cases was a centerpiece of the workshop and there was a consensus that work should proceed on one or more use cases. Recommendations varied. Several suggestions involved picking a use case or two that came from the workshop, and then forming pilot projects around the use cases to actually do something that shows the value of an open data system. One group suggested focusing on drought, because in Texas drought tends to be a key driver of innovation. Another suggestion focused on past decision making, to show how people have used data for practical real-world decisions benefiting people.

In providing synthesis of sessions, Dr. Doyle suggested building a use case centered around a high-profile action taken in Texas where available data were used in decision making, but where results would have been more beneficial if additional data had been available and accessible.

He suggested considering classes of use cases and possible advantages of developing use cases to illustrate classes of water data usage. Among advantages of this kind of approach is the potential to evaluate the costs and benefits of putting resources into one class of use case versus another. Through strategic consideration of action, Texas can be intentional about creating forces that push and pull data systems and understanding such systems in a defined fashion.

NEXT STEPS

Participants were asked to envision concrete next steps as a final part of their "springboard to the future" discussions. This was the end phase of workgroup discussion as the main discussion topic among participants at the final plenary session.

The following list aggregates the key recommendations into common categories and a sequence for action. There was considerable excitement among participants when presenting this final and perhaps most direct action-focused part of the workshop.

Start With Consensus

- Establish areas of agreement on standards for open data sources.
- Find out who has what data already.
- Find out who agrees with the idea of open data sources and hubs.

Plant a Flag

- Initiate one or more use cases.
- Establish an advisory task force to identify and support next steps.
- Establish the network structure for an open data system.
- Establish who will be "anchor tenants." These will be the key users of the initial data hubs.
- Establish which agency(s) or "who" will lead in developing and hosting the initial data hub(s). (Note: A general consensus of work groups is that the agency best suited to lead in developing and hosting the initial data hub is the TWDB's through the Texas Natural Resources Information System.

Tell Everyone

- Share information about open data experiences and best practices.
- Publish articles about the internet of water in media outlets such as Texas+Water and the Texas Water Journal.

Establish Lines of Support

- Identify funding sources.
- Develop incentives for sharing data.
- Gain legislative support, and seek funding and a policy mandate.

CONCLUSION AND THANK YOU

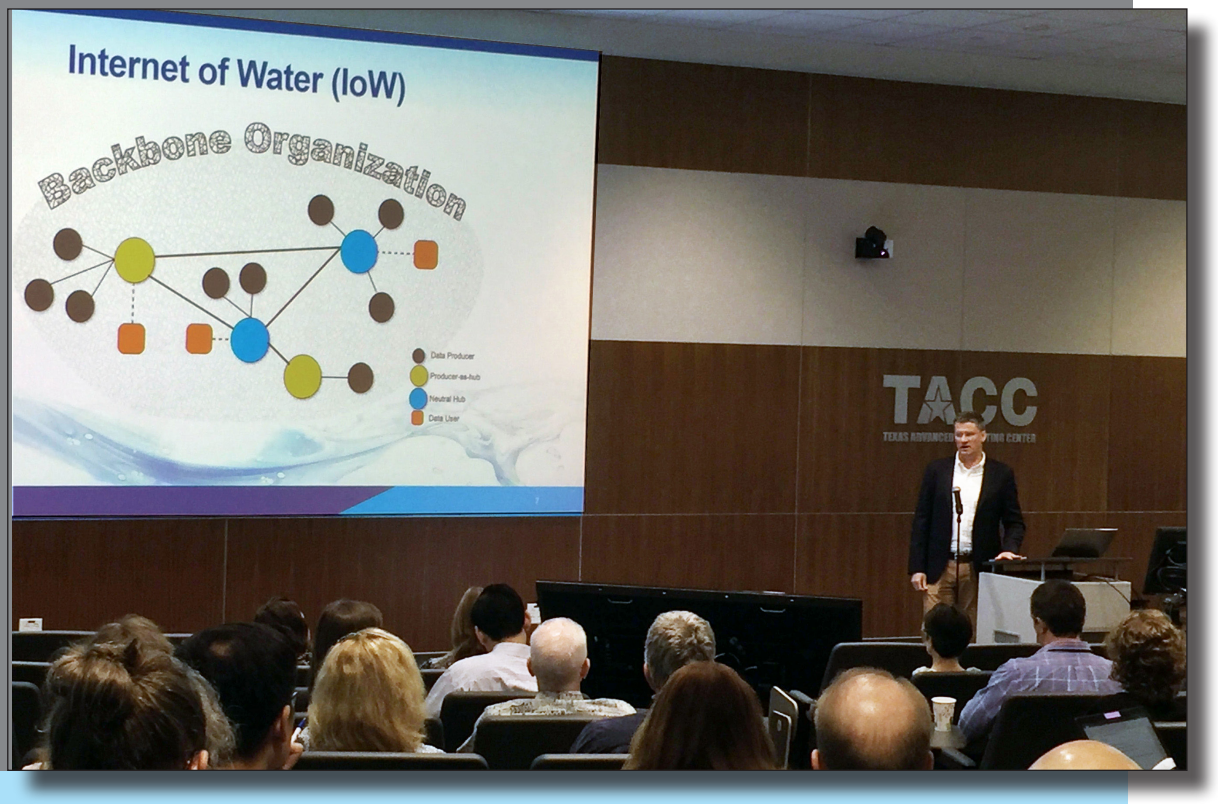
Most participants expressed satisfaction with the workshop (Appendix VII). The workshop achieved its objectives, with anticipated outcomes well covered by participant dialogue. Results of the workshop will help align ideas, underpin development of use cases, educate decision makers, and promote other first steps toward building a comprehensive, open access, water data information system capable of informing comprehensive water management decisions.

The sponsors and organizers are grateful to all participants for taking their time to meet with us and join with each other to help create a better vision for the future of data management and access in Texas and nationally. This dialogue must continue in various forms for work at the workshop to be relevant and useful. We thank all who participated and intend to follow up with all participants in the future.

APPENDICES

APPENDIX I

PROCEEDINGS, PROCESS, SYNTHESIS



"While it may not make sense to have a national water policy, participants at the Aspen Institute dialogue series concluded that it does make sense to have a national water data policy.." -- Dr. Martin Doyle

AGENDA

Opening Plenary Session. (9:00 AM – 10:00 AM)

- Welcome and introduction: Sam Hermitte, Assistant Deputy Executive Administrator, Texas Water Development Board.
- Introduction to the Texas Advanced Computing Center: Dan Stanzione, Executive Director of TACC and Assistant VP for Research at UT-Austin.
- Background/Orientation to the Internet of Water Initiative: Martin Doyle, Director of Water Policy Program, Nicholas Institute for Environmental Policy.
- Instructions/Workshop Process: Rudy Rosen and Susan Roberts, Director, Institute for Water Resources Science and Technology, Texas A&M University-San Antonio and Director, Water Systems Division, Texas Center for Applied Technology.

Breakout (Small Group) Work Sessions. (10:00 AM - 11:00 AM)

- **Big Picture:** Identify, describe, and list 1) who needs, 2) what data, 3) in what form, 4) to inform what decisions about water in Texas, including water supply, water quality, and environmental resources.
- **Data Gaps and Access:** Define the desired future water data management and access in Texas, by listing key attributes of a comprehensive open access water data information system capable of informing comprehensive water management decisions.

Plenary Synthesis Session and Group Discussion. (11:15 AM – 11:45 AM)

- Reporting of breakout session results. Facilitators.
- Synthesis and perspectives on morning sessions. Martin Doyle.

Keynote Address and Data Collaboration Networking Lunch. (11:45 PM – 1:00 PM)

- Keynote address. Kathleen Jackson, Board Member, Texas Water Development Board.
- Data Collaboration Networking lunch.

Breakout (Small Group) Work Sessions. (1:00 PM - 2:45 PM)

- **Texas Use Cases:** To initiate development of use cases for Texas water by identifying critical needs of Texas data providers and consumers. Following a template, facilitators will lead workshop participants in developing draft use cases across water topics and objectives.
- **Springboard to the Future**: Speed-list ideas on next steps to further define, design, and build a water data system for Texas.

Plenary Synthesis Session and Group Discussion. (3:00 PM – 4:00 PM)

- Reporting of breakout session results. Facilitators.
- Synthesis and perspectives on sessions. Martin Doyle.
- Open discussion: consensus building ideas and “next steps.” Rudy Rosen.
- Summary and closing statements: Sam Hermitte.

Guided Tours of the Texas Advanced Computing Center (4:00, 4:15, 4:30 PM)

CONNECTING TEXAS WATER DATA

WORKSHOP TEAM

ORGANIZING COMMITTEE

Sam Marie Hermitte - Texas Water Development Board

Suzanne Pierce - Texas Advanced Computing Center, University of Texas

Sarah Richards - The Cynthia and George Mitchell Foundation

Rudolph Rosen - Institute for Water Resources Science and Technology, Texas A&M Univ.-San Antonio

Susan Roberts - Texas Center for Applied Technology, Texas A&M Engineering Experiment Station

FACILITATION TEAMS

Group A
Mike Myatt - Water Foundation
Emily Warren - Meadows Center for Water and the Environment, Texas State University

Group B
John Tracy - Texas Water Resources Institute, Texas A&M University
Lauren Patterson - Nicholas Institute for Environmental Policy Solutions, Duke University

Group C
Robert Mace - Meadows Center for Water and the Environment, Texas State University
Natalie Freed - Texas Advanced Computing Center, University of Texas

Group D
Todd Votteler - Collaborative Water Resolution
Carrie Thompson - Water Table Consulting

Group E
Glen Low - The Earth Genome
Corinne Wong - Environmental Science Institute, University of Texas

Group F
Dorina Murgulet - Texas A&M University-Corpus Christi
Susan Roberts - Texas Center for Applied Technology, Texas A&M Engineering Experiment Station

WORKSHOP OVERVIEW

THE OBJECTIVES

1. **Big Picture**: To identify, describe, and list 1) who needs, 2) what data, 3) in what form, 4) to inform what decisions about water in Texas, including water supply, water quality, and environmental resources.
2. **Data Gaps, Management, and Access**: To define the desired future of water data management and access in Texas by listing data gaps, accessibility options, and key attributes of a comprehensive open access water data information system.
3. **Texas Use Cases**: To initiate development of use cases for Texas water by identifying critical needs of Texas data providers and consumers. Following a template, facilitators will lead workshop participants in developing draft use cases across water topics and objectives.
4. **Springboard to the Future**: To speed-list ideas on next steps to further define, design, and build a water data system for Texas.

THE PROCESS

CONNECTING TEXAS WATER DATA

A VISION

A water data system for Texas will support access to an accurate accounting of supply, quality, and use of water to better support decision makers in their efforts to enhance sustainable water use. Improved access to and standardization and integration of data, will provide water managers and decision makers a better basis for data-driven decisions, enabling them to more confidently meet urban, agricultural, ecological, and industrial needs for water.

Workshop planning was conducted by a team of organizers representing the following sponsors: Texas Water Development Board; The Cynthia and George Mitchell Foundation; Institute for Water Resources Science and Technology, Texas A&M University-San Antonio; Texas Advanced Computing Center, and; National Science Foundation Research Coordination Network for Climate, Energy, Environment and Engagement in Semiarid Regions. Representatives of The Aspen Institute and Texas Water Research Network also supported the planning team.

A dedicated website supported registration and communication between registrants and organizers. The website also presented background information, reference materials, interactive templates, the agenda, and details about the workshop.

Upon arrival at the workshop, participants were welcomed with an overview of objectives, an introduction to the Texas Advanced Computing Center, and a history of recent work nationally on the concept of developing an "internet of water." Participants then heard about the facilitation process to be followed in morning and afternoon small group sessions, and were introduced to the twelve facilitators who worked in teams of two. Participants received a package of templates, a glossary for use during facilitated sessions (Appendices IV and VIII), and a link to interactive templates. During a networking lunch, participants heard from a member of the board of the Texas Water Development Board about the Board's interest in making water data more accessible.

A post-workshop survey was conducted to allow for follow-up questions and input, as well as gauge participant opinion and satisfaction of the workshop and results. A final report of workshop transactions, results, recommendations, survey results, and proposed actions was published. This document is that publication.

OPENING SESSION

WELCOME

The workshop opened with a welcoming address by Sam Marie Hermitte of the Texas Water Development Board (TWDB). Ms. Hermitte described the reasons behind the workshop and expectations for attendees. After briefly describing early initiatives to create an "Internet of Water" in a few other states, she welcomed participants from the Aspen Institute and the State of California where an open data initiative is already underway. She indicated that participation at today's workshop by water data experts who have experience dealing with development of open data systems elsewhere may help add some perspective to the day's outcomes. Finally she thanked the sponsors and attendees for supporting the ambitious goals for the day.

THE TACC

Workshop participants were welcomed and introduced to the Texas Advanced Computing Center by Dr. Dan Stanzione, center director and Assistant Vice-President for Research at the University of Texas. The center designs and operates some of the world's most powerful computing resources. He stated that the center's mission is to enable discoveries that advance science and society through the application of advanced computing technologies. Dr. Stanzione emphasized the availability of the center's resources to researchers and invited all participants to tour the facility at the end of the day.

INTERNET OF WATER INITIATIVE

Dr. Martin Doyle of the Nicholas Institute for Environmental Policy Solutions at Duke University started his presentation with a story that illustrates in clear language the great importance to water users of open and easy access to existing water data. He stated the difficulty in terms of complexity and time spent that one encounters throughout most of the United States today when seeking information about the safety and exact quality of water coming from one's own tap. Today so much decision is made on the basis of instantly available data, but for water, which is the most important ingredient for life on Earth, access to data for most Americans is far from instant.

One of the primary drivers of access to water data is the fracturing of where water data come from and the diversity of organizations behind generating and storing data. He mentioned the US Geological Survey (USGS) and the associated water science centers in the various states that maintain the database and stream gauge network for the National Water Information System (NWIS). Dr. Doyle described that as the "gold standard" for surface water data and open access. This system presents a ready foundation for building a nationwide open network for water data. But the policy driver behind the NWIS is the mission and mandate of the USGS. Those data are foundational to how we think about surface water availability in the US. However, water quality data are collected for different purposes than to provide understanding about flows and quantities. These water quality data are generally used to address regulatory requirements for monitoring water and meeting set standards to remain in compliance with water discharge permits administered by environmental protection agencies. Pending on the industry reporting, data are collected and managed by different federal, state,

and local agencies with different purposes. Examples include agencies responsible for energy, agriculture, community development, forestry, fisheries, endangered species, watersheds, and so on.

Dr. Doyle and collaborators are seeking a means to have data that come from all of these various sources made available and viewable on a real-time basis. This has been termed, the "internet of water."

The internet of water was initially formed in the course of a water forum held by the Aspen Institute involving about 50 people who came together to talk about water and big data. Following the forum, a group of funders came forward to support the initiative, in particular to seek means to make water data more useful to society. A dialogue series then pulled together a highly diverse group of water experts from utilities, state and federal government, oil and gas, philanthropy, academia, nongovernmental organizations, software companies, and other sectors. The result was a consensus formed by people with very different perspectives around key findings and recommendations for going forward. While it may not make sense to have a national water policy, participants at the Aspen Institute dialogue series concluded that it does make sense to have a national water data policy.

The complete findings of the dialogue series were included in reference materials made available to all participants in today's workshop (Appendix V). There were three key findings:

First, there needs to be a vision about how water data will be used, a notion of usefulness, and a quantification of its value. Water is commonly known to be undervalued, but water data are generally even less valued. Without a sound value proposition for water and water data, it is hard to obtain sustained financial investment in water data infrastructure. The group recommended prioritizing value propositions and understanding how water data can help various sectors meet their mission and gain a return on investment.

Second, there needs to be a series of regional pilots, or use cases, that solve real-time real-world water management problems. This is also a way to show the value of water data. Decisions are being made without data, so pilots will bring data and their value in front of decision makers and to the forefront of underpinning solutions. The group agreed that public or government curated data should be a priority for attention by the data initiative's proponents. These data are collected using taxpayer dollars, should already be publicly available, and the federal government has expressed a commitment to make its public data more open and discoverable. Public data historically have been trusted and seen as authoritative, providing a framework on which other data may be leveraged or validated. Yet large portions of government water data remain inaccessible and lack interoperability. This public data can form a common framework for building a comprehensive open data system. Such government data combined with data from other sources represent a huge store of water data. While much of the non-governmental data also remain undiscoverable and inaccessible, with access that too could be used to build an open data network and help improve the nation's water security.

Third, there needs to be created an internet of water using these data. This would be a framework that enables data systems to talk with one another. However, participants at the Aspen Institute dialogue series concluded that this not be done through a centralized system or a system managed by any one governmental agency. They recommended networking through a federated system of data producers, users, and hubs such as the USGS National

Water Information System and National Ground-Water Monitoring Network. This allows data producers to maintain control over their own data, which proved to be of paramount importance. This concept was called "The Internet of Water."

Dr. Doyle stated that once data hubs are up and running, new water users and water data uses will emerge, and new kinds of data hubs will form. He specifically mentioned new proprietary and private data hubs forming that would provide targeted access and support the needs of validated users. The network will grow organically, with the value of the data and new accessibility increasing as people discover its existence.

One of the key ideas to be explored by today's workshop is the development of use cases, tied to specific beneficial uses of data to solve problems. Dr. Doyle urged participants to think about management decisions made that could have been made better with better access to data on a real-time basis.

Finally, Dr. Doyle explained that the day's workshop will fit well into the series of roundtable discussions that the Aspen Institute is supporting in a few other states and locations. So far roundtables have been held in Texas, California where there was a focus on water policy, Detroit with an emphasis on the Great Lakes, and St. Louis where agriculture received greatest attention. Roundtables in Colorado and Seattle are scheduled for the near future. In selecting sites for roundtables, there has been an effort to include a diversity of geographies, economies, and sector demographics.

WORKSHOP INSTRUCTIONS

Drs. Rudy Rosen and Susan Roberts of the Institute for Water Resources Science and Technology and Texas Center for Applied Technology, Texas A&M University System, introduced workshop participants to the agenda for the day. Participants heard that workshop activities will take place in small group facilitated work sessions in the morning and afternoon, immediately followed by plenary sessions where facilitators will report on the work of the small groups and a summarizer will add perspective to the reports. Participants also heard that at noon there will be a keynote presentation by TWDB board member Kathleen Jackson followed by a networking lunch. After hearing about the agenda, participants were introduced to members of the facilitation team and assigned to one of six groups for work sessions.

KEYNOTE ADDRESS

Kathleen Jackson
Texas Water Development Board

"The better the data, the better the science. And the better the science, the better the policy." -- Kathleen Jackson

To kick off the keynote address, TWDB Board member Kathleen Jackson introduced a former TWDB Board chair in attendance, Carlos Rubinstein, and several former and current members of the TWDB staff. She thanked them for their contributions and recognized staff's important role in contributing to the agency's success. Ms. Jackson then described her engineering background and former work with ExxonMobil. She explained how this background often motivates her to focus on objective measurement of success. She shared examples of TWDB success and how that success has been measured.

Her work in the oil and gas industry often involved managing risk. She related that to current efforts by the TWDB and the state, in general, to manage risk associated with water availability in the face of Texas' recurring droughts. She said, "It seems as though Texas is in a state of perpetual drought punctuated by brief periods of extreme flooding."

She then turned to demand for water supply. The TWDB works in an environment in which groundwater supplies are being depleted as the agency works hard to research and potentially identify new water sources for communities. We seek new supplies not just to ensure current residents have the water they need, but also to supply the needs of a growing population. She said, "Anywhere from 1,000 to 1,200 people are moving to Texas every day and not one of them is bringing any water with them." She continued, "The TWDB plans for drought and to meet the needs of a growing population."

Director Jackson then spoke of her experience traveling around Texas talking to people with "boots on the ground." She shares what the TWDB is and does and always makes the point that the TWDB is the data repository for all water data for Texas. She considers that role vitally important, especially from the standpoint of supporting the science mission of the agency and the use of those data by others. She emphasized that it is important to make raw data available so people can access and use the data for new purposes and reach their own conclusions.

Ms. Jackson also spoke of the TWDB's role as a bank and lender. She stated, "We have money to loan, and you won't get a better interest rate for water project funding than at the TWDB. This is a message delivered all around the state."

Much of what the TWDB does is water planning. The regional water planning process involves more than 450 volunteers across the state with diverse backgrounds, representing big cities, small communities, agriculture, manufacturing, and all the other water users. The volunteers come together to compile strategies to address future water needs and determine "how much water we have today, what we need to do for tomorrow, and what strategies or projects we need to put in place to get us where we need to be in the future." Director Jackson spoke of the TWDB's extensive water planning and regional water plans that come together to form the state water plan, looking out 50 years and updated every 5 years. She spoke of how the plans are data- and science-driven and done cooperatively with the 16 regional water planning groups. She said, "We use the best data available and make the data transparent and usable on multiple platforms." But in spite of all the work on water plans, "We don't plan to plan, we plan to build."

She said that if you look at where we are today, our success is measured by the quality of projects that are moving forward in Texas. At this time, the TWDB has committed $6.2 billion for projects in the SWIFT program. These include projects such as the $3.3 billion Houston-area water supply project, one of the largest water infrastructure projects underway in the nation.

Director Jackson also described her affinity to the land, having been involved in rice farming as a family business. As a result, she understands the critical role water conservation plays in Texas' water past, present, and future. The TWDB's role in managing and sharing state water data reaches across all water initiatives, including water conservation. She emphasized that role in helping to create a culture of conservation among people throughout the state, as well as funding big construction projects. She stated, "We need to instill a culture of conservation so it's an everyday part of what we do." To make this happen, she emphasized that we need to empower people by providing access to data about their own water usage so they can take personal action based on sound data. When people understand where their water comes from and learn what it takes to get water to them, they are more likely to conserve.

Ms. Jackson used a data-sharing initiative with the oil and gas industry as a final example of the TWDB's ongoing efforts to develop open water data systems. House Bill 30, passed by the Texas Legislature in 2015, created a charge to develop brackish groundwater productivity zones and determine ways Texas' brackish groundwater can be harvested. The first step for the TWDB was to review available information. While some data sets were already available to the TWDB, the agency staff understood that other valuable data might have been collected elsewhere but were not readily accessible. Staff believed that the oil and gas industry, in particular, had additional data because of its extensive use of brackish groundwater in production activities and was uniquely positioned to provide well log and corresponding brackish water quality data. Through a collaborative effort, the initiative gained momentum and moved forward successfully after identifying targeted technical objectives, ensuring the data transfer was not burdensome to industry personnel, and determining there were no adverse unintended outcomes as a result of opening access to these data. While directly addressing the charge of House Bill 30, opening access to this set of raw data also directly benefited the oil and gas industry. Once aggregated, the raw data formed a larger database than any one company had access to and can now be used to further the use of brackish groundwater by industry and the public. The TWDB gained additional data and strengthened a continuing collaborative relationship with the Texas oil and gas industry.

Finally, Ms. Jackson stated that the workshop brought together key players and then urged participants to form enduring collaborative relationships during the day, in addition to sharing information and ideas about open data systems for Texas. She thanked all attendees for participating and commended them for their engagement and support, which allows Texas to continue to be an economic leader in this nation and the world.

SYNTHESIS

DR. MARTIN DOYLE

MORNING SESSIONS

During morning group sessions participants consistently expressed an overarching belief that in Texas today one needs to be an expert to find data that exist and to access those data and integrate them for practical use. They stated that much of the data that do exist are not actionable. They defined water data in a highly broad comprehensive fashion, because it was apparent from the participants' long list of data users, needs, and uses that the primary user or "audience" is not clearly in focus. Virtually every need, possible use, and everyone made the list at one time or another in discussion. Some groups simply described the user as "everyone" or "the public." All needs, all uses, were at one point or another expressed as possible additions to the list of water uses. In synthesizing the session, Dr. Doyle stated that, "if you are speaking to everybody about everything, then you aren't speaking to anybody about anything."

One group was an exception. Participants in that discussion made it clear that for them the key user of water data is the water resources expert. Researchers, analysts, managers, and water decision makers fit into the category of expert.

There is a need to segregate work on data systems to focus on the type of user expected to access the system or particular data sets, whether that's for an average citizen or for a water analyst. Dr. Doyle reminded participants that as we begin to form plans for building data hubs and accessible data systems, that we need to be explicit about the end user. It's simply not realistic to build a single water data system for use by the average citizen and the water expert.

Dr. Doyle used the Weather Channel as an illustration. He stated, "while the average citizen can access and use the Weather Channel and accompanying online resources, the average citizen can't use the USGS stream gauge network in the same way."

Consider the different technical resources and investments required to form up an equivalent to the Weather Channel for a particular data set versus forming something like the USGS stream gauge network. The investment in technology and human resources differs in developing a system for average citizens with a heavy emphasis on synthesis and visualization dashboards versus a data system for water experts who may desire raw and accompanying metadata.

Dr. Doyle stated that almost every group mentioned a Google of water, but what they really meant was a Google of water databases. This would be an open source for links to and information about databases that exist. Such a system would be seen as a desired first step toward a comprehensive open data system.

He also mentioned an emerging realization that time delay in use of one data set versus another would greatly influence data applicability in decision making and thus interest by one group of users versus another. For example, decision making such as, "how many acres should I plant?" will require data collected over a different time scale than decisions about changing the way a major utility is operating to meet projected population increases. Participants talked about near real-time data providing early indicators that can be used to make near instant decisions of immediate consequence. They stated that data useful for "hour-by-hour" and day-by-day" decision making are probably beyond the scope of current discussion. However "week-to-week" and "month-to-month" data and decision making seem to be an immediately attainable sweet spot.

AFTERNOON SESSIONS

While there was consistency in discussion from group to group during the morning sessions, session summarizer Dr. Doyle sensed that discussion by afternoon groups started out in somewhat similar directions, but by the end of the sessions discussions varied greatly from group to group. That prompted Dr. Doyle to suggest that as we start thinking about how to proceed in developing use cases, where we begin considering where to apply resources, and when designing data systems that we consider who is in the room. Why? Because who is in the room and party to discussions and decisions matters greatly. It did at the workshop and it will wherever a group of individuals with diverse backgrounds who represent varied interests is convened. This advice was not offered as a value judgment on outcomes, it was just a recognition of the reality of group dynamics.

There was considerable discussion about incentives and policy requirements that may support the evolution of data systems and markets. These may also help further drive data system use and expansion. Dr. Doyle suggested that through strategic consideration of our actions, we can be intentional about creating forces that push and pull data systems in a defined desired fashion and direction.

While observing groups in the afternoon as they developed use case ideas, one idea in particular captured Dr. Doyle's imagination. This was to build a use case centered around a high-profile action taken in Texas within the past year where available data were used in decision making, but where results would have been different and better if additional data had been available and accessible.

Dr. Doyle also suggested to participants that when experts, such as attendees at the workshop, get together and consider questions such as those posed during the day's sessions that they have a strong tendency to identify and discuss items in a top down fashion. That may overly complicate understanding. He suggested an alternative approach is to ask people in the trenches of day-to-day decision making, "what are you now doing with water data and how are you actually making decisions with those data." This would cast a wider net in a search for instances of Texas' water managers taking action using data that are already available.

Moving from an assessment of the day's group discussions and thinking more broadly had Dr. Doyle compare the discussions in Texas with similar activities in California, Missouri, and Michigan. He suggested that in addition to considering isolated use cases illustrating an action taken or desired, that we think more broadly. He suggested considering classes of use cases and possible advantages of developing use cases to illustrate classes of use. Among possible advantages of this kind of approach, it may be possible to evaluate the costs and benefits of putting resources into one class of use case versus another.

Here are examples of possible classes that Dr. Doyle suggested could be used to categorize possible use cases:

1. Events, such as floods, droughts, and water supply ups and downs.
2. Markets, can market forces be used directly or indirectly to drive new data and more access?
3. Unusual but important users, such as insurance companies, real estate developers, and banks.
4. Better decisions on costs or investments, such as building new infrastructure and timing of reservoir releases.
5. Public engagement, such as user-friendly dashboards, delivery of personal or neighborhood water usage information, and public shaming campaigns.
6. Already happening uses where improvements or additions to existing data will provide quick results.
7. Conflicts coming or ongoing, including a use case associated with legal action contemplated or ongoing.
8. Locally-relevant successes showing where a small amount of data were used to change decisions affecting a local area or limited group.


AT&T Wifi Coupon Code

APPENDIX II

The Big Picture

Data Gaps and Desired Future

Breakout Session Details

ACTIVITY 1
BIG PICTURE

Identify, describe, and list 1) who needs, 2) what data, 3) in what form, 4) to inform what decisions about water in Texas.

ACTIVITY 2
DATA GAPS & FUTURE

Define the desired future water data management and access in Texas, by listing key attributes of a comprehensive open access water data information system capable of informing comprehensive water management decisions.

GROUP A

BIG PICTURE

WHO NEEDS ?

- Academic researchers
- Engineering firms
- Regulatory agencies
- Oil and gas companies
- Farmers
- General public
- Regional water plans

WHAT DATA?

- Water use
- Groundwater
- Brackish groundwater
- Flooding information (pre, during, and post event)
- Groundwater ownership
- Rights of way

GAPS, DESIRED FUTURE, ADDITIONAL POINTS

GAPS IN WATER DATA

- Water loss/leaks in systems
- Analysis of water allocation
- Fracking water
- Unstructured digitized data
  - Water rights
  - Discharge permits
- Groundwater
- Saltwater disposal-oil and gas
- Alternative water sources
- Economics-value vs. price
- Modeled data
- Monitoring data
- Brackish vs. freshwater availability
- Groundwater ownership
- Rights-of-way
- Flooding
- Water quality
- Hazardous/industrial wastes
- Water supply reservoirs
- Monitoring sites
- Groundwater
- Sharing between federal and local agencies
- Integration of citizen science data

IMAGINE THE FUTURE

Participants imagined a future of open access and ease of accessibility that included an ability to access lots of information, including legacy data in a digitized form. That information would be user friendly, robust, complete with metadata, and moving more to real-time information available on web services with visualization tools built in so that the average person can actually get something useful out of it. They perceived that data and information would be free, and in consistent reporting formats so that the data would "talk to each other" as its being accessed by the user. They also envisioned a future where there would be adequate funding to sustain the data systems over time. Participants also got into a discussion about citizen science. There were deferring opinions on the value of data derived from citizen scientists, especially on matters of quality control of data for it to be safe and useful. They believed that there would need to be a way to place such data into a context for viable use. Finally, participants discussed the ideal form of a data system. They suggested that several integrated data hubs specialized by sector was most preferable, with incentives for people to add to and share data into the hubs.

GROUP B

BIG PICTURE

WHO NEEDS ?

- Water experts
- Analysts
- General Public

WHAT DATA?

- Project specific
- Decision related

WHAT FORM?

- Raw
- Meta
- Cataloged

FOR WHAT?

- Curated
- Derived

- Analysis
- Synthesis
- Decision making

GAPS, DESIRED FUTURE, ADDITIONAL POINTS

The data that would be most useful in any data hub would be the metadata, associated with who has what data, for what purpose, and the provenance of the data. Participants stated that the users of the data would be individuals involved in research studies and analysts seeking to access specific data or studies. They would benefit simply by having a source to be able to find data by subject and by having a data hub that would provide them with a catalog or curated listing that directs them to a location in a data hub where they could access appropriate raw data or curated data. Participants believed that this would aid research and synthesis of activities related to water and providing input and planning advice to decision makers.

Participant discussion focused on gaining access to and a critical need for raw data, or data as close to raw data as possible. However, participants cautioned that it's not always possible to obtain all data in a raw state, but they emphasized access to data as unaltered as possible. Who would use this? It would be used in the course of work by experts for analysis and synthesis, and passed on to others for decision making on water resources, including water users, water utility managers, and so on. They felt that this would be focused on the water resources professional, but not the general public.

Participants also discussed existing data hubs. They mentioned the US Geological Survey's National Water Information System (NWIS)as an example that participants' access. However, participants mentioned that even though the NWIS is useful, it is limited in use in the built environment, i.e., where water has been removed from the environment and modified by treatment or use). Participants then used the built environment as an area where there is a data gap and an area for future focus on providing access to or more quantitative data that would be useful for people in the water resources profession addressing questions in the built environment.

Participants used a common acronym to describe the desirable state of water data. Water data should be FAIR:

F - Findable
R - Reusable
A - Accessible
I - Interoperable

GROUP C

BIG PICTURE

WHO NEEDS ?

- The public
- Researchers
- Emergency responders
- Regulators
  - State
  - Federal
- Teachers
- Real-estate developers
- Farmers
- Ranchers
- Insurers
- News media
- Water districts
- Industries
  - Oil and gas
  - Technology
  - Energy
- Cities/communities

WHAT DATA?

- Where does MY water come from?
- Historic
- Real-time
- Predictive
- Metadata

WHAT FORM?

- Scale:
  - Individual
  - Community
  - Local
  - State
  - Federal
- Detail/granularity:
  - Raw
  - Tiered
- Access:
  - Easy to find
  - Easy to navigate
  - Available online
  - Summarized
  - Ability to drill down and disaggregate

FOR WHAT?

- To understand:
  - How much water I can use
  - How much water I am using
  - How clean is my water
  - Where is my water
  - My cost of water
  - Local restrictions
- Determine flooding risks
- Rainwater collection

GAPS, DESIRED FUTURE, ADDITIONAL POINTS

GAPS IN WATER DATA

- Real time data:
  - Missing stream gauge data
  - Infrequency of observation
  - Automated meter readings
- Rural vs. urban
- Permit information
- Water quality
- Disparity between needing both water quality and quantity data for use, but having only one or the other
- Continuous data and privacy/liability issues
- Limited access to biological data

Participants agreed that data should be accessible, easily navigable, interoperable, and failure safe. Participants spent considerable time talking about multiple levels of granularity and the quality of data, from broad data to distilled data, geographic indexing, sources, credits for who generated data, historic context, metadata, and curated quality. No agreement was reached on what should be done is data is of poor quality. Suggestions included allowing for users to add information or comments into the data" on the side" as well as to provide feedback to data mangers of the data hub on problems and, if possible, how to address problems with the data.

GROUP D

BIG PICTURE

WHO NEEDS ?

- Parents
- Natural resource managers
- Farmers
- Producers
- Flood control districts
- Everyone
- Groundwater districts: Utilities, Well owners, Agencies
- First responders
- Planners
- Developers
- Weather Service

WHAT DATA?

- Stream flow
- Water quality and quantity
- Salinity
- Temperature
- Soil moisture
- Evapotranspiration
- Rain gauge
- Well data
- Water crossings
- Flood data
- Depth and velocity
- Reservoir storage
- Agricultural fields
- Land use
- Land cover
- Conjunctive use
- Return flow

WHAT FORM?

- When needed
- Real time
- On demand
- Universally scaled for layering/sharing

FOR WHAT?

- To explain/educate: Recreation, Safety, Quality, Flooding
- Resource management
- To protect sensitive ecosystems
- Flood control districts

GAPS, DESIRED FUTURE, ADDITIONAL POINTS

GAPS IN WATER DATA

- Lack of integrated flood mapping, for emergency response, low water crossings, and shared across counties
- Climate forecasting related to groundwater-surface water interface, recharge, temporal/ spatial variability
- Precipitation data that is accessible to nontechnical users
- Soil moisture and evapotranspiration
- Real-time bay and estuary inflow
- Biological, agencies need to share
- Quality & flow piecemeal difficult interfaces
- Water rights, needs to be online and accessible
- Water availability
- Dye tracing data from groundwater districts
- Well logs, existing Railroad Commission data needs to be made accessible

IMAGINE THE FUTURE

Participants Participants suggested possibly downscaling the USGS national water model for application in Texas by adding state data to it. That would fill out the model for Texas, with added state water quality data creating a clearing house for Texas water information.

Participants also talked about integrating into the model remote sensing data available through the National Aeronautics and Space Administration (NASA).. Participants mentioned a series of satellites producing different data sets that could provide a source of data for a Texas water model.

GROUP E

BIG PICTURE

WHO NEEDS ?

- Utilities
- Consultants
- Agencies
- Legislators
- Agriculture producers
- Water users
- Public
- Watch-dog groups
- Courts/people in court
- Permit applicants
- Industry
- Financial institutions
- Insurers
- Researchers
- National Weather Service
- First Responders
- Oil and gas
- Public health agencies
- City planners
- Rights-of-way
- Flooding
  - Water quality
  - Hazardous/industrial wastes
  - Water supply reservoirs

WHAT DATA?

- Metered usage:
  - Reservoirs
  - Irrigation
  - Commercial
- Utility usage
- Losses/leakage
- Water source
  - Cost of water
- Stream flow
- Reservoir levels
- Water rights
- Water availability
- Wells:
  - Location
  - Quality
  - Quantity
  - Geology
- Groundwater surface water interaction
- Real-time water quality
- Output from models
- Future scenarios:
  - Climate
  - Demographics
  - Water security
  - Demand
- Trends:
  - Change in land use
  - Runoff
  - Precipitation soil moisture

WHAT FORM?

- Free or inexpensive
- Queryable
- Manipulatable
- Accessible
- From known source
- Verifiable
- Metadata
- Supported by a policy framework
- In one place
- In standard format
- Downloadable
- Includes legal context
- Layered for different users

FOR WHAT?

- Update water availability models
- Public access to models
- Emergency response
- Recreation decisions
- Border protection
- Property valuation
- Use prioritization
- Address unintended consequences of water decisions
- Food security
- Know water footprint
- Access alternate sources of water
- Understand energy needs
- Mitigation decisions
- Environmental impact assessment
- Water availability and allocation
- Understand regulatory successes and failures
- Conservation
- Protection
- Funding decisions
- Infrastructure decisions
- Know water quality

GROUP E (CONT'D)

GAPS, DESIRED FUTURE, ADDITIONAL POINTS

Participants were asked, "On a scale from 1 to 7, with 7 being the highest, where is Texas today on overall water data availability for decision making?" The answers scattered around 3 to 4.

A second follow-on question was asked, "How easy will it be for Texas to get to 7?" Answers were more scattered, but trended a bit higher with the midpoint between 4 and 5. Participants felt the key challenges to getting to 7 are the heterogeneity of data and the human component (i.e., the political will and ability to make data sharable).

Participants divided users into four general meta categories. (1) legislators and policy makers, (2) government agencies, (3) researches who help inform legislators and policy makers in the agencies, and (4) planners and the people who are actual users of water.

Participants stated that the most critical water decisions that would require new data or better access to existing data mostly had to do with the best use of water in the state. There was a strong focus on gaining access to actionable data. There was discussion on the relative value of raw data versus data from models, also expressed as raw data versus processed data. They stated that some end users require insights, not actual data. There was also considerable conversation about future scenarios, especially regarding data that will allow users to predict what might happen in the future. Data from the past may be indicative of what may happen in the future. There was also discussion about trends as indicators, and aggregating available information in a fashion that ensured it is not just data, but that it is actionable information.

Participants identified four areas or instances where data gaps -- lack of data and/or access to data – have created problems in the state:

1. Actual events like Hurricane Harvey.
2. Lawsuits, and how data can help inform understanding and decisions.
3. Suboptimal decision making at almost every level throughout the state, whether involving a utility, agency, or other.
4. For much decision making on water, not only do we often not have data to know what the problem is, we don't have the data to know how to make the right investments to fix the problem.

Participants' vision for the future for data is that it be open, real-time, accessible, free, interoperable, simple, user friendly, and fully integrated.

GROUP F

BIG PICTURE

WHO NEEDS ?

- Water conservation managers and decision makers
- Water Utilities
- Consultants – watershed, permitting
- Counties
- Educators
- Planners
- Developers
- Agriculture
- Nonprofits
- Groundwater Conservation Districts
- Landowners
- Land use planners
- Water resource managers
- Academics
- Research Community

WHAT DATA?

- surface water and groundwater quality
- Groundwater level data per aquifer
- Groundwater-surface water interaction (gaining or loosing water systems)
- Water supply quality
- Water use (surface water and groundwater)
- Stream-flow gauge data (i.e. flooding)
- Soil moisture throughout the state
- Biological stream data
- Water rights, ecological/ biological planning, adaptive management
- Water supply & flood control:
  - Reliable
  - How much supply is available
  - Change WAM from monthly to daily
  - Change models to deterministic
  - Analytical methods
  - Water & transportation infrastructure
  - Establish needs based on the type of problem: dynamic versus static

• WHAT FORM?

- Granular
- Platform-based
- Benchmark
- Agreed-upon structures & standards
- Agreed-upon and acceptable methods & measurements
- Pdfs and digital data
- Raw
- Processed & synthesized
- Already analyzed
- Visual forms (graphs, images)
- Retrievable
- Shape files
- Well logs
- “Private” data
- Texas Mesonet
- Value/quality for a fee
- Usable QA/QC
- Interoperable
- Determine what can be aggregated
- Statistically rigorous
- With metadata
- Continuous at temporal and spatial scales
- Basin-scale

FOR WHAT?

- Long-range planning
- Drought
- Prediction
- Real-time needs
- Public & policy visualization
- Improved analysis
- Responsiveness to regulations
- Access for specific uses such as permits
- Alerts/emergency detection
- Research
- Improve capacity to integrate data
- Hub for connectivity and improvements; address protection & risk
- Information to sectors by use
- Move from static to dynamic monitoring
- Ecologic responses to water quality and availability
- Monitor water rights and see results
- Better decisions
- Scalable actions
- Leverage other data sets
- Cost of data sets

GROUP F (CONT'D)

GAPS, DESIRED FUTURE, ADDITIONAL POINTS

Participants were asked about gaps in Texas water data and implications to management decisions and their visions of future data systems to begin bridging those gaps, and key attributes of open access data systems.

Gaps included the knowledge of available data and the ability to access connected, real-time water management data sets. Sufficient time to update databases was also felt to be a gap. Achieving granularity of data sets could be better accomplished via aggregation. At the river basin scale, lack of real-time views into the state of the basin, and access to connected data sets, are missing. Participants defined other gaps in the relative ease of using data and models for any basin.

Vision: Real-time data sets and databases will be dynamically linked. Data catalog(s) and viewers will be available in a central platform that also allows decentralized input. Water data management can learn from other sectors; for example, use of a "GitHub" type of open, community-wide management will also open transparency. Community-wide involvement and management of shared data sets will ensure that users can see origins of data and actions in view. An annotated collection of data about water derived from existing and heterogeneous databases/datasets with the goal of uniformity and coherence. A virtual data set/database to transparently view and query other databases?

APPENDIX III

Texas Use Cases Springboard to the Future

Breakout Session Details

ACTIVITY 1
TEXAS USE CASES

Identify critical needs of Texas data providers and consumers, describe, and list as potential use cases for Texas water across topics and objectives.

ACTIVITY 2
SPRINGBOARD TO THE FUTURE

Next steps to further define, design, and build a water data system for Texas

GROUP A

TEXAS USE CASE

WATERSHED PLANNING, WATER ALLOCATION, FLOODING

Participants formed three subgroups to work on Group A's top three recommendations.

Subject	**Watershed planning**
Objective	To develop a water budget for a river basin using science-based planning
Participants	TWDB, public utilities
Data	Land use over time and water use, actual groundwater and surface water use, metered data, water quality, endangered species data, surface water diversions, discharges, stream gauges, geologic data, soil moisture
Sources	Cities, counties, Natural Resources Conservation Service, Texas Natural Resources Information System, Google Earth, planning firms/organizations

Subject	**Water allocation**
Objective	To ensure that basic water needs are met, then use above that will be charged at full cost
Participants	Groundwater Conservation Districts, Texas Water Development Board, Texas Commission on Environmental Quality, businesses
Description	The more one pays, the more one cares

Subject	**Flooding**
Objective	To develop a collection of data sets that can be used to reduce risk, increase response, and set priorities on projects
Participants	Federal Emergency Management Agency, US Environmental Protection Agency, Texas Water Development Board, local public works agencies
Description	Flood mapping, reservoir levels and discharge

SPRINGBOARD TO THE FUTURE – NEXT STEPS

Next steps to move toward open data for Texas were listed. First is to gain legislative support, such as gain funding support or a policy mandate. Next is to identify the network structure, who will do what, determine partners, who will pay, and who will take leadership roles. Another is to pick an interesting use case or two that came from the workshop, and then form pilot projects around the use cases and actually do something that shows the value an open data system. The participants also recognized that Texas does have existing open data sources in place and operating. They suggested sharing information about the existing open data experiences and best practices, thus getting the word out about the value of open data sources. Finally, participants discussed establishing standards or guidance for open data sources so that people understand how and in what form to make data available, so it can be integrated better, and so people who may be apprehensive about open data can better understand what it means. This could help reduce barriers, along with meeting opponents of open data to help address fears.

GROUP B

TEXAS USE CASE

WATER UTILITY REPORTING TO THE TWDB

Subject	**Water utility reporting to the Texas Water Development Board (TWDB)**
Objective	To provide enhanced open access to water utility reporting data already sent to and logged into databases by the TWDB.
Description	Water utilities are legally required to submit three reports to TWDB: (1) Water use Survey, (2) Water Loss Audit, and (3) a Conservation Report. Those data are reviewed and processed, and entered in database format on the TWDB website. A PDF is then generated. However, if anyone wants to use the data across Texas they need to get all of the reports, read through the relevant ones and select desired data, and then reprocess the information into digital data for any kind of actionable use (i.e., data that were originally actionable, actionable again). This use case will be to make these data sets searchable and downloadable. There will be no privacy issues because all the data are public information to start with, it goes directly to a public agency, and it's being collected in database format. The use case project would make this data readily accessible through an open interface or interactive application. Emphasis will be on raw data, as opposed to exact uses of the data. Then those who access the data would synthesize the data as they felt most appropriate to meet their own needs.
Uses	• Industrial water use during drought • Better decision making on water-related investments • Higher visibility for addressing water loss and conservation actions • Explore utility billing structures • Many users for general research into and analysis of water use in Texas: o Innovation; Target setting for science and policy; Real-time data source; Engagement for education and consumer information sharing
Participants	TWDB, public utilities
Regulatory	• Legislative statutes and agency rules trigger reporting • Standardized by regulation
Workflow	• Utilities upload reports online • Design-build open access user interface • Determine extent of historical data to include for access • Translate data from forms to new accessible interface
Sources	• Public utilities, Texas Commission on Environmental Quality, water rights use, water sales, water flows, climate related, recharge rates • TWDB Water Use Survey, Water Loss Audit, and Conservation Report

SPRINGBOARD TO THE FUTURE – NEXT STEPS

Participants had several ideas for moving forward, including the recognition that droughts are a key driver of innovation, that an inventory of where data now resides would be a natural first step, that an advisory task force for next steps could be useful, and that a clearing house for water quality information would be welcome. A final idea was mentioned by participants that may be implemented immediately. This was taking immediate initiative to write editorial and opinion items to the public and water community stakeholders about the internet of water. In discussing the idea, use of Texas+Water and the Texas Water Journal, were suggested as currently available venues for such outreach and communication to stakeholders.

GROUP C

TEXAS USE CASE

ENVIRONMENTAL FLOW TRANSACTIONS

Subject	**Environmental flow transactions**
Objective	To have the greatest positive impact on environmental flows at the lowest cost
Data Gaps	Environmental flow study raw data, cost data for transactions, biological data, water availability (what's on the market), historical data at temporal and spatial levels
Participants	Lawyers, Texas Commission on Environmental Quality (TCEQ), Texas Water Development Board (TWDB), Texas Parks and Wildlife, river authorities, purchasers, sellers
Workflow	1. Identify potential funding sources 2. Identify possible sellers 3. Identify areas of need, e.g., threatened species 4. Compare historic to current flows Additional actions in no order that may be taken: • Review water rights seniority • Do cost-benefit analysis • Study prior cases • Assess water quality and impacts • Review predictive models • Review TCEQ process for amending water rights • Identify existing environmental flow rights • Estimate flows needed to make a difference
Sources	US Geological Survey, TCEQ, regulations/requirements, river authorities, wa-ter rights, environmental flow studies, stream flow including historical data (SB 2), water quality, existing environmental flow rights, water availability models, threatened species

SPRINGBOARD TO THE FUTURE – NEXT STEPS

Participants had a series of potential next steps, lead off by a need to identify funding sources for establishing the data hub, followed by an identification of "anchor tenants" which would be the key users and supporters of the hub. There was also discussion about creating an initial support group called, "Cooperating Agencies for the Temporal and Spatial Management of Environmental Occurrences of Water," or as participants affectionately labeled it, the CATS MEOW. But whatever it may be called, the idea participants voiced is to create a group or organization to work on data standards and communicate on data in Texas. Related to that was the notion of creating a users' forum to allow for feedback discussions between super users, help with general education, and use it to create a community of users at all levels. Participants also discussed the need to address barriers to participation for certain institutions, better understand what the barriers are, identify resistance (including who may be opposed), and address the barriers. This effort may include identifying a neutral broker for data to support whichever entity takes the lead on the overall effort, and find and motivate political champions so that some barriers may be reduced or removed by statute, for example by requiting some kinds of data from some sources be openly available.

Finally, participants considered which agency, or "who," would be best suited to lead in developing and hosting the key data hub. The conclusion of the group was that the TWDB's Texas Natural Resources Information System (TNRIS). Reasons for the choice included that TNRIS is neutral, public, supported by statute, and has a stable source of funding.

GROUP D

TEXAS USE CASE

FLOOD WATER MANAGEMENT IN EPHEMERAL STREAMS

Subject	**Flood water management in ephemeral streams**
Objective	To better prepare for flood water management and emergency response in ephemeral streams in Texas
Description	Flash floods occur in ephemeral streams, sometimes even at low levels of rainfall. Emergency and natural resource managers need to prepare of unanticipated flood scenarios.
Data Gaps	Need rain map for the ground (i.e., how water moves and accumulates once it hits ground)
Uses	• Produce data for immediate use in emergency • Many data resources must work together immediately and flawlessly on public health and safety • Way to access real-time inundation conditions, spatially and temporally • Understand how waters will recede • Determine opportunities to divert water off-channel for storage and flood reduction • Placement of flood control structures • Identify biological areas that benefit from flooding • Post-flood damage assessment • Baseline data on impacts on soils (erosion) and nutrients
Participants	County government, National Weather Service, US Geological Survey, citizens, local media, first responders, Texas Water Development Board, Texas Water Development Board, Texas Division of Emergency Management, Federal Emergency Management Agency, flood management districts, Natural Resources Conservation Service, cities, landowners, nongovernmental organizations, conservation districts, engineering consulting firms, river authorities, water utilities, wastewater facilities, resorts
Regulatory	• FEMA flood plain mapping drives insurance • Tort law • Federal and state designation of "State of Emergency" • Legally required reporting, including industrial spills from treatment facilities • Local codes and ordnances • Local, state and federal determinations of evacuation and other orders for health and safety
Workflow	• Need a mechanism to bring together data from many sources immediately • Need an organization (assigned or created) to answer data questions for Texas flood emergencies
Sources	Same as Participants (above)

SPRINGBOARD TO THE FUTURE – NEXT STEPS

Participants emphasized developing examples of how people have used data for practical decisions, i.e., real world examples of benefits to people. The suggested conducting a survey to determine, "who has what data already." Participants stated that there may be more data available than generally assumed, possibly because there may be few or no incentives for collectors of data to share with others what data they have and to support making data sets available. Participants asked, "what are the incentives for organizations to share given already strained budgets and a lack of time to do basic work?" They also asked about disincentives to sharing access to data, especially for the private sector. There was even discussion about how some public organizations may be reluctant to open and share data because of fear of legal action against the agency. All this discussion focused on addressing incentives and disincentives as an important step forward. One idea even involved awarding a prize, or public challenge, to use TWDB data and demonstrate positive impacts to decision making for a project in Texas.

Finally, participants concluded that the agency in Texas best suited to lead in developing and hosting the key data hub is the TWDB's Texas Natural Resources Information System.

GROUP E

TEXAS USE CASE

INTEGRATE AND UPDATE THE TEXAS WAM AND GAM

Subject	**Integrate and update the Texas Water Availability Models (WAM) and Groundwater Availability Models (GAM)**
Objective	To integrate and update the WAM and GAM to better understand water availability across surface water and groundwater, and across the interface between the two.
Description	Separate models are often outdated, sometimes reverse engineered, and lead to suboptimal results by design. Current models for surface water and groundwater in Texas can be integrated for better results leading to better decision making about water in Texas.
Data Gaps	Need rain map for the ground (i.e., how water moves and accumulates once it hits ground)
Users	All users of state, regional, and local water management plans
Uses	• Provide better tools for decision making and reduce/avoid some costs • Improve state water planning and plans • Provide for more adaptive management • Assist real-estate planning and reduce costs
Participants	See sources
Regulatory	State, regional, and local water management planning
Workflow	• Need a mechanism to bring together data from many sources immediately • Need an organization (assigned or created) to answer data questions for Texas flood emergencies
Sources	Groundwater conservation districts, Texas Water Development Board State Water Plan and Texas Natural Resources Information System, US Geological Survey, floodplain mapping, US Geological Survey Texas water dashboard, Texas Commission on Environmental Quality, The Nature Conservancy Living Waters, Texas Railroad Commission, Texas General Land Office, Texas Department of Licensing and Regulation well licensing, Lower Colorado River Authority Hydromet, TexMesonet, National Weather Service river forecast, US Bureau of Reclamation, US Army Corps of Engineers
Characteristics	Data are available and ready for use today dispersed across many agencies and organizations. These data may be hard to find for most potential users.

SPRINGBOARD TO THE FUTURE – NEXT STEPS

Participants first listed existing data sources and then, considering the list, asked, "what can we do to or with this existing data to improve outcomes for Texans the most." Their answer was to integrate and update the Texas Water Availability Models (WAM) and Groundwater Availability Models (GAM). Participants summed up their reasoning with a problem statement: separate models are often outdated, sometimes reverse engineered, and lead to suboptimal results by design. Thus, the key for the group was not to just have more data, but to have more research, more models, better models, better data sets, maps, and a tool. That tool will allow people to see water availability across surface water and groundwater, and across the interface between the two. This interface is where the greatest optimization of the models will be achieved. With that, the data sets will be optimized and the improvement sought by the participants will be achieved.

The end result is that there will be updated WAM and GAM, and with better models over time the end users, including policy makers, regulators, and water rights holders, will be served better. Participants stated that it is important that this effort be positioned as not changing how water is regulated in Texas. This project would be framed to honor and protect property rights and how water is already being managed in Texas. The tool would allow for better evaluations and decisions; better state, regional, and local water planning and plans; more adaptive and integrated management, and; better tools to avoid costs. This would be a tool that serves a specific purpose. It would also drive traffic to existing data portals from which data will be drawn.

GROUP F

TEXAS USE CASE

PROBABILITY OF RESERVOIR WATER SUPPLIES FALLING

Subject	**Risk management of the probability of reservoir water supplies falling below criteria at 3, 6 , 9, and 12 months**
Objective	Risk management: identify risk of communities' water supplies falling below critical levels
Participants	Primary users: Water Resource Managers, utilities, power agencies – any group that may need to take action based on risk and "triggers"
Regulatory	Water rights in reservoirs and placed in Water Management Plan. The plan is stochastic with water rights defined by TCEQ oversight of court-based adjudication. Focus on permission with constraints.
Workflow	1. Identify potential funding sources 2. Identify possible sellers 3. Identify areas of need, e.g., threatened species 4. Compare historic to current flows Additional actions in no order that may be taken: • Review water rights seniority • Do cost-benefit analysis • Study prior cases • Assess water quality and impacts • Review predictive models • Review TCEQ process for amending water rights • Identify existing environmental flow rights • Estimate flows needed to make a difference
Sources	Texas Commission on Environmental Quality, water rights use, water State river flows and related data sets. Water sources = run of river data

SPRINGBOARD TO THE FUTURE – NEXT STEPS

Participants observed that two key questions need to be addressed in order to form a "springboard" to the future of Texas water data management:

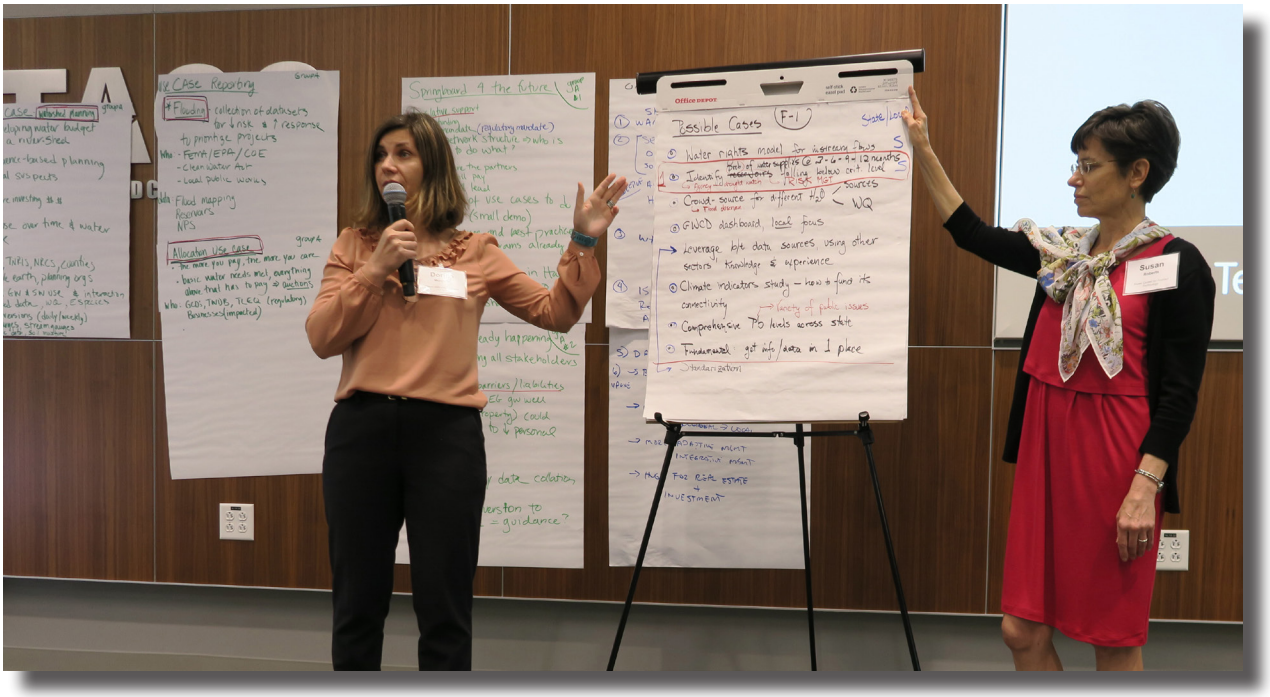
- What agency will be the overseeing entity?
- What entity is going to pay for changes to the existing data management systems?

Participants discussed TWDB and TCEQ, with TWDB's Texas Natural Resources Information System (TNRIS), to lead in developing and hosting the key data hub. USGS was also suggested.

Possible process to form the "springboard" might entail the following:

- Pick one topic / one need that drives an open, connected system.
- Start with the current responsible data agency.
- Build data and metadata of similar quality.
- Survey Texas water agencies and users to find coalescing point and "bundle" an approach to connecting currently unconnected data sets and databases.

COMPLETE LIST OF POTENTIAL USE CASE SUBJECTS



GROUP A

1. Regional water planning
2. Allocation of surface water during drought
3. Flooding (catastrophic) impacts – ecological, economic, social
4. Watershed protection planning (e.g., Rio Grande and interboundary)
5. Options for community water supplies
6. Interbasin water transfer (i.e., San Antonio Water Systems Vista Ridge Project) vs. brackish groundwater desalination vs. new reservoirs
7. Industrial water use during drought
8. Need for more data and transparency of data
9. Water rights - priority of contracts, seniority of right, supply variability, diversion, beneficial use
10. Sales transactions
11. Harris-Galveston Subsidence District and integrated support from regulatory agencies
12. Conservation data
13. Utilities connections

GROUP B

1. Water utility reporting to the Texas Water Development Board (TWDB)

GROUP C

1. Environmental flow transactions
2. Nonpoint source pollution
3. Determination of appropriate groundwater withdrawal and impact on aquifers
4. Best management practices for conservation
5. Recreational use attainability analysis
6. Flood prediction and emergency response
7. Desired future condition for groundwater and predictions
8. Estimation of groundwater availability
9. Impervious cover and regulation

GROUP D

1. Flood water management in ephemeral streams
2. Planning for drought
3. Environmental flows
4. Climate impacts to Texas hydrology
5. Water quality in the context of consumptive use

GROUP E

1. Integrate and update the Texas Water Availability Models (WAM) and Groundwater Availability Models (GAM)

GROUP F

1. Water rights model for instream flows
2. Flood observations: crowd-source for different water sources and water quality
3. Groundwater Conservation District dashboard
4. Standardization - leverage between data sources using other sectors' knowledge and experience
5. Climate indicators study – how to fund its connectivity to statewide water resources concerns
6. Comprehensive lead (or other potential contaminants) across the state

WATER DATA WORKSHOP

APPENDIX IV

GLOSSARY

GLOSSARY

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

Data hub - An independent location or system where data is stored that connects to data from multiple sources, while maintaining the autonomy of the independent location or system.

Data gap - Where information critical to decision making is either not available at all, or where information exists or is available but is not in a suitable format or accessible for decision making processes or other uses.

Data system - Software or hardware that is used to collect, organize, archive, distribute, or integrate data.

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

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

Information system - Software or hardware that is used in the processing, analysis, or synthesis of data so they can be used to answer questions.

Interoperability, interoperable - The ability of multiple computing or other information management systems to operate on the same data and produce the same analysis or results.

Metadata - Data that describe and give information about other data.

Open - The ability to have access to data using open-source and open-architecture protocols and methods.

Stakeholder - Anyone with an interest in the outcomes of Texas' progress on water data, including data users and data producers from relevant sectors of government, industry and civil society.

Water security - The ability to access water at sufficient quantity and quality to sustainably meet agricultural, ecological, industrial, military, public health, sanitary, and urban needs.

Water data - Quantitative or qualitative representations or measurements of properties of water or water related measurements.

Use case - A short summary organized in a fashion that helps list in a concise and consistent format the data gaps, needs, and uses for a particular objective. The objective is what decision, action, or other thing needs to be accomplished. For the workshop this can be a need of data managers, providers and/or data consumers. A use case communicates a set of answers to the question of, *who* needs *what type* of data in *what form* to make *what decision*(s). Use cases will support display of a water decision making process and the data needs associated with that process.

WATER DATA WORKSHOP

APPENDIX V

REFERENCES

Reference Material

Imagine an Internet of Water

If we've learned anything from the Internet, it is that we are not likely to imagine how it will be used nor what people will find valuable and important. In the same way, it is more likely that the Internet of Water will enable innovations that are not imaginable now, hopefully toward a far more sustainable water future.

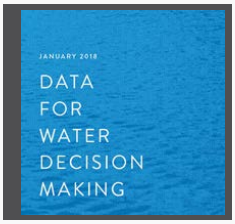
This web article sheds light on what the future may hold.

Click to Read Web Article

https://www.aspeninstitute.org/aspen-journal-of-ideas/imagine-internet-water/

Data for Water Decision Making:

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



A lack of data and information has limited our ability to understand, let alone better manage, all aspects of our water resources. This report and case studies published in January 2018 support California's efforts to develop modern water data systems. It argues that simply providing more data is not enough, and that generating useful and useable information hinges on the development of data systems based on end users' needs. The report describes lessons learned from a process of stakeholder engagement focused on defining and clarifying uses of water data, and how knowledge of these uses can inform the development of water data systems.

Click to Download the Report

Click to Download Use Cases

Report -- https://www.law.berkeley.edu/wp-content/uploads/2018/01/DataForWaterDecisionMaking.pdf
Use Cases -- https://www.law.berkeley.edu/wp-content/uploads/2018/01/DFWD-Use-Cases.pdf

Aspen Institute Report – Internet of Water:

Sharing and Integrating Water Data for Sustainability

Between May 2016 and February 2017, the Aspen Institute Dialogue Series hosted several roundtables with a select group of water experts, managers, policy makers, regulators, and representatives from the private and social sectors to focus on how to create better water data infrastructure to access and connect publicly collected and reported sources for data, beginning with quantity, quality, and use information.

This report highlights and provides a principle-based blueprint recommending a 3-step plan for how to design and launch a feasible and operable "Internet of Water."

Click to Download the Aspen Report

https://assets.aspeninstitute.org/content/uploads/2017/05/Internet-of-Water-Report-May-2017.pdf

Texas Water Roadmap Forum:

Workforce Education, Data, and Research

Three forums were held between February 2015 and November 2016, bringing together Texas water experts from business, industry, government, academia, research, and the investment community in impartially facilitated sessions to determine ways to secure Texas' water future through accelerating growth of infrastructure, technologies, research, education, and sustainable use. The final forum focused in on data access and management, with recommendations and a suggested path forward.

This report details the findings of Texas water experts.

Click to Download the Water Forum Report

http://libguides.tamusa.edu/ld.php?content_id=28446621

WATER DATA WORKSHOP

APPENDIX VI

WORKSHOP PARTICIPANTS

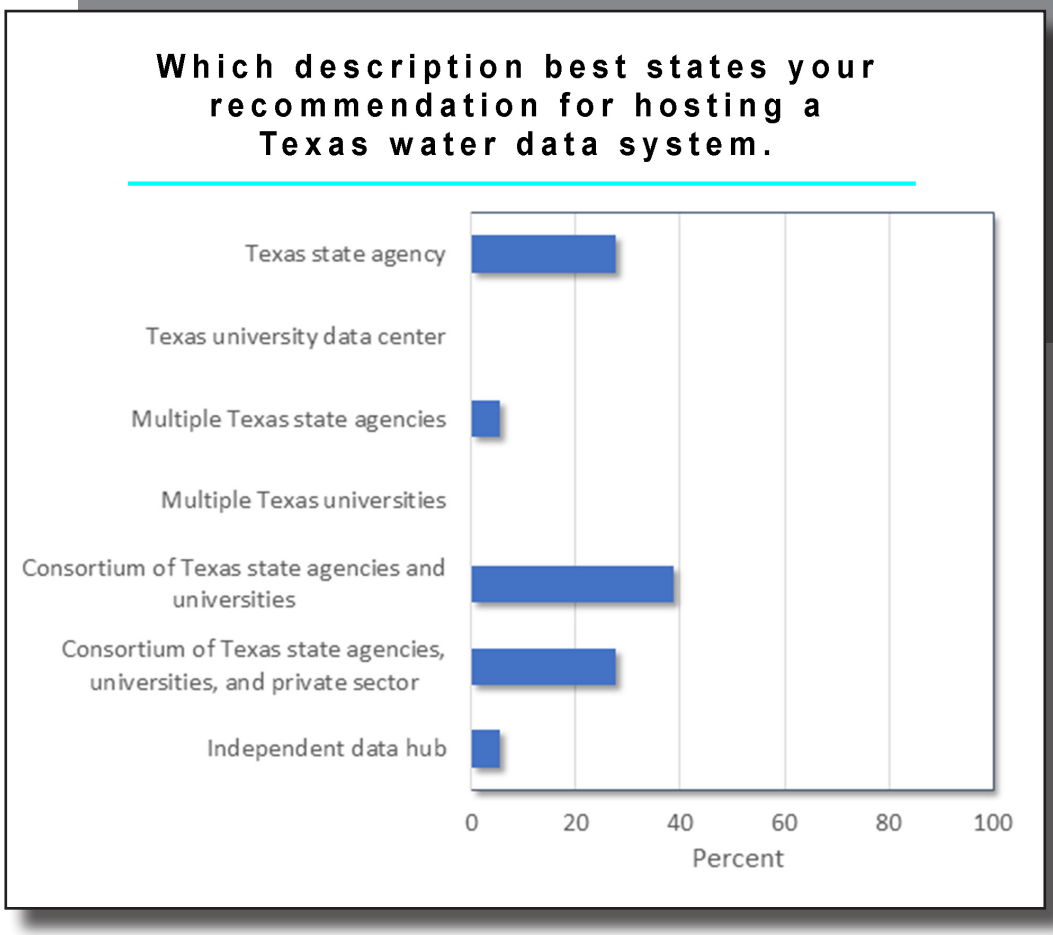
First Name	Last Name	Representing	Email
Dirk	Aaron	Clearwater UCD	daaron@cuwcd.org
Josh	Adler	Source Water	Josh@sourcewater.com
Shumon	Alumon	Prairie View A&M University	shalam@PVAMU.EDU
Kip	Averitt	Averitt & Associates	kip@averittandassociates.com
Carole	Baker	Texas Water Foundation	cbaker@texaswater.org
Ryan	Bare	Houston Advanced Research Center	rbare@harcresearch.org
Kelly	Bennett	B3	Kbennett@b3insight.com
Bill	Billingsly	Texas Water Development Board	Bill.Billingsly@twdb.texas.gov
Jamie	Burke	AECOM	jaime.burke@aecom.com
Todd	Burrer	Inframark	tburrer@inframark.com
Susan	Butler	CH2M	Susan.Butler@CH2M.com
Justin	Camp	Barton Springs Edwards Aquifer Conservation Dist.	jcamp@bseacd.org
Adele	Cardenas	USEPA Region 6	cardenas.adele@epa.gov
Keith	Cole	Water Lens	kcole@waterlensusa.com
Margaret	Cook	The University of Texas at Austin	margaretcook@utexas.edu
Quenton	Dokken	Tarleton State University	dokken@tiaer.tarleton.edu
Andrew	Donnelly	Daniel B. Stephens & Assoc., Inc.	adonnelly@dbstephens.com
Chris	Dorow	BASF	christopher.dorow@basf.com
Martin	Doyle	Duke University	martin.doyle@duke.edu
Paul	Faeth	Cadmus Group	Paul.Faeth@cadmusgroup.com
Adeline	Fox	Texas Water Conservation Association	afox@twca.org
Natalie	Freed	University of Texas Austin, TACC	saplingsonwheels@gmail.com
Jordan	Furnans	LRE Water LLC	jordan.furnans@lrewater.com
Marcus	Gary	Edwards Aquifer Authority	mgary@edwardsaquifer.org
Tom	Gerik	Texas A&M AgriLife Research-Blackland	t-gerik@tamu.edu
Yolanda	Gil	Information Sciences Institute	gil@isi.edu
James	Gray	Data.World	james.gray@data.world
Karen	Guz	San Antonio Water Systems	Karen.Guz@saws.org
Maurice	Hall	Environmental Defense	mhall@edf.org
Rich	Haut	Houston Advanced Research Center	rhaut@harcresearch.org
Sam	Hermitte	Texas Water Development Board	Sam.Hermitte@twdb.texas.gov
Ben	Hodges	Univ. of Texas - Austin	hodges@utexas.edu
Kathleen	Jackson	Texas Water Development Board	Kathleen.Jackson@twdb.texas.gov
Narayanan	Kannan	Tarleton State University	kannan@tiaer.tarleton.edu
Kathy	King	Redstone Strategy	KathyKing@redstonestrategy.com
Brant	Konetchy	WSP	brant.konetchy@wsp.com
Ken	Kramer	Sierra Club	kenwkramer@aol.com
Sara	Larsen	Western States Water Council	saralarsen@wswc.utah.gov
Sharlene	Leurig	Meadows Center	eFlows@txstate.edu
Cindy	Loeffler	Texas Parks and Wildlife Dept.	Cindy.Loeffler@tpwd.texas.gov
Glen	Low	Earth Genome	glen@earthgenome.org
Robert	Mace	Meadows Center	rem142@txstate.edu
David	Maidment	Univ. of Texas - Austin	maidment@utexas.edu
Justin	McInnis	Hays County	justin.mcinnis@co.hays.tx.us
Jordan	Merson	San Antonio River Authority	jmerson@sara-tx.org
Binayak	Mohanty	Texas A&M University	bmohanty@tamu.edu
Stephanie	Moore	Daniel B. Stephens & Associates	smoore@dbstephens.com
Dan	Mueller	Environmental Defense Fund	dmueller@edf.org
Dorina	Murgulet	Texas A&M Univ. - Corpus Chraisti	dorina.murgulet@tamucc.edu
Mike	Myatt	Water Foundation	Mmyatt@waterfdn.org
Mike	Ouimet	Texas Department of Public Safety	michael.ouimet@dps.texas.gov
Lauren	Patterson	Duke University	lauren.patterson@duke.edu
Leslie	Patterson	Texas Commission on Environmental Quality	leslie.patterson@tceq.texas.gov
Maguel	Pavon	Texas Water Development Board	miguel.pavon@twdb.texas.gov
Daniel	Pearson	US Geological Survey	dpearson@usgs.gov
Trino	Pedraza	New Braunfels Utilities	tpedraza@nbutexas.com
Suzanne	Pierce	Texas Water Research Network	spierce@tacc.utexas.edu

Herman	Ramsden	UT Rio Grande Valley	herman.ramsden1@utrgv.edu
Ruthie	Redmond	Sierra Club	ruthie.redmond@sierraclub.org
Sarah	Richards	Cynthia & George Mitchell Found.	srichards@cgmf.org
Susan	Roberts	Texas Center for Applied Technology	svroberts@tamu.edu
Rudolph	Rosen	Inst. Water Resource Science and Tech.	rudy.rosen@tamusa.edu
Carlos	Rubinstein	RSAH2O	carlos@rsah2o.com
Leslie	Savage	Railroad Commission of Texas	leslie.savage@rrc.texas.gov
Sarah	Schlessinger	Texas Alliance of Groundwater Districts	sarah@texasgroundwater.org
Spencer	Schnier	Freese and Nichols	spencer.schnier@freese.com
Stefan	Schuster	SWCA Environmental Consultants	sschuster@swca.com
Julie	Sommerfeld	Bastrop County	julie.sommerfeld@co.bastrop.tx.us
Robert	Stefani	Austin Water Utility	Robert.Stefani@austintexas.gov
Carrie	Thompson	Collaborative Water Resolution	carrie@watertbl.com
Ernest	To	Alan Plummer Associates	eto@apaienv.com
John	Tracy	Texas Water Research Institute	john.tracy@ag.tamu.edu
Mark	Treviño	Bureau of Reclamation	mtrevino@usbr.gov
Joe	Trungale	Trungale Engineering & Science	joe@trungaleengineering.com
Charlie	Upshaw	Webber Energy Group	crupshaw@utexas.edu
Michael	Urrutia	Guadalupe-Blanco RIver Authoirty	murrutia@gbra.org
Anastasia	Valdes	Water Markets LLC	anastasia@watermarkets.us
Todd	Votteler	Collaborative Water Resolution	votteler@waterdisputes.org
David	Walker	Lower Colorado River Authority	david.walker@lcra.org
Jennifer	Walker	Texas Living Waters Project	jennifermwalker@earthlink.net
Emily	Warren	Meadows Center	EmilyW@txstate.edu
Carl	Westergard	Guadalupe-Blanco RIver Authotity	cwestergard@gbra.org
Jennifer	White	Texas Water Development Board	Jennifer.White@twdb.texas.gov
June	Wolfe	Texas A&M AgriLife Research-Blackland	jwolfe@brc.tamus.edu
Corinne	Wong	Univ. of Texas - Austin	ciwong@austin.utexas.edu
Mike	Woodside	U.S. Geological Survey	mdwoodsi@usgs.gov
Michael	Young	Bureau of Economic Geology - UT	michael.young@beg.utexas.edu

APPENDIX VII

PARTICIPANT SURVEY

Summary



Following the workshop, participants were invited to participate in a survey where they were presented three questions.

The first sought recommendations on the next steps we should take as a group or as individuals, agencies, companies, or institutions. There were 19 responses. Participants gave varied answers ranging from keeping the workshop group together and refining or organizing what was initiated, to involving more participants from agencies, more computer scientists, more private sector representation, and having agencies form water data working groups. Some participants stated that better defining or a narrowing of the intended audience is needed, while others suggested starting with small steps and expanding over time. Several participants stated that proper planning processes needs to begin with reviews conducted and goals, measurable benchmarks, and protocols set. A need for funding was mentioned by several participants. Recommendations on technical aspects of creating data hubs were given.

Continuing the development of use cases and initiating example projects to demonstrate the value of open data hubs came up in several recommendations. One participant stated, "Develop a program that directly involves and engages all stakeholders in a way where sharing data provide benefits to all. Start small on a project(s) that are doable and show value in broader data collation and distribution." Another stated, "I felt that the use cases discussed at the workshop are very insightful and I think it would be helpful to take a closer [look] on the technical side of how agencies are implementing data-sharing." One simply stated, "Pilot a couple of use cases to show the value."

The second question sought to assess opinion on hosting options for open data hubs or systems. The results clearly show respondents rejected hosting by a single university or multiple universities. Instead, respondents were almost evenly spilt over hosing by (1) a Texas state agency, (2) a consortium of Texas state agencies and universities, and (3) a consortium of Texas state agencies, universities and the private sector.

Respondents were also invited to explain their choice should they desire. Of the 18 participants who answered this question, 13 added an explanation. Many stated that state and federal agencies already serve to host data, thus any answer to this question must include agencies as a host. This may explain why universities alone were not selected as the preferred host by any respondent. Following is a response that supports that conclusion: "My opinion is that each data provider needs to maintain the fidelity of their own information on an ongoing basis." Please refer to the raw answers for listings of data hosting agencies mentioned. This response may further explain the rationale, "I think a state agency such as TNRIS would be the natural choice. Data-sharing involves curation and database maintenance and may not fit into the research agenda of universities. Universities can however serve as a technological partner."

Finally, participants were asked for additional thoughts on any matter they felt appropriate. One respondent reiterated continuing on with development of use cases. Others suggested expanding participation, especially involving members of the general public, and carrying the workshop to other areas of the state. There were several comments accompanied by a sense of urgency to carry on.

Finally, several respondents simply stated their appreciation for the workshop and the organized manner in which the workshop proceeded.

WATER DATA WORKSHOP

PARTICIPANT SURVEY RESPONSES

SURVEY QUESTION 1

What comes next is of critical importance to furthering efforts to connect water data in Texas, and all workshop participants may not have had the opportunity to share all their ideas. Please offer your recommendations on what we – as a group or as individuals, agencies, companies, or institutions – should do next.

RESPONSES

1. The workshop gave a few excellent answers to that question. Since the workshop was well organized to get such answers, continue with the group and narrow down the suggestions, then organize to try it.
2. As a group, we need to make sure we have clearly defined the intended audience(s) for the data and then consider the ways in which the data should/must be presented to each audience. While raw data is great for those that understand it, raw data is useless to an audience that requires context and a bit of interpretation. We may find that we need to provide the same data in multiple ways, depending on the audience. As a group, data collection methods will need to be determined. Will only existing data be harvested from their sources to create the new repository? Will we ask/require individuals, agencies, companies, and/or institutions to begin submitting data in a new format? Or, maybe the existing data isn't harvested at all and instead, API's are written to query the data already out there and present it in a meaningful way to the audience? With an API-only approach, the hosting question is solved. The data stays where it is already located. The comment, raised during the workshop, that "you have to be a data expert to get to the data now" would be absorbed with an all API approach. However, that would require a ton of sophisticated code to be written.
3. We need to involve computer scientists. Instead of trying to roll out an "internet of water" all at once, it is prudent to start small and imminently doable. For example, a simple webpage with links to where to download existing datasets. This doesn't require much technical know-how. The very first websites on the internet were no more complicated than this. Once we start scratching the surface, the next steps become easier to envision and execute. This also guarantees that the undertaking is not an all-or-nothing proposition; we will get usable results from the 'Connecting Texas Water Data' project immediately.
4. My suggestion would be to query the agencies in Texas that provide the most water-related natural resource information on whether or not they would be amenable to forming an "open water data" workgroup that could lay the groundwork for more discussion and greater integration of efforts on this front. Everyone seems to be coming around to the notion of transparency, but it's getting a meaningful conversation going that is the hard part. I bet TWDB, TCEQ, et al., would be willing to convene on some ongoing basis on this topic and jointly address the issue.
5. Continue having open conversations but work towards a goal. Have measurable benchmarks to meet.

6. Recommend: 1) Identify key datasets and hubs. Inventory 2) Understand which agencies/ institutions are essential for create a scale effort on water data in Texas and have them be a founding coalition
7. TCEQ and TWDB should work on systems to automate the submittal of annual Water Use Reports to TCEQ and annual Water Use Surveys to TWDB by water users so that data can be automatically loaded into agency databases. The state should look at ways to cross reference data from the two reports to get a better picture of what sources and uses of water in the state.
8. The hosting groups should help narrow the top 5 needs for Texas water data that were presented during the workshop. From there, subcommittees could be formed for each topic and those groups could focus on one particular issue to address. Participants could express interest in working on a subcommittee and rank their first-fifth choice.
9. Develop guidelines/protocols for integrating data, to allow for federation among datasets; pilot a couple of use cases to show the value.
10. I'm struck by how much of the discussion was public policy driven. Any solution needs to go beyond government only and actively seek participation from the private sector.
11. Expand/improve existing state water data hubs at TWDB/TNRIS/TCEQ. Link information between these agencies. Provide a dashboard portal for other agencies (GCDs, River Authorities, universities, etc.) to upload data to existing hub and include some primary level of data quality review.
12. I think there should be a review of what technologies have been developed in sharing data, e.g. web services, interactive maps, etc. Advances have been made in the past decade through organizations like CUAHSI, ESRI, etc. Even though the general public may not be aware of those technologies, many agencies have been adopting them. I felt that the use cases discussed at the workshop are very insightful and I think it would be helpful to take a closer on the technical side of how agencies are implementing data-sharing. This can help audience understand what everyone else is doing, where the low hanging fruits are, and how to prioritize strategies for data-sharing.
13. Texas Agencies and Universities should be funded and spearhead the effort. Consistent and long-term funding must be available to collect specific data across Texas and process all kinds of water data.
14. Develop a program that directly involves and engages all stakeholders in a way where sharing data provides benefits to all. Start small on a project(s) that are doable and show value in broader data collation and distribution. A single warehouse of data will extremely difficult to manage and indeed those data are already housed at different state agencies. Perhaps a web-based system would work, in which agencies keep their data, but in a format that's accessible to anybody with internet and correct scripting languages.
15. Identify early wins - prioritize action items.
16. Create a crosssectional committee to review the recommendations produced by the meeting and propose the next steps.
17. Link all available data through one portal. Include critical metadata describing data source, range, quality, appropriate uses, and cautions; note question 2.
18. How can we increase the water supply in unconventional ways?

SURVEY QUESTION 2

Several participants suggested hosting a data system at the Texas Water Development Board, such as through the TWDB's Texas Natural Resources Information System, or through the Texas Commission on Environmental Quality, or at a university data center such as the Texas Advanced Computing Center, or through some combination of state agencies and universities. Others suggested a much more distributed approach to hosting. Which description(s) below best states your recommendation?

ANSWER CHOICES	RESPONSES	
Texas state agency (If so, which one or ones? Please specify.)	27.78%	5
Texas university data center (If so, which one or ones? Please specify.)	0.00%	0
Shared responsibilities between Texas state agencies	5.56%	1
Shared responsibilities between Texas universities	0.00%	0
Consortium of Texas state agencies and universities	38.89%	7
Consortium of Texas state agencies, universities, and private sector	27.78%	5
An independent data hub (If so, please describe the nature of such a hub.)	5.56%	1
Total Respondents: 18		

ADDED SUGGESTIONS

1. I am uncomfortable suggesting a data host until a full understanding of the intended audience is presented. The data host should somehow be aligned with, and responsible to, the audience.
2. Of all the state agencies providing data in Texas, the TWDB does the best job in terms of ease of accessing the data (finding it, downloading it, and getting it in a usable format). But even within the TWDB, the data is stored in disparate places and can be difficult to find if you don't know exactly what you're looking for. The data storage mechanism for the "internet of water" should mirror the internet itself, as distributed as possible. There are many concerns with storing the data at a centralized site such as a state agency, not least of which is vulnerability to political whims (as we have seen recently at the national level).
3. My opinion is that each data provider needs to maintain the fidelity of their own information on an ongoing basis. They can do this in many ways, including hosting from each

agency/university and making the data accessible to inclusion in other hubs via web services. Other alternatives can include provisioning of a "shared space" for data with the TACC (or other), or a trusted cloud vendor (although this brings with it other concerns). It helps to have a governmental body or advisory body that is viewed as a trusted partner to assemble the different options and present them to the data providing parties.

4. Use the brand and the backing of the State by using agencies like TWDB/TNRIS with the power of TACC.
5. Many different agencies and organizations already host their own datasets. That's unlikely to change, especially since many of them may have already invested or are considered the 'authoritative' source for that info. Better to connect these in a federated approach
6. Federal - EPA, FEMA, USGS, NOAA State - TWDB, TCEQ. Any university and private sector entity that is interested in participating. Probably other agencies/entities that I am not aware of.
7. Some sort of coordinated effort between TWDB (quantity) and TCEQ (quality)
8. Raw data can be hosted and supported by a public entity (state agency, university), they are public records after all. The key challenge is not in an index of public raw data it's in the processing and standardizing of the data which effectively takes the data from public to proprietary. It's a lot of work to standardize data, which is essential for comparable study. I'd be surprised if any private venture would participate if providing data would make it effectively available to anyone via an open records request. An independent data hub could be more successful at attracting data wrangling and analytic solutions with contractual clauses limiting access. There may be a tendency of agencies and universities to view this as unimportant when measured against the public resources available to subsidize this kind of effort. I suggest 2 reasons to consider proprietary data. First, if the data doesn't have a market then it begs the question why spend the money to host it in the first place. Second, some of the most critically needed information will come from private actors and they will need the confidence that sharing the data will not produce liability for them, for example water quality. Only an independent data hub can satisfy these needs...
9. TWDB/TNRIS. They already have systems in place. These systems can be expanded/improved, but already have a foundation to build from.
10. I think a state agency such as TNRIS would be the natural choice. Data-sharing involves curation and database maintenance and may not fit into the research agenda of universities. Universities can however serve as a technological partner.
11. TWDB - they will need to staff-up to do this. Having one unbiased place that already receives some of the data will minimize overlap and provide focus direction.
12. TNRIS
13. Texas Water Development Board Already represents most complete and best documented source. Seek to expand both scope and funding support.

SURVEY QUESTION 3

Please share any additional thoughts or suggestions that you may have regarding the topics discussed at the workshop.

ADDITIONAL THOUGHTS

1. If members of the general public are going to use the Internet of Water, should they be represented in and of the workshops going forward?
2. Would very much like to see additional and fleshed-out documentation for the case studies that were identified in the workshop. It would be good to identify 2-3 top priority or low-hanging-fruit use cases that could be jointly worked on by Texas data providers to show interest and momentum. They don't have to be terribly complex, and can take advantage of existing data sources if they're available, just need to demonstrate that working together to satisfy a use case or two is possible.
3. For such a short period of time a lot happened. Well organized and focused, unlike so many meetings. Looking forward to promised follow up or report.
4. Great job. Thanks for putting this together.
5. Need very clear next steps and clarity on participants, roles, and funding. Need concrete progress to keep the momentum going.
6. A contact list of all who attended would be very helpful as this event was very much a networking and brainstorming workshop. Also, a summary of discussion topics would be great.
7. I was very impressed with the summit. Thank you for allowing me to participate.
8. Texas water is critical and so investment must be made to preserve and utilize it for future purposes.
9. Start simple; show value; broaden the engagement of groups across the state.
10. Much work needs to be planned out and set in motion. TWDB will be a good data repository. Still need to build the framework to make data-driven decisions possible.
11. Reiteration: build a data portal that links to other data sources and provides strong descriptions of data available at the linked site; note answer one.
12. Perhaps to hold in Houston?

WATER DATA WORKSHOP

APPENDIX VIII

PARTICIPANT WORKSHOP TEMPLATES, USE CASE GUIDANCE

Workshop Templates

Use Case Guidance

Texas Water Data Workshop

- Work Group Templates
  Download Interactive Templates Here:
  https://data.water-texas.org/interactivetemplates.pdf

BIG PICTURE

Participant Name	_____		
Who Needs	**What Data**	**In What Form**	**For What Decision**

DATA GAPS & ACCESS

Participant Name _____

Your experience with situations in Texas water that arise from lack of data, or lack of accessible data?

Your vision for desired future of water data management in Texas:

Your list: key attribute of an open access data / information system:

Use Case Template

To help organize and make a clear case for improved access and use of data to manage water supplies in the future, efforts of the Aspen Institute and others have developed a "use case" model that serves as a useful tool for organizing and assessing stakeholder data needs and communicating those needs to decision makers.

To begin working in this direction, participants in the Connecting Texas Water Data Workshop will begin the process of building use cases by helping identify the top ten or twenty possible examples of gaps in data availability, access, and integration that impede decision-making. To achieve this, workshop participants should have a good conceptual understanding of use cases designed to inform decision making. Participants will be supplied with a model template to build a well-organized use case and will have opportunity to look through samples of use cases already developed for application elsewhere.

Definition, Model, Examples, and Template

A use case is a short summary organized in a fashion that helps list in a concise and consistent format the data gaps, needs, and uses for a particular objective. It communicates a set of answers to the question of, who needs what type of data in what form to make what decision(s)? They also provide a way to identify critical data sources or sets where interoperability is important. We envision that use cases will be responsive to stakeholder data needs, as well as useful for technical developers seeking to better understand the data needs of system users. While there can be numerous ways to display a use case, we will follow the model below and provide a blank use case template for use by participants at the workshop.

Examples

https://www.law.berkeley.edu/wp-content/uploads/2018/01/DFWD-Use-Cases.pdf)

Model use Case Template and Explanation (see following pages)

<table>
<tr><td>Objective</td><td>The objective is the decision, goal, or desired action to be achieved. The objective describes what is to be accomplished.</td></tr>
<tr><td>Description</td><td>The description is any defining information about context and background that might help a reader understand the objective or added details of the topic in general.</td></tr>
<tr><td>Participants</td><td>The participants section provides a list of the main decision-maker (s) and other key parties involved or affected. Attributes or contact information for participants may be listed here, if desired.</td></tr>
<tr><td>Regulatory Context</td><td>The regulatory context lists any laws, statutes, rules, regulations, reporting requirements, legal operational constraints, and governmental agency programs either existing or under development. This category may also include boundaries, for example geographic borders, time-based reporting requirements, and financial limits.</td></tr>
<tr><td>Workflow</td><td>Workflow describes the steps, listed as specific actions and in order of occurrence if possible, to be taken by the participants in order to accomplish the objective.</td></tr>
<tr><td>Data Sources</td><td>Data sources are the repositories, locations and holders of recorded measurements or properties collected and assembled about water. Data sources may be well defined and assembled or largely unconsolidated. Data gaps desired to be filled may be listed here as well as existing data. The sources should be listed in sufficient detail to be identified and located. Here is a table format that can be used for this purpose and nested here or added as a separate table.
<table>
<tr><th>Data Category</th><th>Description</th><th>Data source</th><th>Access Method</th></tr>
<tr><td>Water</td><td>Water availability</td><td>USGS web site for gage data</td><td>https://wdr.water.usgs.gov/</td></tr>
<tr><td>Agriculture</td><td>Evapotranspiration</td><td>Texas Water Development Board Precipitation and Lake Evaporation Data</td><td>http://www.twdb.texas.gov/ surface water/conditions/evaporation/</td></tr>
<tr><td>Infrastructure and utilities</td><td>Records of electricity used for pumping</td><td>Data collected by permittee</td><td>Not available at aggregate level—data collected for each individual case</td></tr>
<tr><td>Land use</td><td>Aerial photos</td><td>Satellite imagery - Google Earth</td><td>https://www.google.com/earth/</td></tr>
</table></td></tr>
<tr><td>Data Characteristics</td><td>Data characteristics includes notes about the type, form, and format of data that would be most useful for making decisions, including anything out of the ordinary about the data.</td></tr>
</table>

Workshop Participant Name	
Objective	
Description	
Participants	
Regulatory Context	
Workflow	
Data Sources	
Data Characteristics	

Data Category	Description	Data source	Access

SPRINGBOARD TO THE FUTURE

Participant Name _____

In your view, what are the next steps for water data management in Texas?

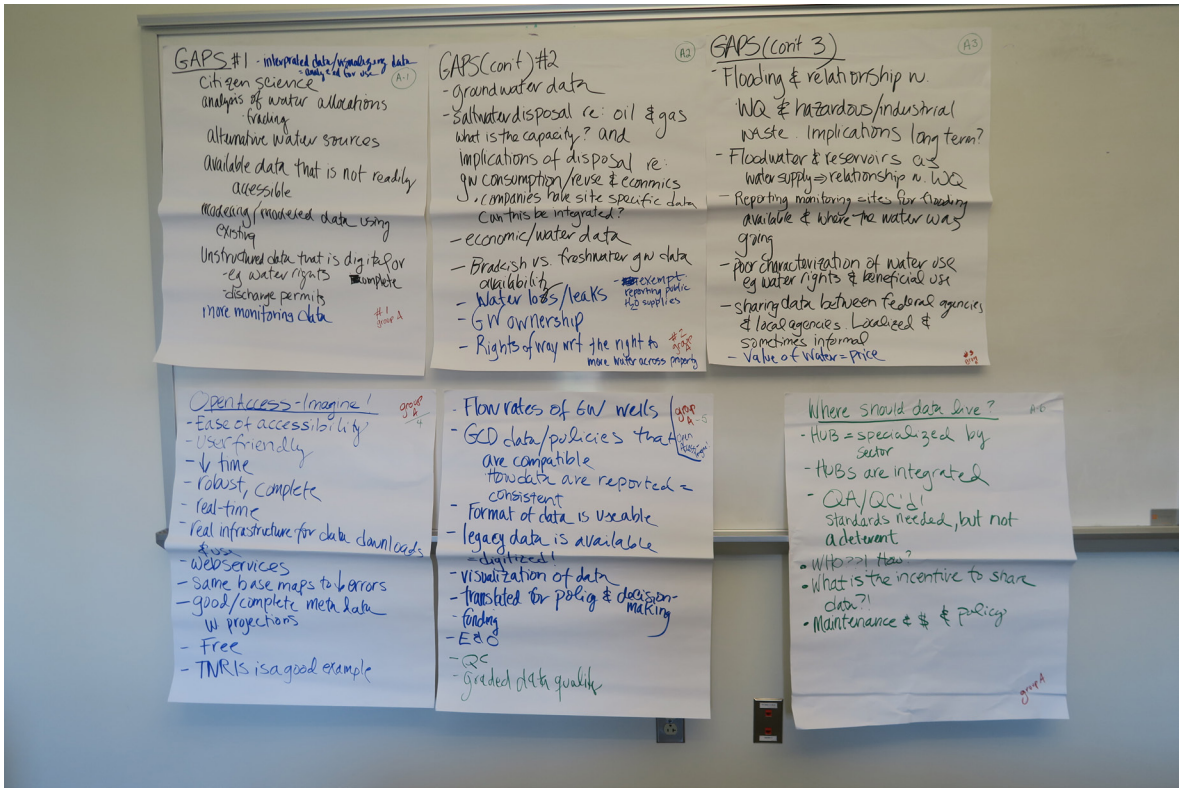
WATER DATA WORKSHOP

APPENDIX IX

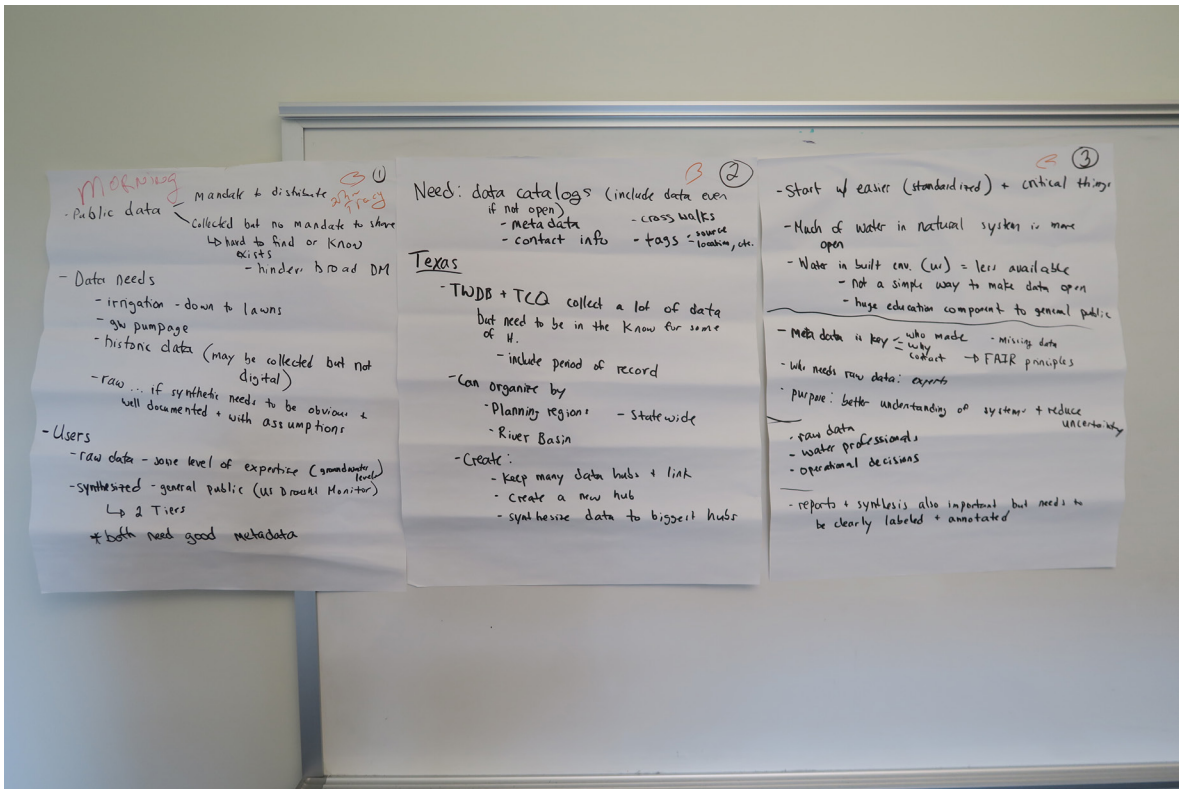
RAW DATA FROM BREAKOUT SESSIONS

GROUP SESSIONS I

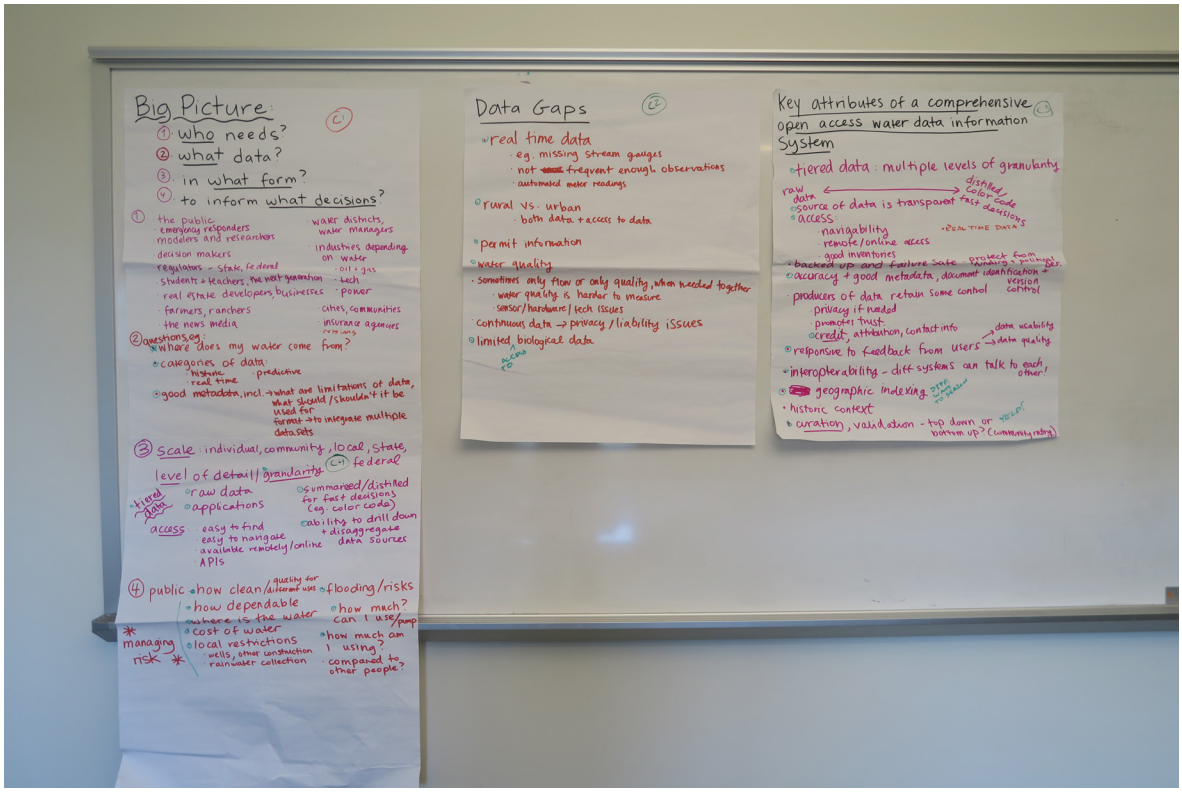
A



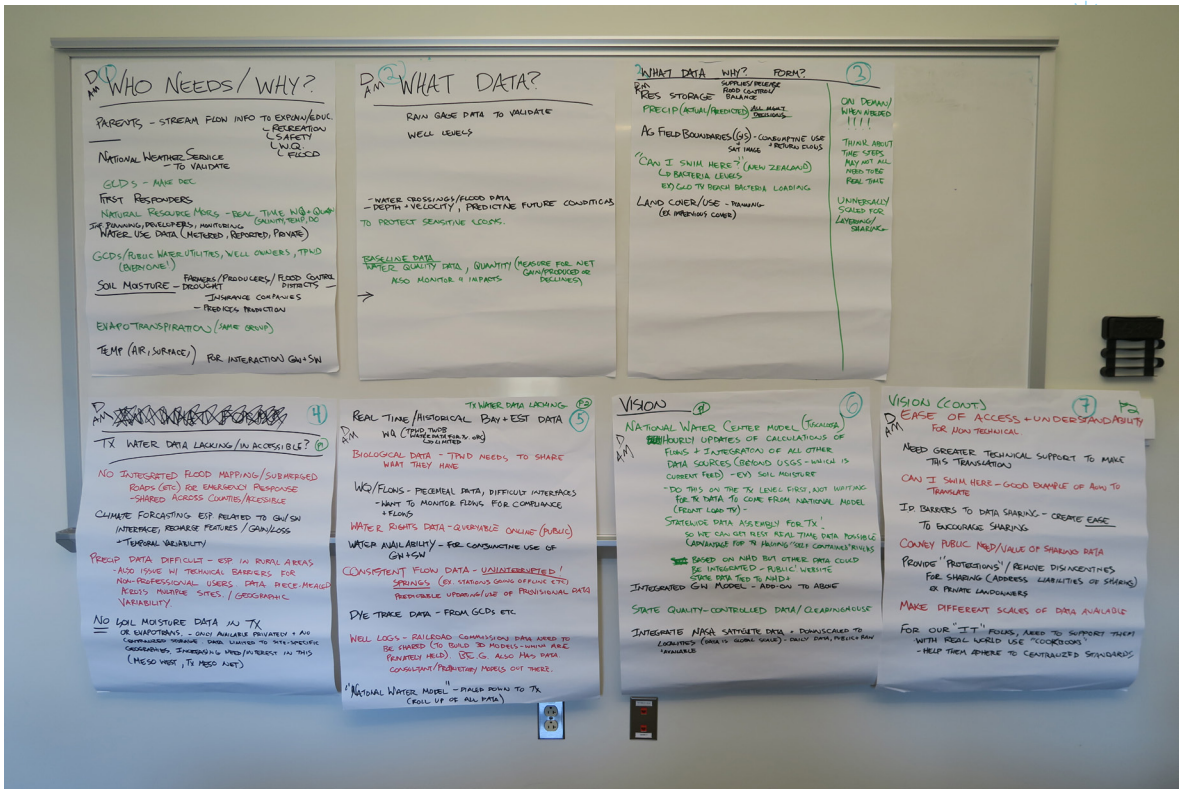
B



C

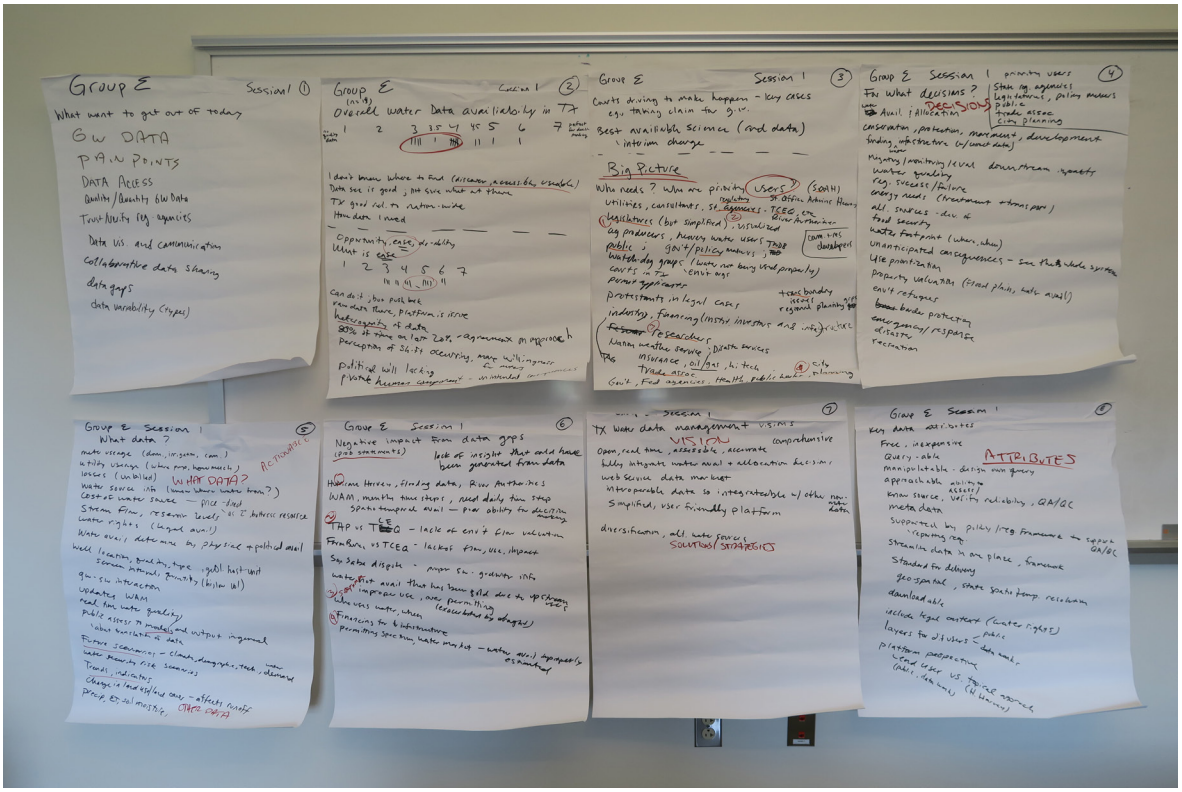

Big Picture:
① who needs?
② what data?
③ in what form?
④ to inform what decisions?
Data Gaps
real time data
eg. missing stream gauges
not frequent enough observations
automated meter readings
rural vs. urban
both data + access to data
permit information
water quality
Sometimes only flow or only quality, when needed together
water quality is harder to measure
sensor/hardware/tech issues
continuous data → privacy/liability issues
limited biological data
Key attributes of a comprehensive open access water data information System
tiered data: multiple levels of granularity
access:
navigability
remote/online access
good inventories
accuracy + good metadata, document identification + version control
producers of data retain some control
privacy if needed
promotes trust
credit, attribution, contact info
responsive to feedback from users
interoperability - diff systems can talk to each other!
geographic indexing
historic context
curation, validation - top down or bottom up?

D

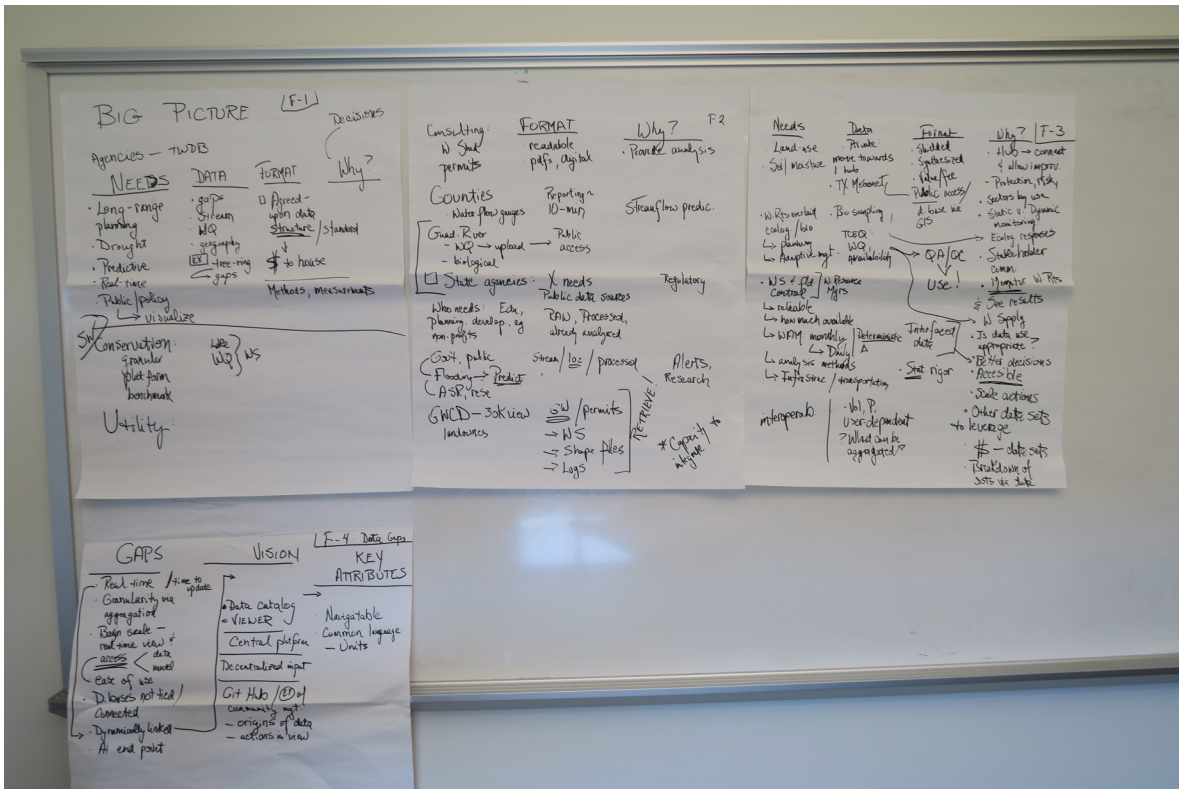

WHO NEEDS/WHY?
WHAT DATA?
WHAT DATA WHY? FORM?
VISION
VISION (CONT)
EASE OF ACCESS + UNDERSTANDABILITY

GROUP SESSIONS I

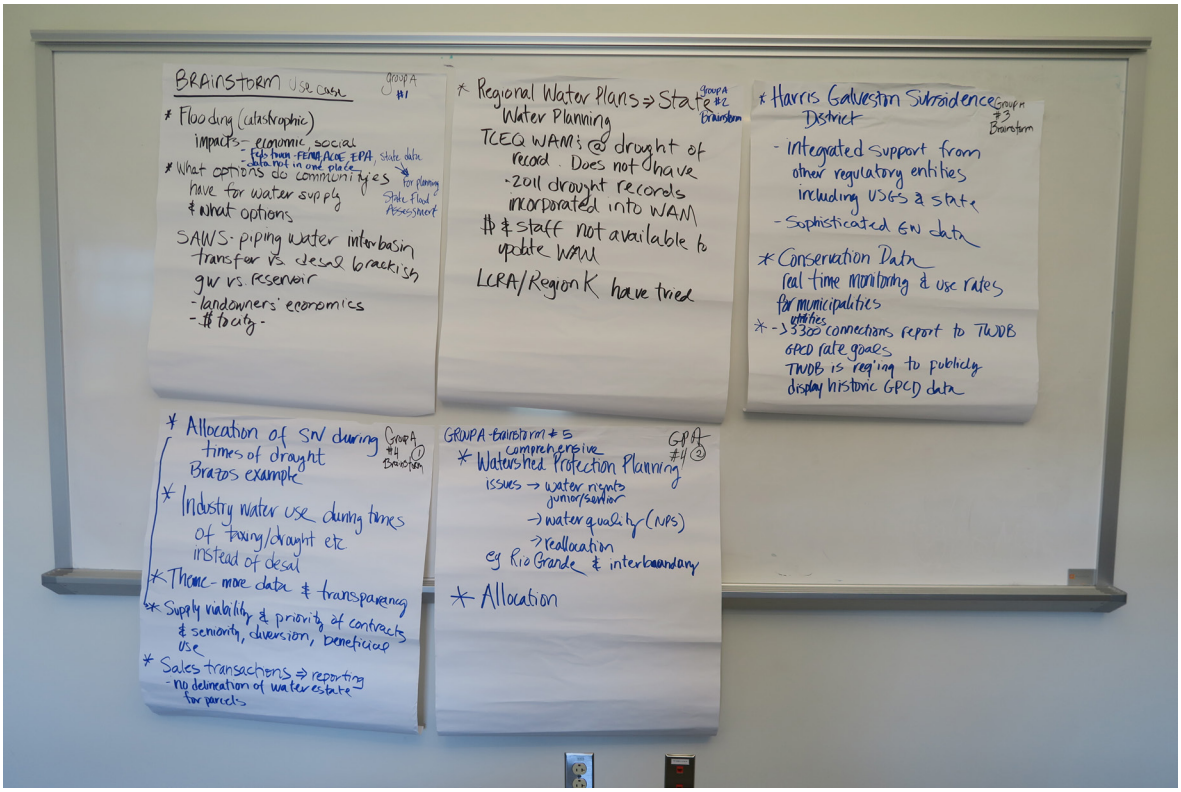
E



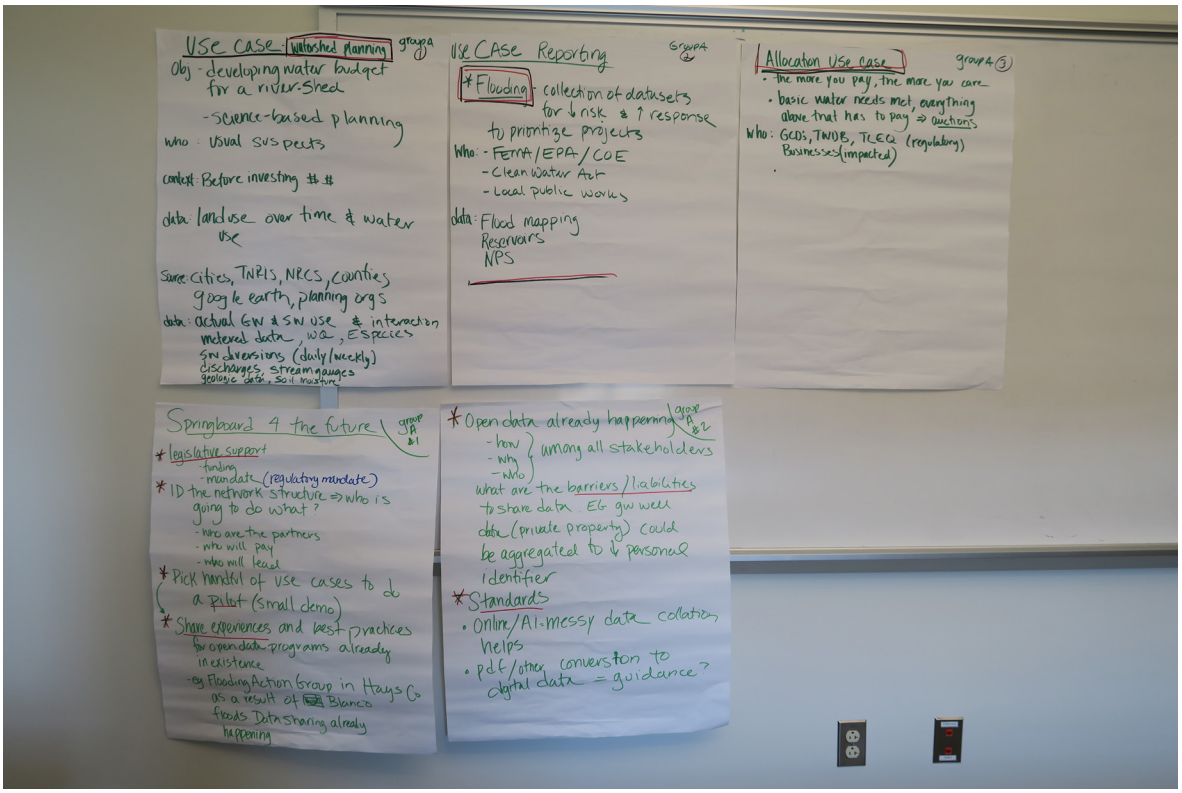
F



A1

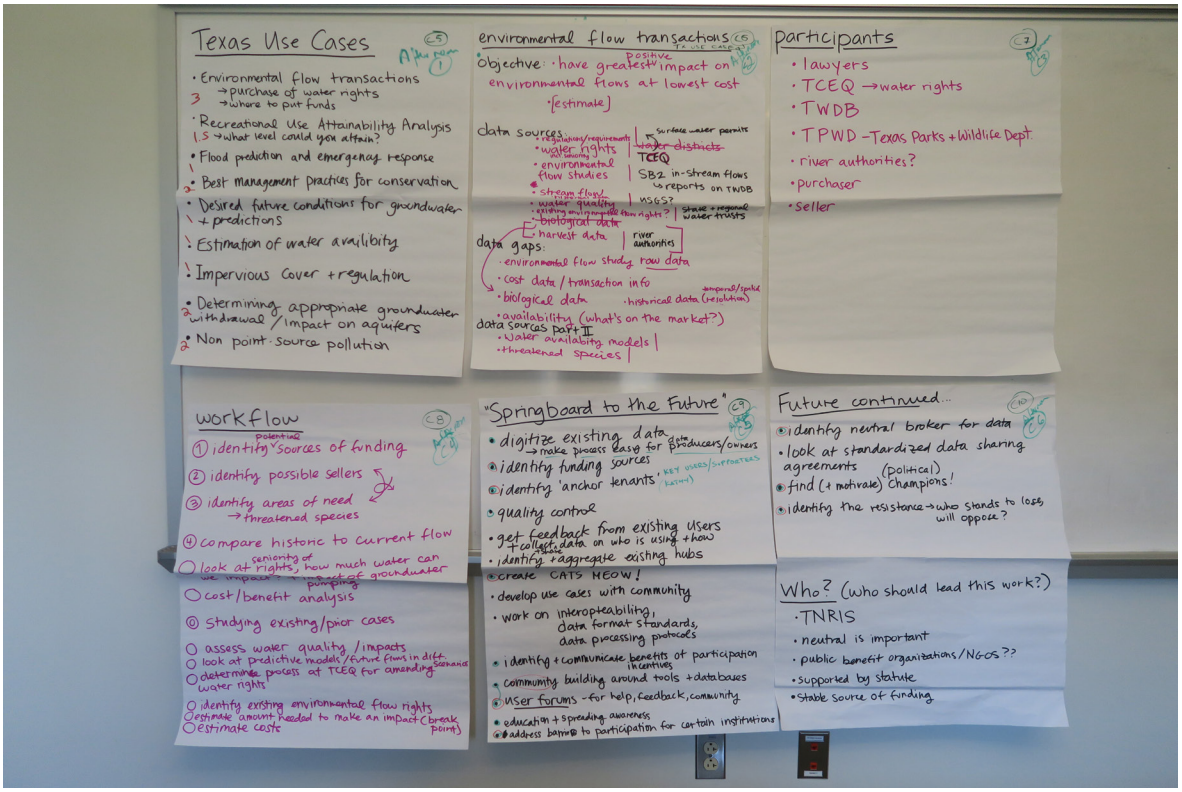

BRAINSTORM Use Case
group A #1
* Flooding (catastrophic)
impacts - economic social
- fed from FEMA/ACE EPA, state data - data not in one place
* What options do communities have for water supply & what options
for planning
State Flood Assessment
SAWS - piping water inter basin transfer vs. desal brackish gw vs. reservoir
- landowners' economics
- $ to city -
* Regional Water Plans ⇒ State Water Planning
group A #2 Brainstorm
TCEQ WAM: @ drought of record. Does not have -2011 drought records incorporated into WAM
$ & staff not available to update WAM
LCRA/Region K have tried
* Harris Galveston Subsidence District
Group A #3 Brainstorm
- Integrated support from other regulatory entities including USGS & state
- Sophisticated GW data
* Conservation Data
real time monitoring & use rates for municipalities
utilities
* ~3300 connections report to TWDB
GPCD rate goals
TWDB is req'ing to publicly display historic GPCD data
* Allocation of SW during times of drought
Group A #1 Brainstorm
Brazos example
* Industry water use during times of boring/drought etc instead of desal
* Theme - more data & transparency
* Supply viability & priority of contracts & seniority, diversion, beneficial use
* Sales transactions ⇒ reporting
- no delineation of water estate for parcels
GROUP A brainstorm # 5
Group A #1 (2)
* Watershed Protection Planning
comprehensive
issues → water rights junior/senior
→ water quality (NPS)
→ reallocation
eg Rio Grande & interboundary
* Allocation

A2

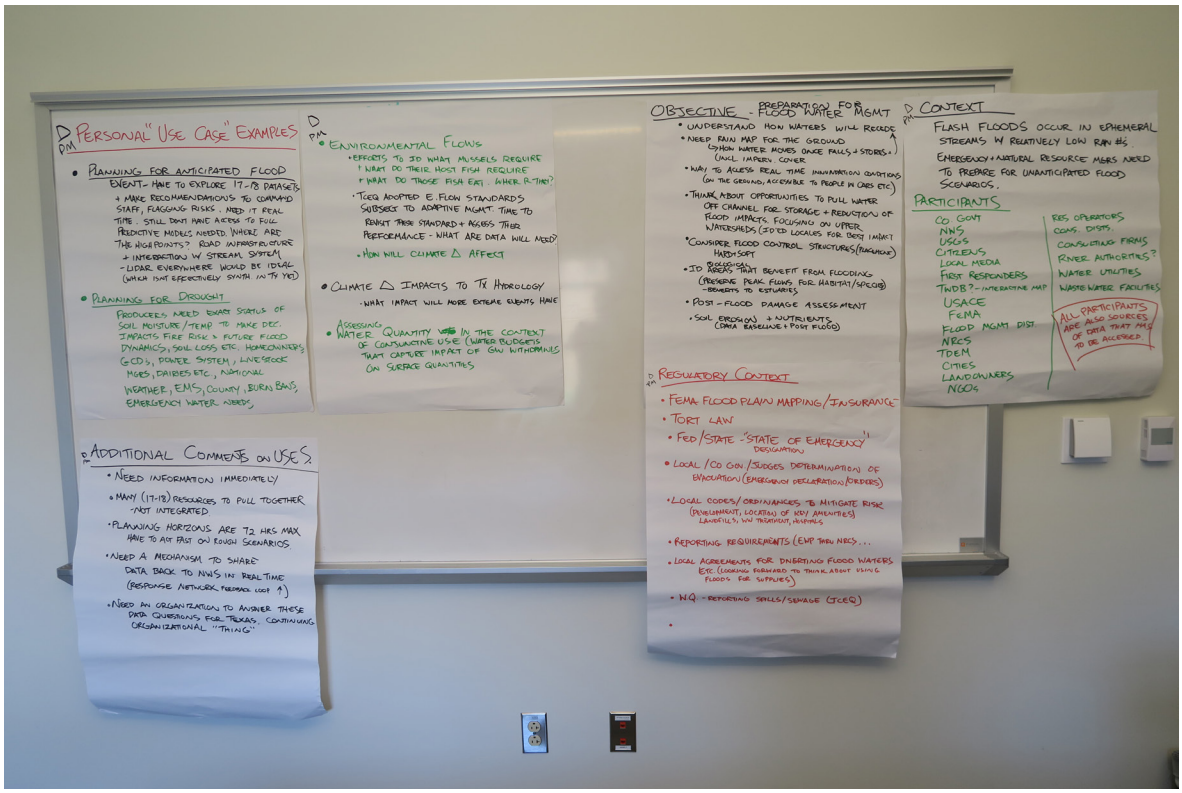

Use case - Watershed planning
group A
Obj - developing water budget for a river-shed
- science-based planning
who: usual suspects
context: Before investing $ $
data: land use over time & water use
Source: cities, TNRIS, NRCS, counties google earth, planning orgs
data: actual GW & SW use & interaction metered data WAS, ESPECIES SW diversions (daily/weekly) discharges, stream gauges geologic data, soil moisture
Use Case Reporting
Group A #2
* Flooding collection of datasets for ↓ risk & ↑ response to prioritize projects
Who: - FEMA/EPA/COE
- Clean Water Act
- Local public works
data: Flood mapping
Reservoirs
NPS
Allocation Use case
group A #3
- the more you pay, the more you care
- basic water needs met, everything above that has to pay ⇒ auctions
Who: GCDs, TWDB, TCEQ (regulatory) Businesses (impacted)
Springboard 4 the future
group A #1
* legislative support
- funding
- mandate (regulatory mandate)
* ID the network structure ⇒ who is going to do what?
- who are the partners
- who will pay
- who will lead
* Pick handful of use cases to do a Pilot (small demo)
* Share experiences and best practices for open data programs already in existence
- eg Flooding Action Group in Hays Co as a result of Blanco
Floods Data Sharing already happening
* Open data already happening
group A #2
- how
- why
- who
among all stakeholders
what are the barriers/liabilities to share data. Eg gw well data (private property) could be aggregated to ↓ personal identifier
* Standards
• Online/AI - messy data collation helps
• pdf/other conversion to digital data = guidance?

GROUP SESSIONS II

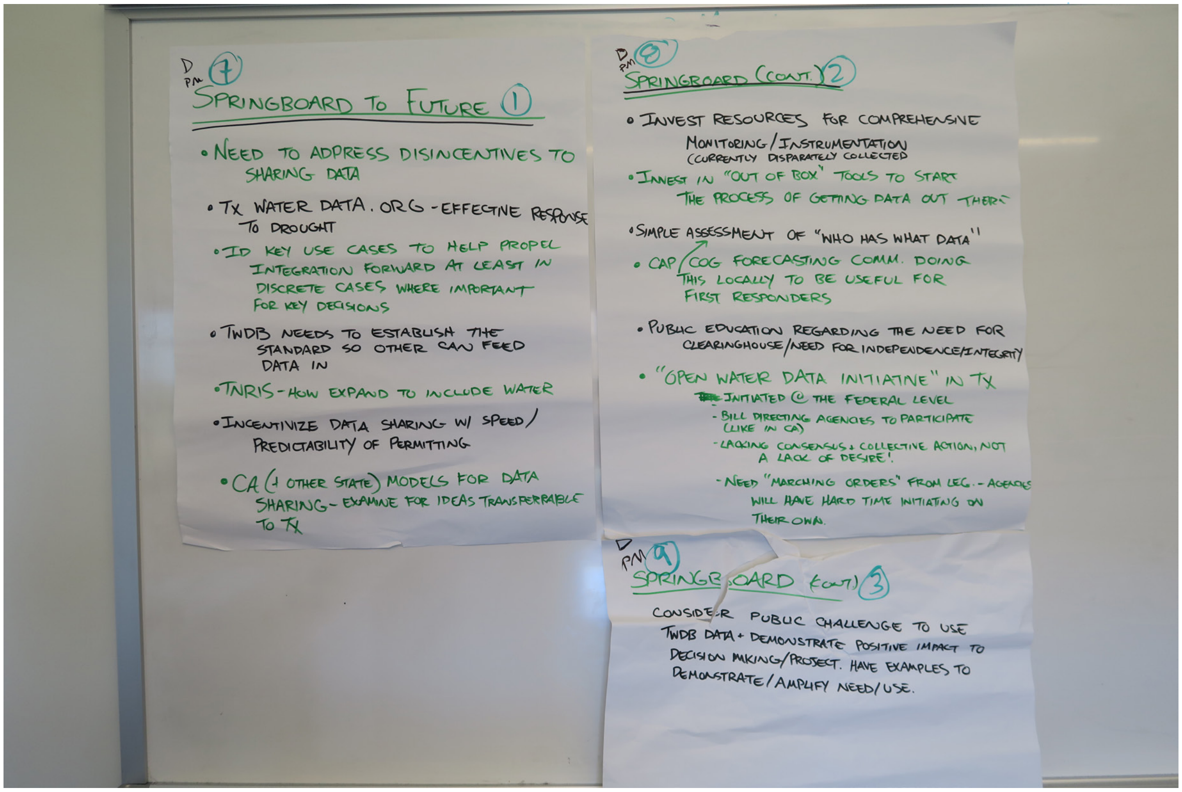
C



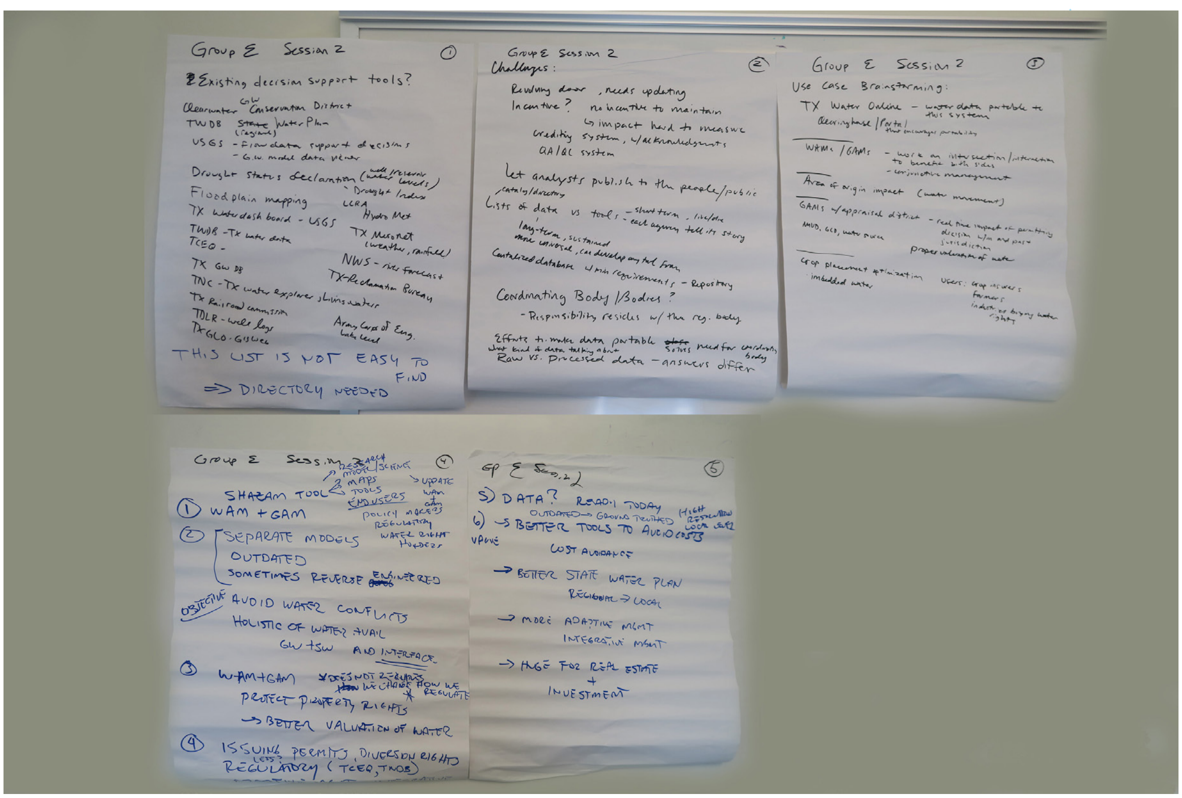
D1



D2

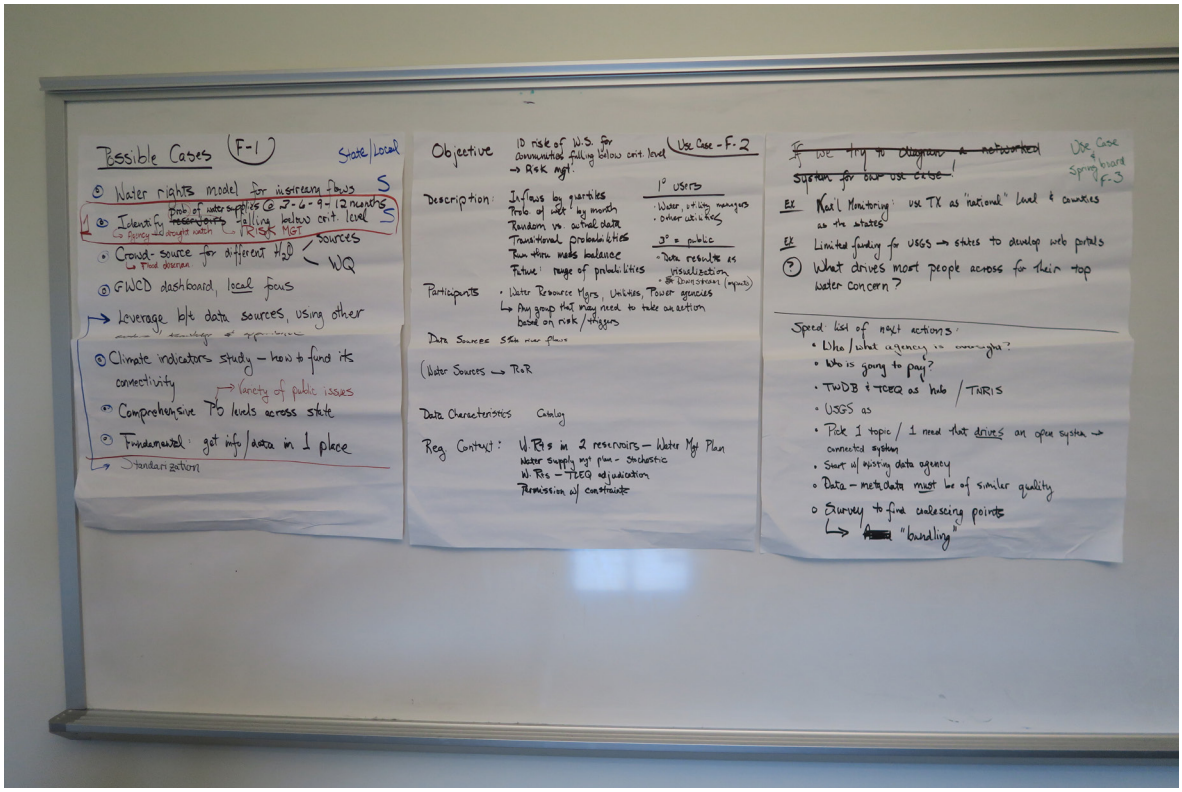

SPRINGBOARD TO FUTURE (1)
• NEED TO ADDRESS DISINCENTIVES TO SHARING DATA
• TX WATER DATA.ORG - EFFECTIVE RESPONSE TO DROUGHT
• ID KEY USE CASES TO HELP PROPEL INTEGRATION FORWARD AT LEAST IN DISCRETE CASES WHERE IMPORTANT FOR KEY DECISIONS
• TWDB NEEDS TO ESTABLISH THE STANDARD SO OTHER CAN FEED DATA IN
• TNRIS - HOW EXPAND TO INCLUDE WATER
• INCENTIVIZE DATA SHARING W/ SPEED/PREDICTABILITY OF PERMITTING
• CA (+ OTHER STATE) MODELS FOR DATA SHARING - EXAMINE FOR IDEAS TRANSFERRABLE TO TX
SPRINGBOARD (CONT.) (2)
• INVEST RESOURCES FOR COMPREHENSIVE MONITORING/INSTRUMENTATION (CURRENTLY DISPARATELY COLLECTED)
• INVEST IN "OUT OF BOX" TOOLS TO START THE PROCESS OF GETTING DATA OUT THERE
• SIMPLE ASSESSMENT OF "WHO HAS WHAT DATA"
• CAP/COG FORECASTING COMM. DOING THIS LOCALLY TO BE USEFUL FOR FIRST RESPONDERS
• PUBLIC EDUCATION REGARDING THE NEED FOR CLEARINGHOUSE/NEED FOR INDEPENDENCE/INTEGRITY
• "OPEN WATER DATA INITIATIVE" IN TX
INITIATED @ THE FEDERAL LEVEL
- BILL DIRECTING AGENCIES TO PARTICIPATE (LIKE IN CA)
- LACKING CONSENSUS + COLLECTIVE ACTION, NOT A LACK OF DESIRE!
- NEED "MARCHING ORDERS" FROM LEG. - AGENCIES WILL HAVE HARD TIME INITIATING ON THEIR OWN
SPRINGBOARD (CONT.) (3)
CONSIDER PUBLIC CHALLENGE TO USE TWDB DATA + DEMONSTRATE POSITIVE IMPACT TO DECISION MAKING/PROJECT. HAVE EXAMPLES TO DEMONSTRATE/AMPLIFY NEED/USE.

E


Group E Session 2
Existing decision support tools?
THIS LIST IS NOT EASY TO FIND
=> DIRECTORY NEEDED
Group E Session 2
Challenges:
Coordinating Body/Bodies?
Group E Session 2
Use Case Brainstorming:
Group E Session 2
SEPARATE MODELS
OUTDATED
AVOID WATER CONFLICTS
PROTECT PROPERTY RIGHTS
BETTER VALUATION OF WATER
→ BETTER STATE WATER PLAN
→ MORE ADAPTIVE MGMT
→ HUGE FOR REAL ESTATE + INVESTMENT

GROUP SESSIONS II

F



Connecting Texas Water Data Workshop

This report may be cited as: Rosen, Rudolph A. and Susan V. Roberts. 2018. Connecting Texas Water Data Workshop. Institute for Water Resources Science and Technology, Texas A&M University-San Antonio, San Antonio, TX 78224. (ISBN-13: 978-0-9986645-4-5) https://libguides.tamusa.edu/ld.php?content_id=42020932

Copies may be obtained at https://libguides.tamusa.edu/ld.php?content_id=42020932

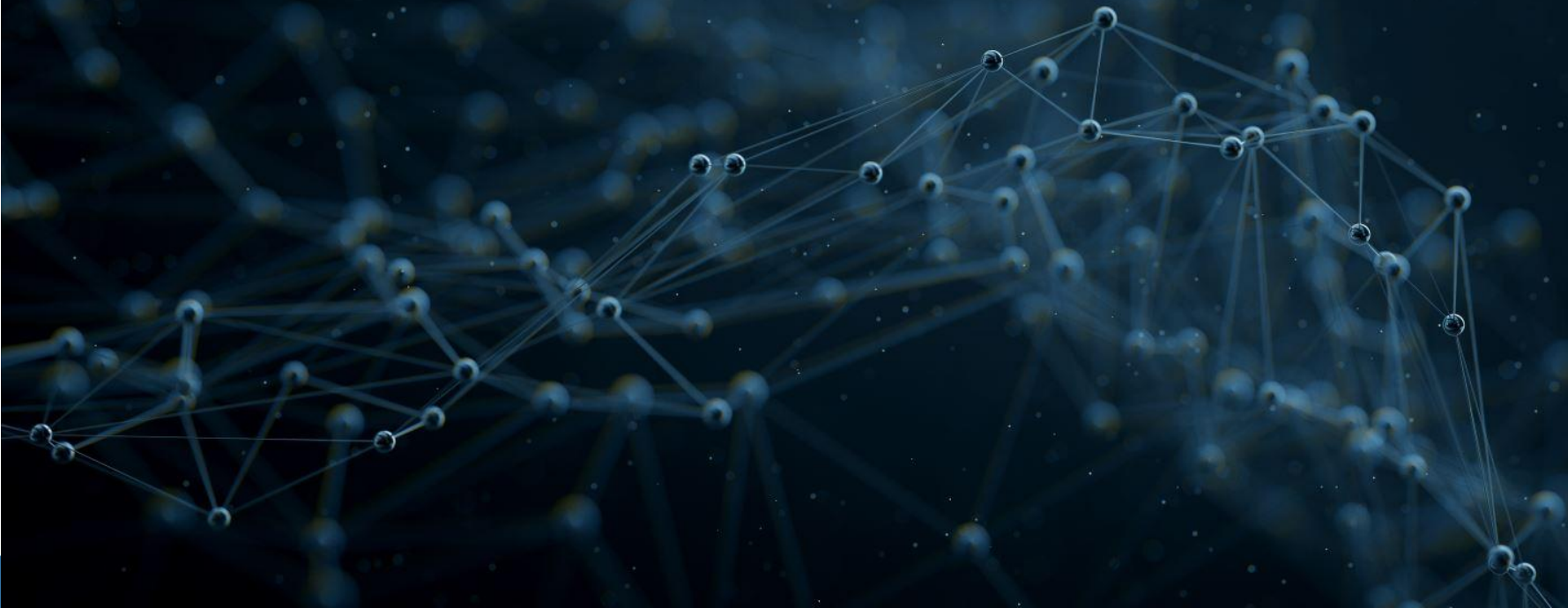
Appendix B

% Cell Count of Agency	
Agency	**% of data**
TWDB	74.61%
USGS	3.37%
TCEQ	3.23%
TexMesonet	1.68%
HIFLD	1.26%
SDR	1.26%
NWS	1.26%
NOAA	1.12%
LCRA	0.98%
TxDOT	0.98%
BRA	0.98%
TPWD	0.70%
USDA	0.42%
NDMC	0.42%
TNRIS	0.42%
CoCoRaHS	0.42%
USACE	0.28%
GLO	0.28%
Austin Water	0.28%
GBRA	0.28%
NRA (Nueces River Authority)	0.28%
BCRAGD	0.28%
USFWS	0.28%
ANRA	0.28%
TSSWCB	0.28%
Texas A&M University	0.28%
SARA	0.28%
NWQMC	0.14%
USFS	0.14%
University of Washington	0.14%
SJRA	0.14%
EPA	0.14%
TAMU	0.14%
HDSC	0.14%
LNRA	0.14%
Office of the Attorney General	0.14%
TDEM	0.14%
FEMA	0.14%
HIFLD	0.14%
NRCS	0.14%
Texas Association of Regional Councils	0.14%
Tx State Soil & Water Conservation Board	0.14%
Texas Demographic Center	0.14%
TxDOT	0.14%

NLCD	0.14%
US Census Bureau	0.14%
The Texas General Land Office	0.14%
Public Utility Commission	0.14%
Hays County	0.14%
RRC	0.14%
TomTom	0.14%
USIBWC	0.14%
Hays Trinity Groundwater Conservation District	0.14%
TRA	0.14%
Grand Total	**100.00%**

Appendix C

Texas Water Data Hub



The TWDB Team

Richard Wade

Advisor

Sam Hermitte

Advisor

Laura Sepulveda

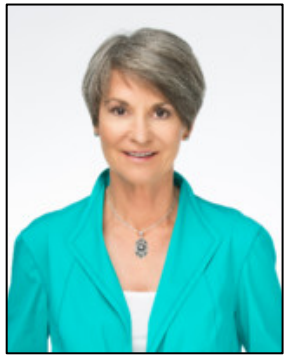
Design

Taylor Christian

Data

John Haney

Development



Kathleen Mack

Contracts

The Team

Texas Water Development Board

Internet of Water

THE CYNTHIA & GEORGE MITCHELL FOUNDATION

Richard Wade

Sam Hermitte

Taylor Christian

John Haney

Laura Sepulveda

Mission

Create an intuitive system to index, document, search and access Texas water data.

Process

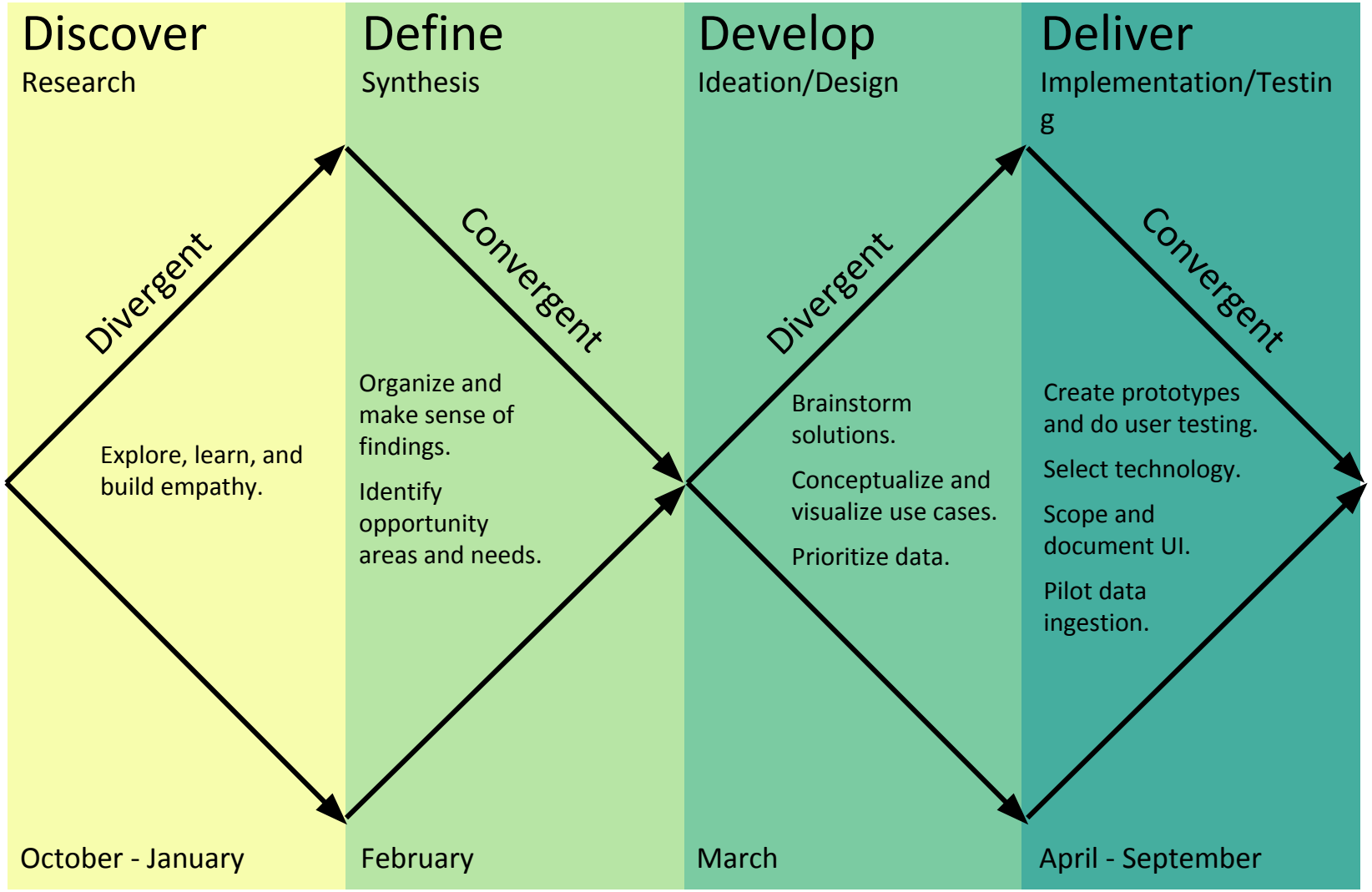
October 2020 – September 2021

Focus on four initial phases:

- Discover
- Define
- Develop
- Deliver

The design process will oscillate between divergent and convergent thinking methods to generate many ideas and possibilities and then refine and narrow them down.

We will use human centered design methods to engage real people throughout the process to make sure we are designing an intuitive system that will work for our users.



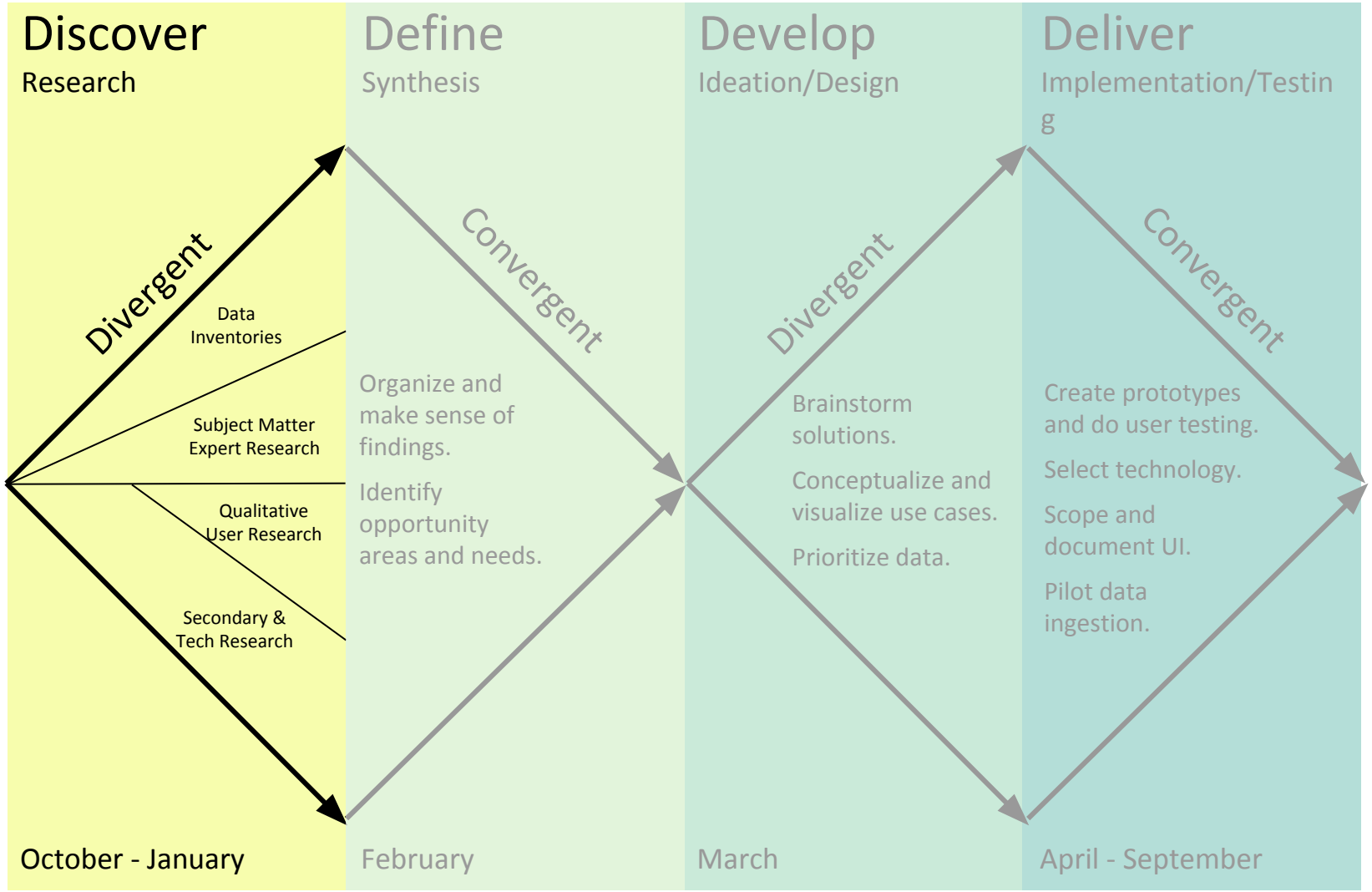
What We've Done

Created an inventory and report of all TWDB water data and began updating an existing statewide water data inventory.

Conducted 10 Subject Matter Expert interviews to learn from others' open data experiences.

Held 11 in-depth qualitative user research interviews to build empathy with (water) data users and better understand their workflows and behaviors for accessing, working with, and sharing data.

Explored existing data hub features and investigated technology options.



Where We Are

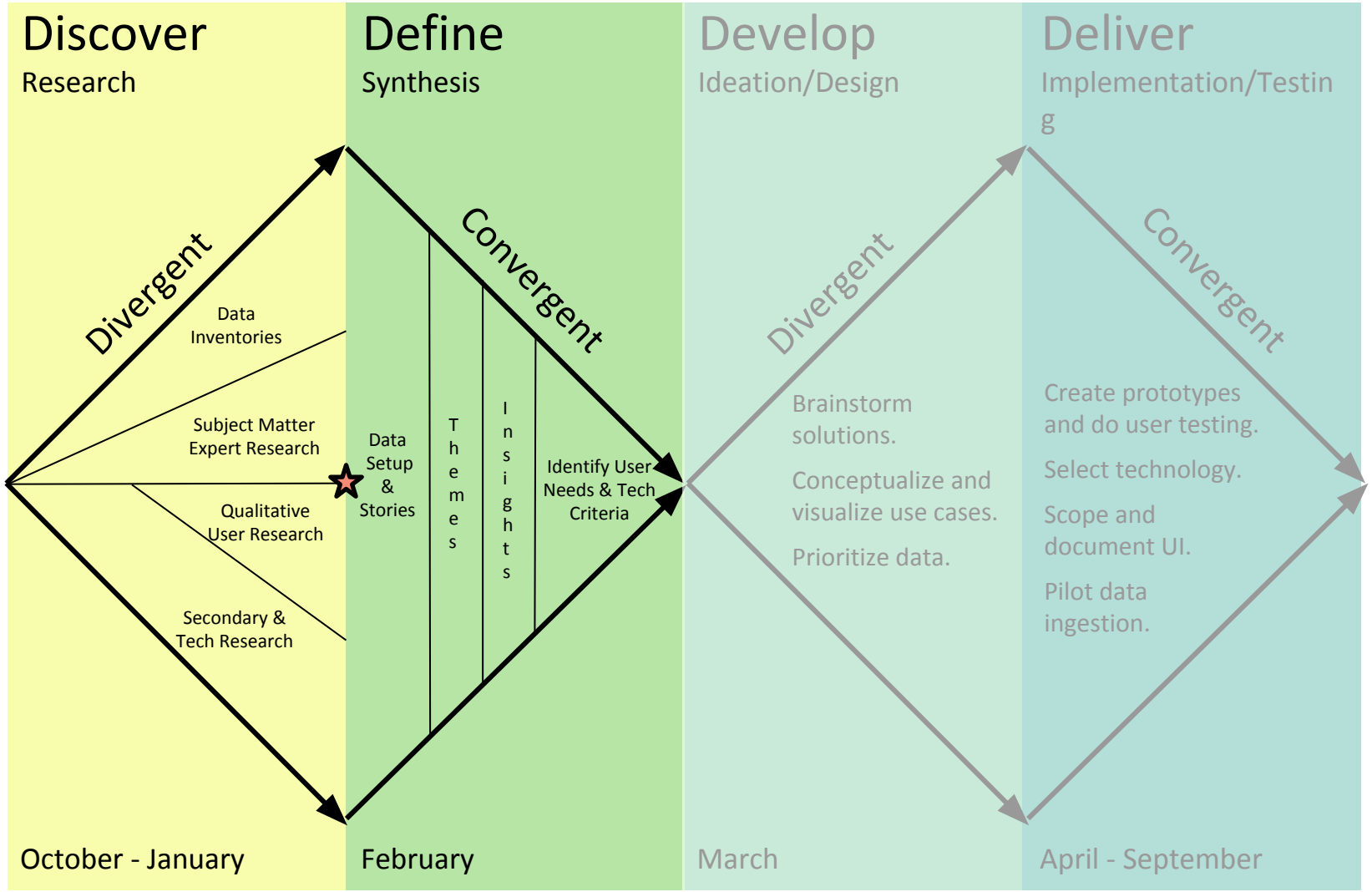
We have accumulated an overwhelming mountain of information and are organizing and making sense of it all.

We are highlighting interesting stories from the field and learnings from subject matter experts.

Cleaning up our data for affinity mapping methods to identify common underlying behaviors and themes.

Our most relevant themes will then be turned into provocative and actionable insights.

From there we will be able to develop requirements around technology and user needs.



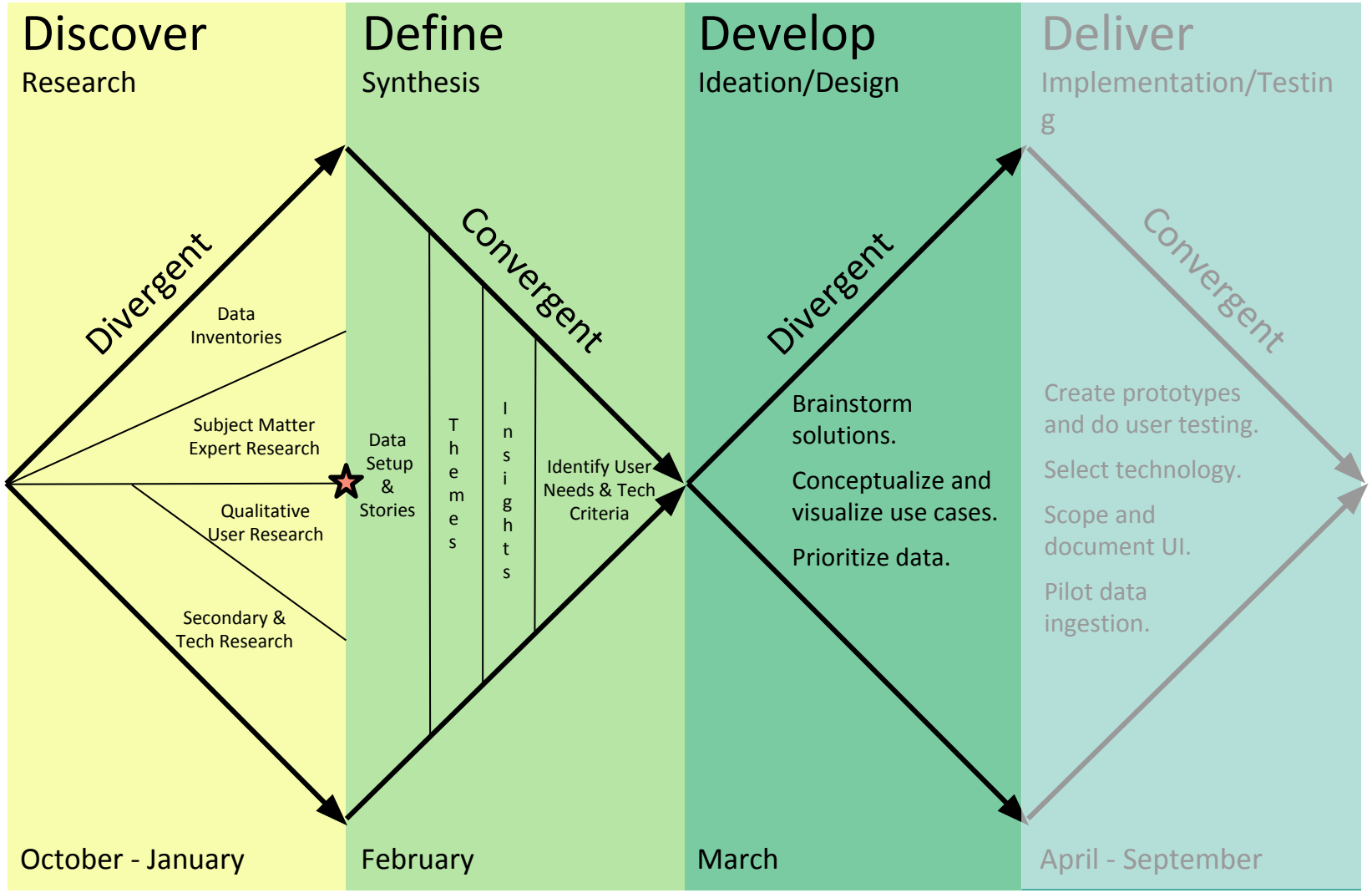
What's Next

After clearly defining the problem space, our users needs, and design criteria to help shape solutions we will switch back to divergent thinking methods.

We will hold brainstorm ideation sessions to create a large number of ideas for solutions- both bad and good.

Top solution ideas will be fleshed out and incorporated into user stories for conceptualization.

We will work to prioritize and further document our inventory data for future inclusion in the hub.



Qualitative Research Goals

We are seeking to understand:

- How people search for and find data and information they need
- How people use (water) data
- What is important to people when evaluating a data source
- How producers update and share data

Research Methods

Contextual Inquiry

- Virtual, in-context, two hour meetings with individual participants
- Structured exercises + interview style questions
- Recorded

Overview

- Consent and explanation
- Background and data questions
- Activity 1 – Workflow shadowing
- Activity 2 – Search and select
- Activity 3 – Card sort of data hub features
- High level closing question

Subject Matter Expert Research Goals

We are seeking to understand:

- The process others used to create data hubs timelines, resources and how they went.
- Drivers behind the creation of hubs (legislation, user demands, etc.)
- Mistakes to avoid and best practices to guide us.
- The workflows and technology others are using.
- How they are handling legacy data, data standards and standardization.
- What next steps are they considering or implementing?

Texas Data Hub
SME Research Plan

Overview

In order to explore and build an innovative and effective statewide hub for water data in Texas, the team will meet with outside subject matter experts on data sharing, governance and hubs to learn from their experience. In doing so we hope to build off work that has already been done by others instead of re-inventing the wheel. These interviews will focus on learning from others' implementation experiences to inform strategic planning, guiding principles and best practices to compliment the user research intended to learn about human behavior and user needs.

Goals

We are seeking to understand:

- The process others used to create data hubs, timelines, resources and how they went.
- Drivers behind the creation of hubs (legislation, user demands, etc.)
- Mistakes to avoid and best practices to guide us.
- The workflows and technology others are using.
- How they are handling legacy data, data standards and standardization.
- What next steps are they considering or implementing?

Participants

1. The Internet of Water
2. United States Geologic Survey (USGS)
3. New Mexico
4. California
5. Colorado
6. Lower Colorado River Authority (LCRA)
7. Texas Advanced Computing Center (TACC)
8. The First Street Foundation - Flood Factor
9. City of Austin

Method

These will be virtual discussion type meetings lasting about 1 hour to get a better understanding of the experience and knowledge others have in building similar data hubs and any advice they may have to offer us. The below is a list of general questions to help guide the conversation.

- What was the process like planning, building and launching your data hub?
    - Timelines
    - Resources (staff and funding)
    - Description of the process (decision making and design/programing)
- How did you decide what data to include in the hub and what technology to use?
- What approach to data standards and data standardization have you taken?

- How do you ingest data into your hub? Do you use data transformations?
- How do you handle legacy data?
- What is working well?
- What challenges did you face? How did you resolve them?
- What do you plan on or wish you could change?
- What advice would you give us?
- What's next for your hub?
- How do you measure success?
- Anything else you think we should know?

Subject Matter Expert Research

In order to explore and build an innovative and effective statewide hub for water data in Texas, the team met with outside subject matter experts on data sharing, governance and hubs to learn from their experience. In doing so we hope to build off work that has already been done by others instead of re-inventing the wheel. These interviews focused on learning from the implementation experiences of others to inform strategic planning, guiding principles and best practices.

Goals

We were seeking to understand:

- The process others used to create data hubs including timelines, resources and how they went.
- Drivers behind the creation of hubs (legislation, user demands, etc.)
- Mistakes to avoid and best practices to guide us.
- The workflows and technologies others are using.
- Approaches to legacy data, data standards and standardization.
- What next steps others are considering or implementing?

Who We Spoke with

- California
- New Mexico
- Colorado
- United States Geologic Survey (USGS)
- Internet of Water
- City of Austin
- Lower Colorado River Authority (LCRA)
- Texas Advanced Computing Center (TACC)
- The First Street Foundation - Flood Factor

Key Learnings and Takeaways

Add value through problem solving

Set clear standards and governance

Build with the future in mind

Take a phased approach

Empower users

Make it a community effort

Add value through problem solving

Meeting the operational and business needs of water data organizations is a foundational component. Providing benefits to water data owners, will increase participation and aid in quicker and more comprehensive data build out.

Linking business use cases to drive development overtime will show direct value and benefit everyone. This will build support and understanding in the water data community and lead to partnerships that naturally grow and sustain the hub in the long term.

Combining water data with data from other realms such as social or demographic data creates high value problem solving and decision-making tools beyond just a data catalog.

Water data is often siloed or spread across various projects, teams, tools, or organizations. Many lack the resources and ability to consolidate and standardize data to create large comprehensive datasets. This type of work is both difficult, and highly valuable.

Set clear standards and governance

Setting clear data standards and governance early in the process will help organize resources, make decision, and development. A balance may need to be struck between ease of entry and strict adherence to ideal standards in the beginning to promote partner contributions and data publishing. Outlining ideal and minimum criteria will allow for some flexibility while still outlining the goal.

Individuals, teams, and organizations have various levels of data literacy and management which impact data and the level of effort/ability to contribute to an open data system. Data is also often influenced by changes in ownership, technology, funding, and relevance over time. These factors must be taken into account when developing resources to support data producers, users, and the public.

Being data creator and data server makes standardization easier. Changes and transformations should ideally be made by the data owner. Otherwise, having your own database with a set schema that can be controlled may also help with standardization so ingested data is made to fit the end point database.

Clear governance around legacy and archived data as well as source and derived data is important to inform users and make sure you are presenting the most relevant and accurate information available. A maintenance and storage plan for data archiving will help ensure optimal performance and historical retention.

Build with the future in mind

Solidifying reliable ongoing funding and a dedicated team is a key component of a successful open data system. The product will require ongoing attention and support as technology, expectations, and needs inevitably evolve over time. This cannot be treated as a side project or an add-on to existing responsibilities and must be able to survive leadership changes.

A focus on meeting existing business needs and slow incremental development overtime is sustainable and effective but needs to include a long-term planning element to ensure cohesive, interoperable, and collaborative results. Without future planning this type of development could result in piecemeal or stand alone features that solve specific problems, but do not work well together and require extra training resources and maintenance.

Flexibility, service integration, and maintenance are important to consider when selecting technology options. Weighing long-term and short-term needs will help avoid having to do a costly refresh down the road. As the system grows in functionality, content, and use the ability to automate repetitive update and maintenance tasks will be valuable and should be considered from the beginning. Initial development may need to focus on specific areas, but technology that can easily accommodate future integrations will be better setup for specialization to meet more advanced user needs and evolve with the times.

The more research and requirement gathering that can be done up-front considering users will allow for more strategic planning and the ability to show impact and value sooner. A system setup with different environments for testing, internal users, and the public will be more adaptable in the long-term allowing for iterative updates and soft releases.

Take a phased approach

A phased approach is important to focus development, but also create set deliverables and deadlines to keep the project on track. For example, research and use case development can become ongoing never-ending tasks that spiral out of control. A phased approach can outline appropriate times to address process step while maintaining progress forward.

Phased development provides the ability to head off problems or performance issues that could come from trying to do everything at once. By building and launching in phases load and timeout issues may better be planned and accounted for resulting in increased trust and satisfaction from users.

Too many "tools" or specialized ways for users to access and interact with data can end up creating data silos, clutter, and confusion in a system. A phased approach may help to better focus on a few broad use cases or areas that will strategically work together to create a cohesive experience.

Empower users

Empowering users to feel in control and manage their data in a way that fits their own workflows will increase comfort, trust, and usability. If data contributors can see immediate and tangible value this will lead to a sustainable cycle of growth and loyalty within the open data system.

When user success is directly linked with the success of the open data system everyone will benefit. Learning and growth should be a core success metric as opposed to the classic model of focusing inward to improve efficiency.

Documentation and tracking is imperative for data practitioners but is not very glamorous and is often overlooked or generalized if not captured in the moment. An open data system with built in tracking and documentation would be able to provide immediate value to the water data community. Especially if this was done in a way that provided increased control to users such as with URIs for accessing uploaded data through specific URLs, data management tools that would allow for immediate adjustments and corrections outside of scheduled update cycles, or site level landing pages to be able to find and access at a more relevant granular level.

Users often working closely with and reach out to others in the water data community for advice, support, and recommendations. The personalized support that comes from interactions with others in the community should be incorporated into the digital environment as well whenever possible. An open data system should allow for two-way communication and input as well as different means of personalized interactions between users to integrate the institutional knowledge and expert skills that exist within the water data community.

Make it a community effort

Development of an open data culture to promote sharing and interaction is key to building a following and success. Outreach and education for staff and the broader community will build trust and understanding. Establishing liaisons for partner organizations and data types to act as stewards can help champion the cause and increase accountability while reducing the data maintenance load.

Clarity in roles and responsibilities is critical to creating an environment of accountability and inclusion. Make sure to look at a diverse user group when evaluating needs, and continually check back in to make sure these are accurate and being addressed appropriately.

Foster a collaborative and supportive relationship with partner organizations by focusing on how to alleviate burdens and reduce compliance pressures, and get people energized with use cases. The more understood users feel the more likely they are to interact with and get value from the open data system.

Obligatory participation can spur action, but also changes the power dynamic and may actually lead to slowing progess if the burden is too heavy.

Appendix D

Please be advised this is a working document. Everyone is welcome to contribute ideas on what should be covered during our meeting and how it should be organized. Please do not overwrite or delete someone else's contributions without coordinating it first. This is going to be a substantial discussion and the more organized we can make it ahead of time, the more efficient and productive it will be.

Ontology Discussion

Agenda Items
2 Hour Session 1 - May 4th 12:00pm-2:00pm central (10:00pm-12:00pm Pacific)

Welcome (Suzanne Pierce)
- Selecting the next date/time Ontology Part 2 Doodle Poll

Getting on the same page
- What do you hope to achieve?

Ontology 101 (Yolanda Gil)
Closing Remarks

Expected Attendees (Confirmed):

TWDB - Taylor Christian, Kelly Swanson, possibly another intern
CSR - Greg Smithhart, Brent Porter–Center for Space Research
CHARM - Steven Mikulencak - A&M Agrilife extension
TIFF - Amin Kiaghadi (TWDB), Samuel Rendon (USGS)
USC - Yolanda Gil, Varun Ratnakar
TDIS - Suzanne Pierce

Tentative Attendees:
Deborah Khider

Agenda Items
(2 Hour Session 2 - Tentative May 20 1pm-3pm central (11am-1pm Pacific)

Ontology Part 2 - Practical Applications

Welcome (Suzanne Pierce)
Successful Case Study of Broad Community Ontology (Deborah Khider)
- Best practices and lessons learned

Closing Remarks

Agenda Items
(3 Hour Session 3 - June ?? Morning Session)

Welcome
Hands-On Ontology Exercise (Varun Ratnakar)
*possibly - Combining Multiple Ontologies (Craig Knoblock) - contact through Karen Rawlins krawlins@isi.edu
Closing Remarks

Appendix E

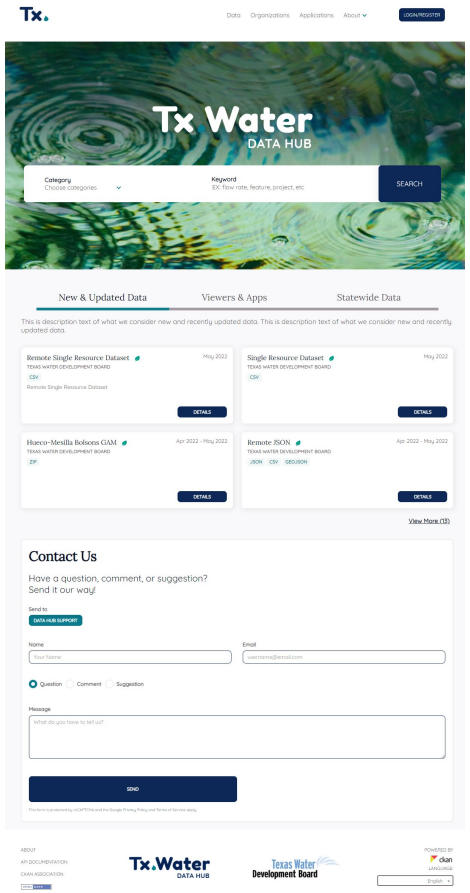
Texas Water Data Hub

Progress Update

Development is underway!

Completed:

- Landing page
- Catalog pages
  - Data
  - Organizations
  - Applications
- Organization details
- Application details
- and more…



Development is underway!

Completed:

- Landing page
- Catalog pages
  - Data
  - Organizations
  - Applications
- Organization details
- Application details
- and more…

In progress and upcoming:

- Data details page
- Download functionality
  - Data
  - Metadata
  - Data dictionary
- Upload process enhancements
- and more…

Launch Plan □

Alpha

Target:

July

Content:

5-15 datasets and applications

Audience:

Limited to data partners for testing filter/search facets, metadata, data details and download functionality.

Get Involved:

Email Taylor if you have publicly downloadable data and are interested in partnering.

Beta

Target:

November

Content:

25-50+ datasets and applications

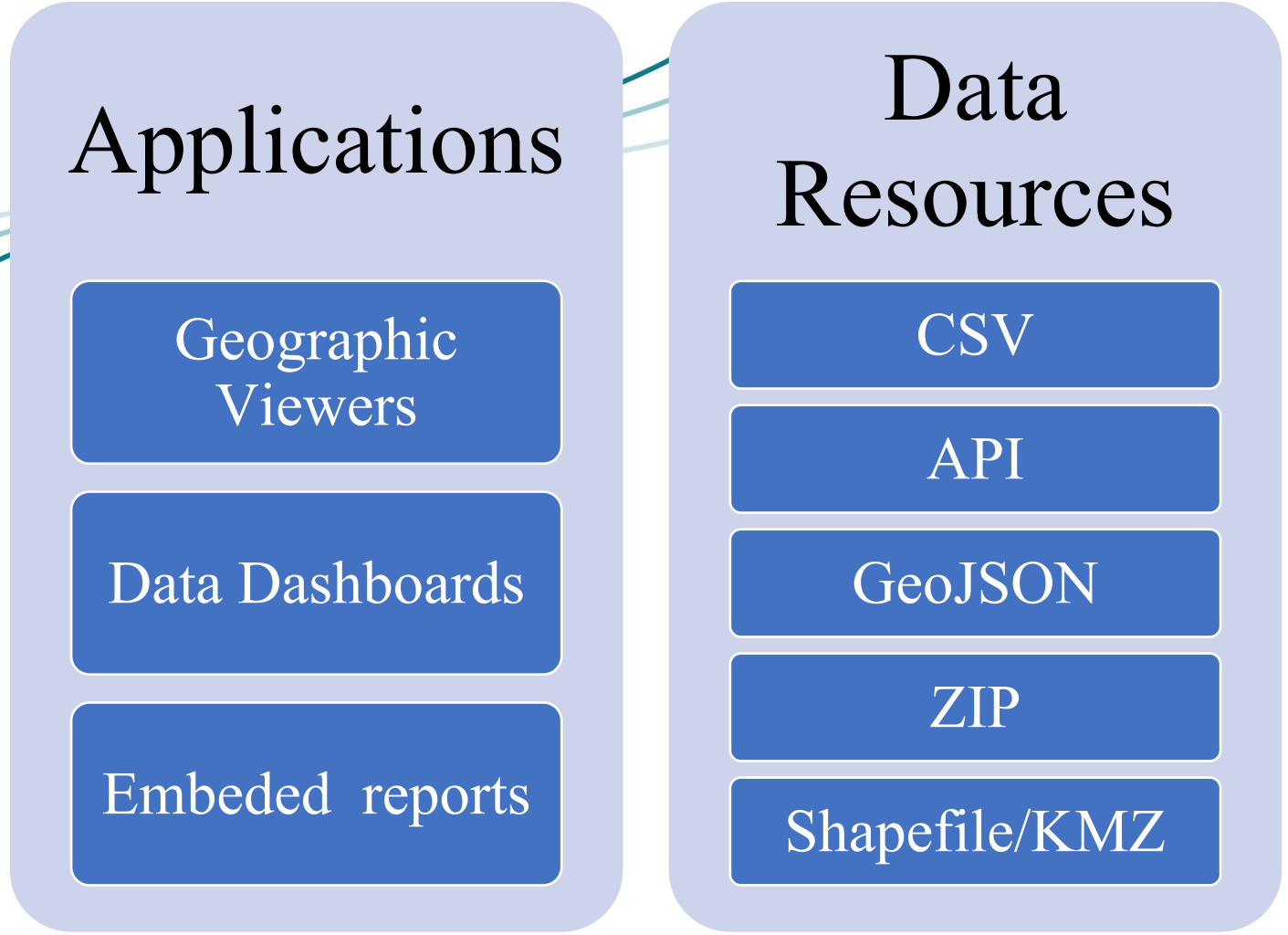
Audience:

Public

Get Involved:

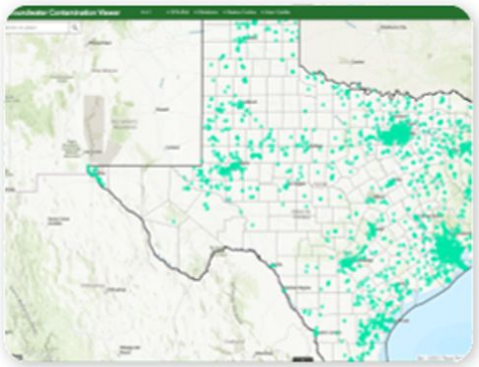
Email Taylor if you have publicly downloadable data and are interested in partnering.

What is a Data Resource vs. Application in the Texas Water Data Hub?



Do you have data you think would be valuable in the hub?

What category would you look in to find the Groundwater Contamination Viewer?



Groundwater Contamination Viewer

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

This interactive, online map allows users to query and obtain spatial relationship information about groundwater contamination cases.

0 Resource(s)

OPEN APP

DETAILS

What category would you look in to find Artificial Reef data and application?

Artificial Reefs Interactive Mapping Application

TEXAS PARKS AND WILDLIFE

Interactive mapping application enables the public to locate and query offshore artificial reefs along the Texas coast.

1 Resource(s)

OPEN APP DETAILS

Artificial Reef Sites

Apr 1968 - Jul 2019

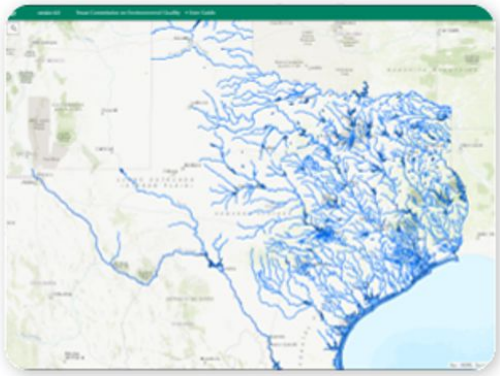
TEXAS PARKS AND WILDLIFE

Offshore artificial reefs along the Texas coast including site names, coordinates, and additional information.

CSV

DETAILS

What category would you look in to find the Surface Water Quality Segments Viewer?



Surface Water Quality Segments Viewer

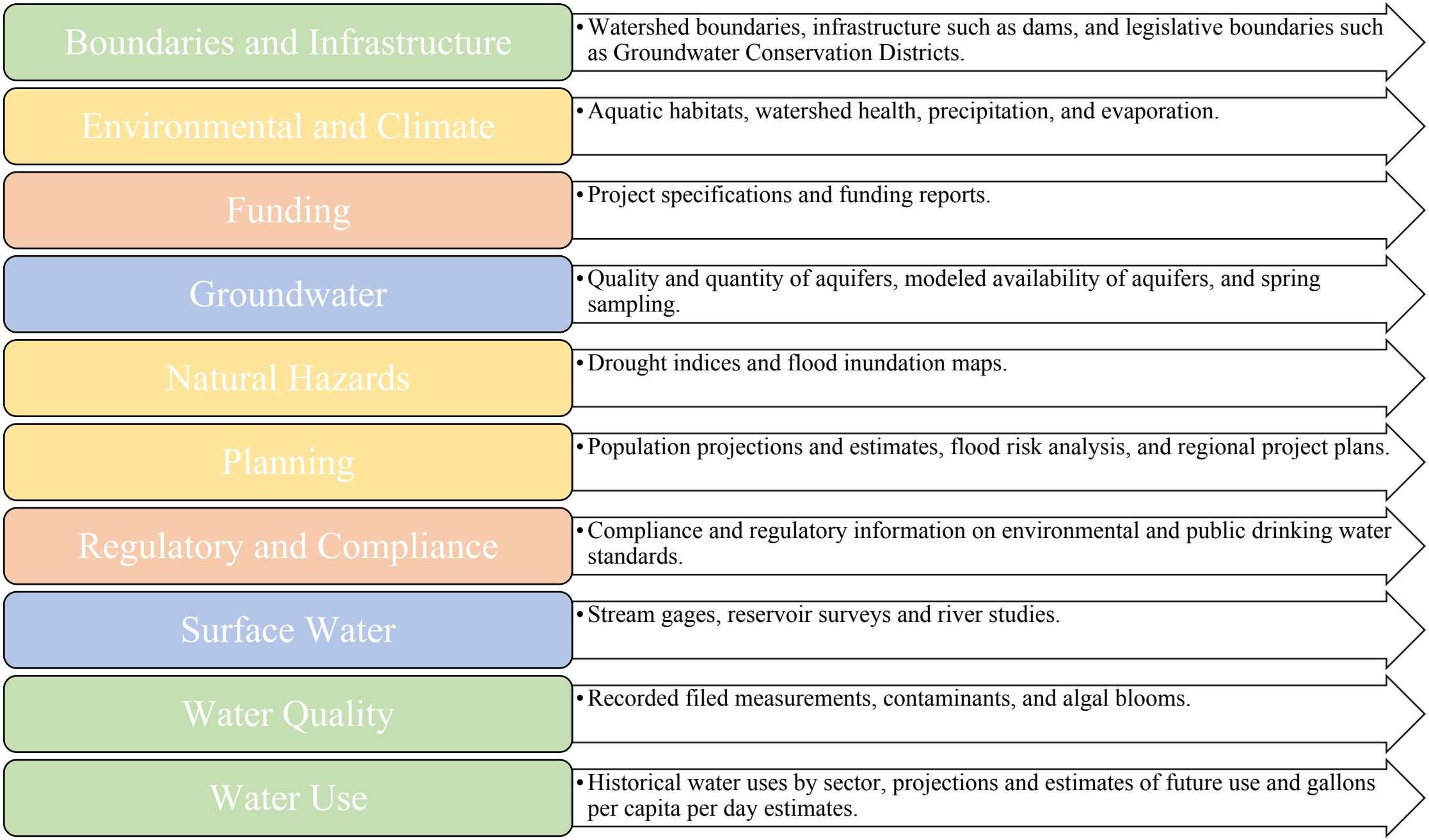
TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

This interactive map viewer allows users to view the stream segments and water bodies (including impairment status) for any body of water in Texas that the TCEQ monitors...

0 Resource(s)

OPEN APP

DETAILS


Boundaries and Infrastructure
• Watershed boundaries, infrastructure such as dams, and legislative boundaries such as Groundwater Conservation Districts.
Environmental and Climate
• Aquatic habitats, watershed health, precipitation, and evaporation.
Funding
• Project specifications and funding reports.
Groundwater
• Quality and quantity of aquifers, modeled availability of aquifers, and spring sampling.
Natural Hazards
• Drought indices and flood inundation maps.
Planning
• Population projections and estimates, flood risk analysis, and regional project plans.
Regulatory and Compliance
• Compliance and regulatory information on environmental and public drinking water standards.
Surface Water
• Stream gages, reservoir surveys and river studies.
Water Quality
• Recorded filed measurements, contaminants, and algal blooms.
Water Use
• Historical water uses by sector, projections and estimates of future use and gallons per capita per day estimates.

Groundwater

Surface Water

Boundaries and Infrastructure

Water Quality

Water Use

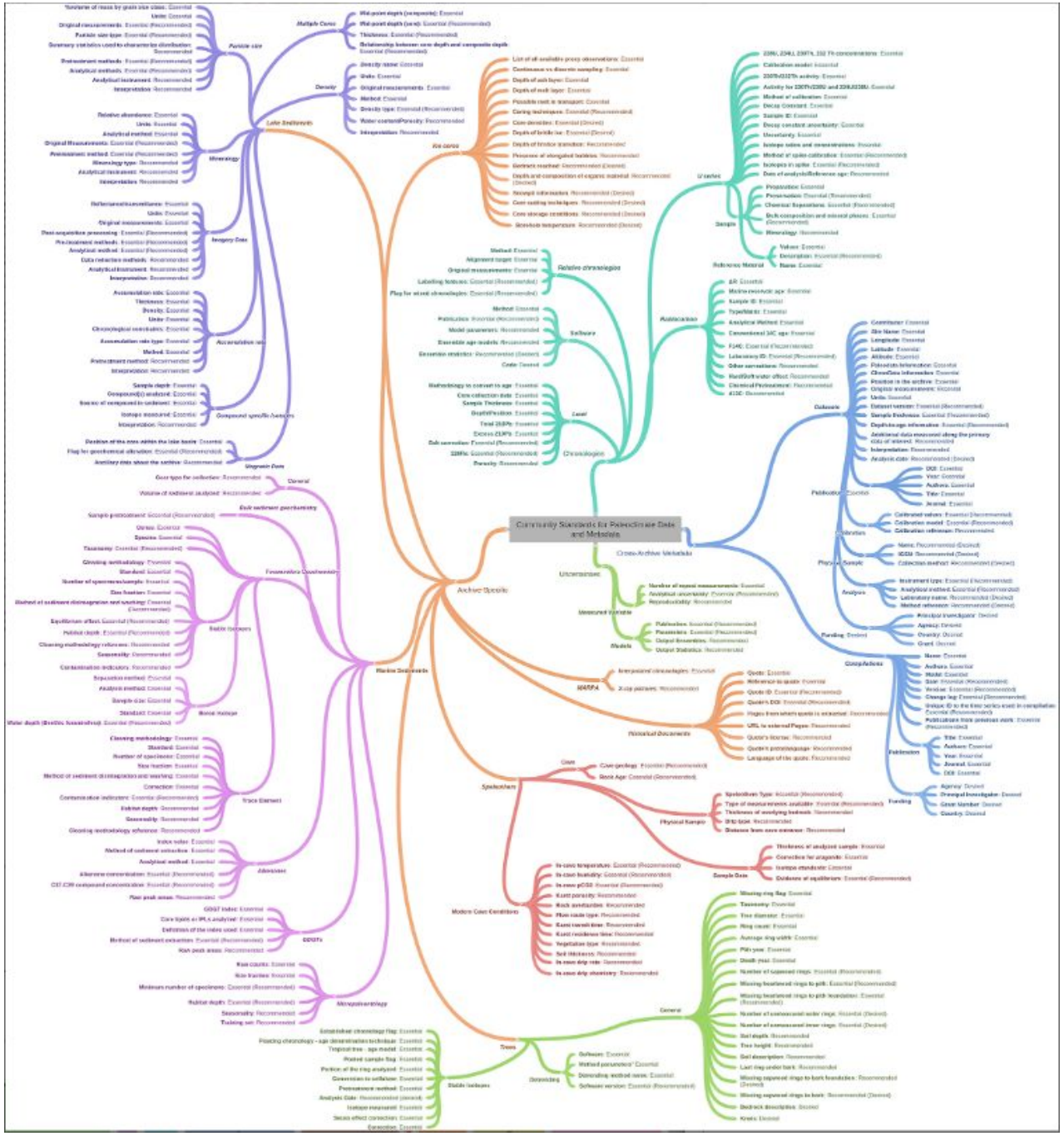
Environmental and Climate

Natural Hazards

Planning

Funding

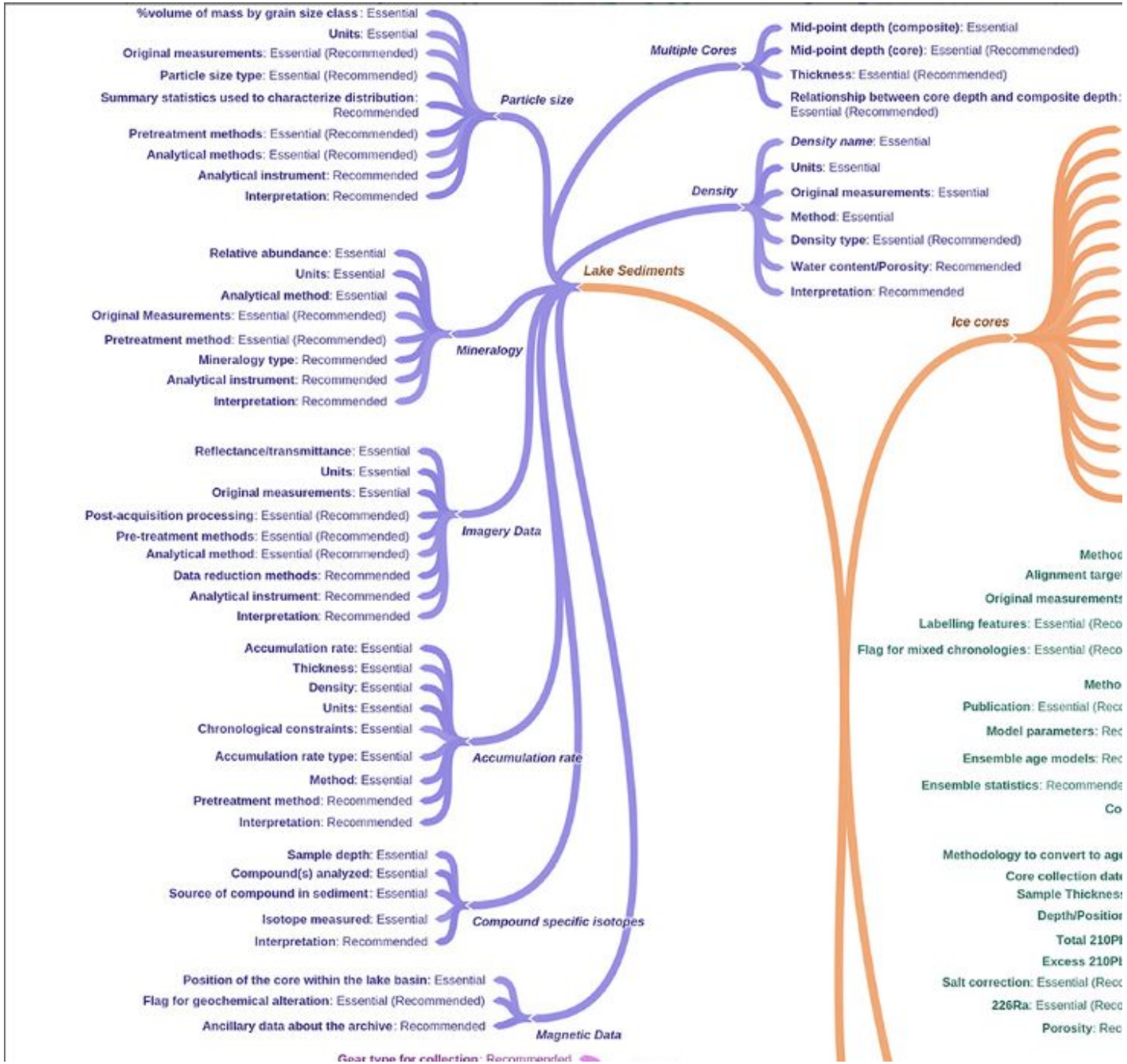
Regulatory and Compliance



PaCTS 1.0: A Crowdsources Reporting Standard for Paleoclimate Data

An example of Ontology: PaCTS: 1.0

"The Paleoclimate Community reporTing Standard (PaCTS) provides guidelines as to which information should be included when reporting data from various paleoclimate archives.....The ultimate goal of this effort is to (1) make these datasets more re-usable over the long term, and (2) provide a roadmap for implementing and revising the standard, as the field of paleoclimatology and its practitioners both evolve. **The requirements are driven by the differing needs of data producers and the data consumers, who often have different goals in mind. Thus, agreeing on and writing up these requirements involves building consensus among the community to decide on their present and future goals.** "

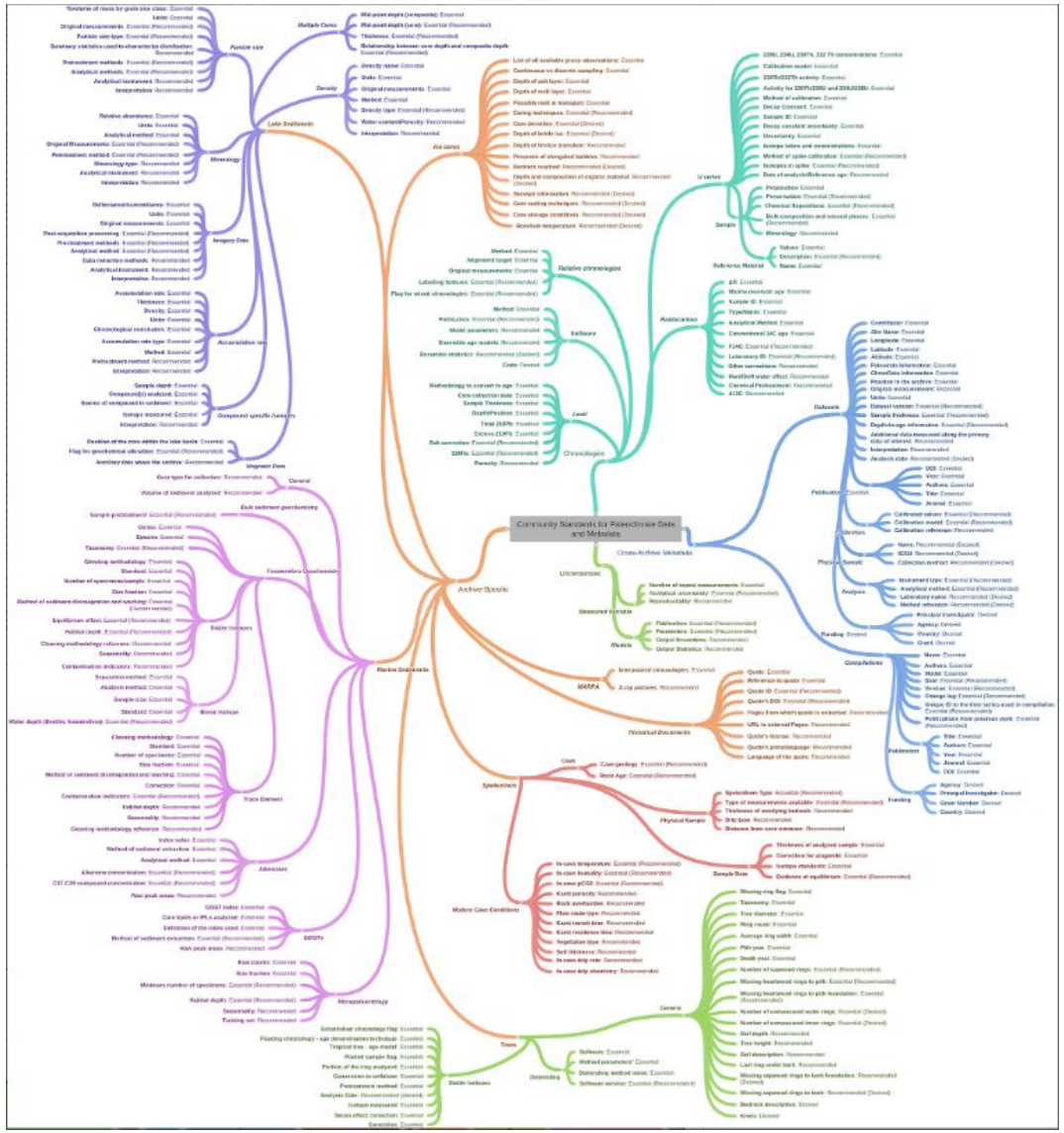


Within Lake Sediments there is data on:

- Magnetic data
- Compounded specific isotopes
- Accumulation rate
- Imagery data
- Minerology
- Particle size
- Multiple cores
- Density

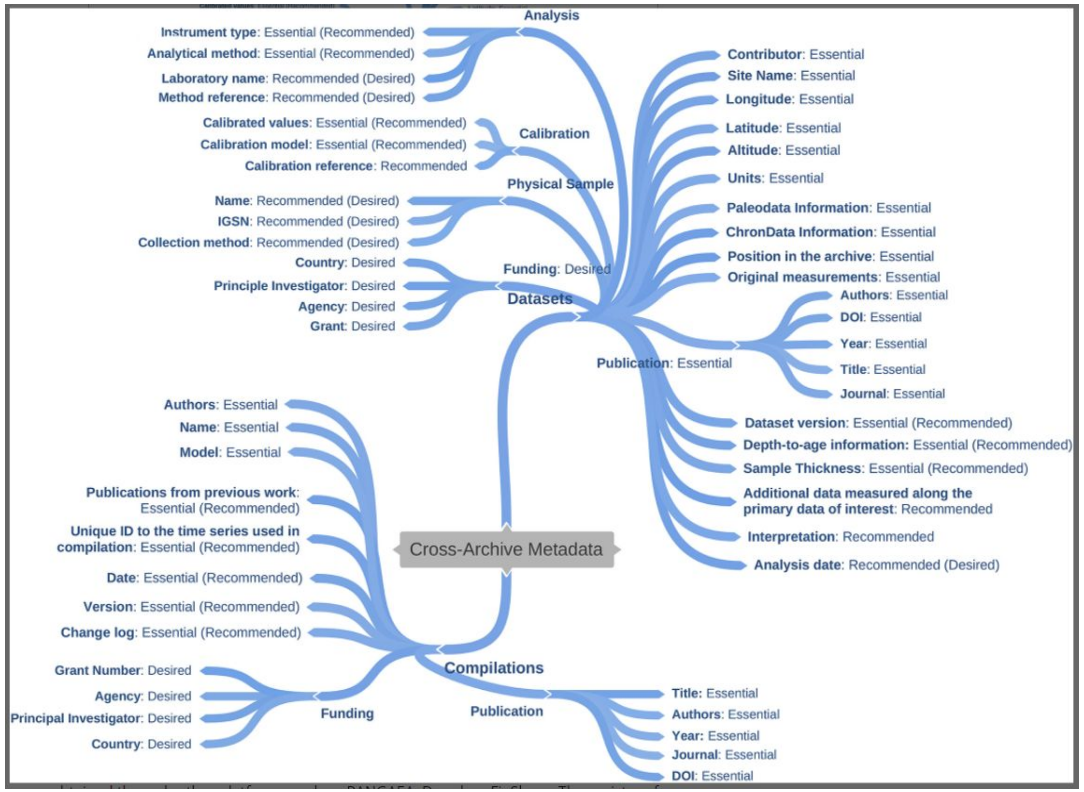
Each with a unique schema on required fields

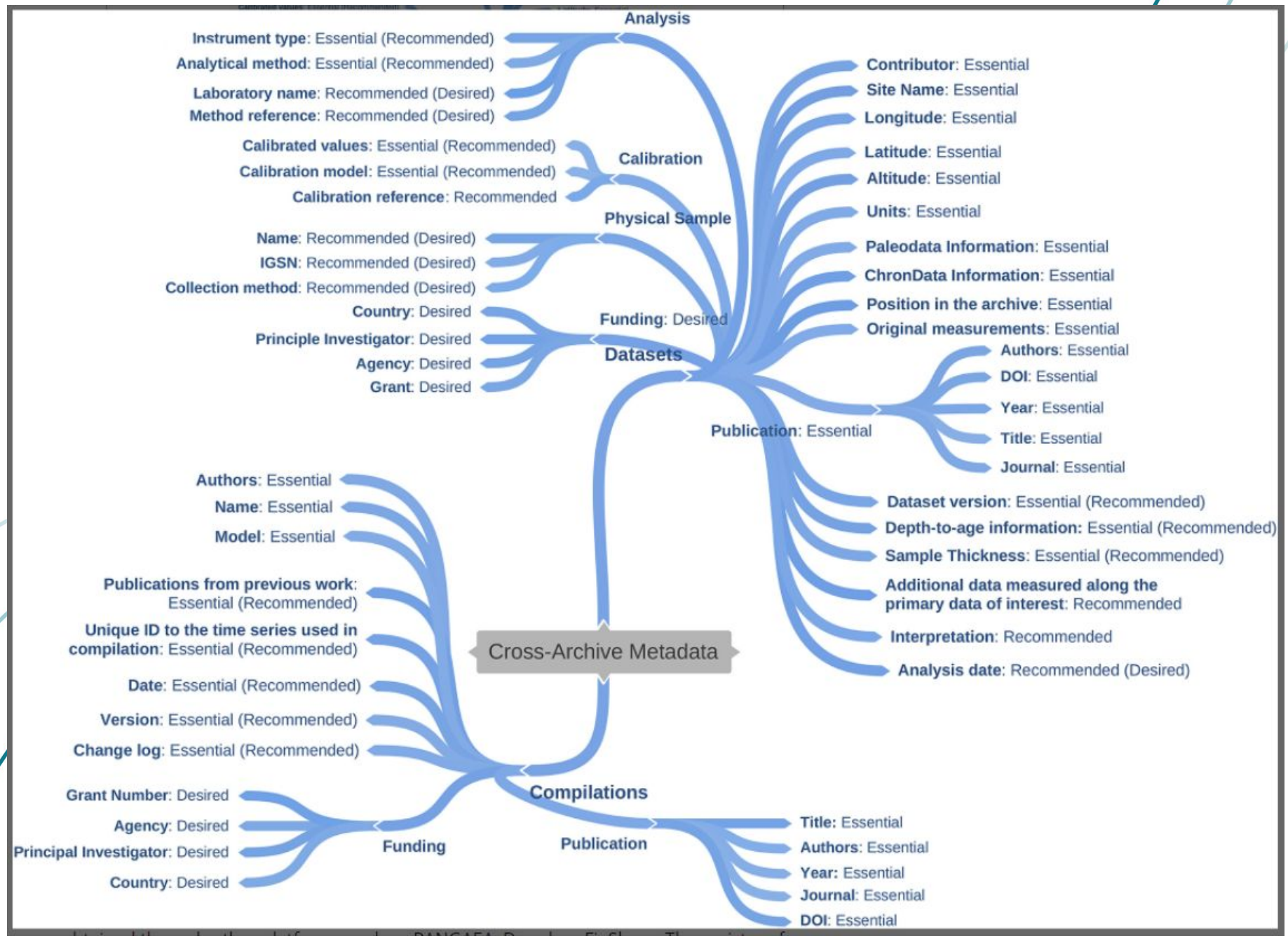
To get here for water data.....



PaCTS 1.0: A Crowdsources Reporting Standard for Paleoclimate Data

..... We need to start here





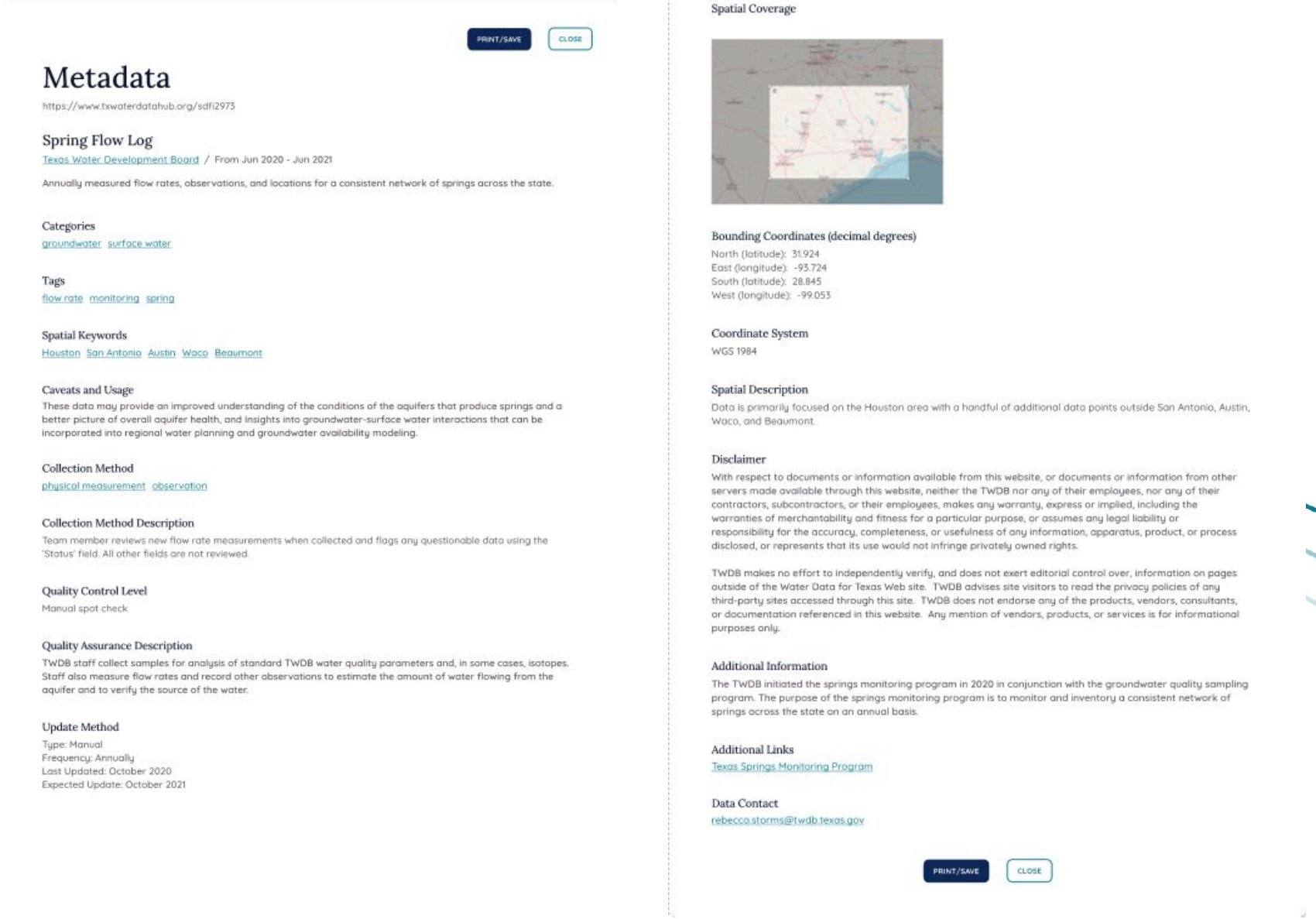
To develop a water ontology for we need to start with cross-archive metad ata.

aka the metadata fields that should be used to describe all datasets

Schema.org

"Is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data.....

A shared vocabulary makes it easier for webmasters and developers to decide on a schema and get the maximum benefits for their efforts. It is in this spirit that the founders, together with the larger community have come together to provide a shared collection of schemas."



PRINT/SAVE CLOSE

Metadata

https://www.txwaterdatahub.org/sdfi2973

Spring Flow Log

Texas Water Development Board / From Jun 2020 - Jun 2021

Annually measured flow rates, observations, and locations for a consistent network of springs across the state.

Categories
groundwater surface water

Tags
flow rate monitoring spring

Spatial Keywords
Houston San Antonio Austin Waco Beaumont

Caveats and Usage
These data may provide an improved understanding of the conditions of the aquifers that produce springs and a better picture of overall aquifer health, and insights into groundwater-surface water interactions that can be incorporated into regional water planning and groundwater availability modeling.

Collection Method
physical measurement observation

Collection Method Description
Team member reviews new flow rate measurements when collected and flags any questionable data using the 'Status' field. All other fields are not reviewed.

Quality Control Level
Manual spot check

Quality Assurance Description
TWDB staff collect samples for analysis of standard TWDB water quality parameters and, in some cases, isotopes. Staff also measure flow rates and record other observations to estimate the amount of water flowing from the aquifer and to verify the source of the water.

Update Method
Type: Manual
Frequency: Annually
Last Updated: October 2020
Expected Update: October 2021

Spatial Coverage

Bounding Coordinates (decimal degrees)
North (latitude): 31.924
East (longitude): -93.724
South (latitude): 28.845
West (longitude): -99.053

Coordinate System
WGS 1984

Spatial Description
Data is primarily focused on the Houston area with a handful of additional data points outside San Antonio, Austin, Waco, and Beaumont.

Disclaimer
With respect to documents or information available from this website, or documents or information from other servers made available through this website, neither the TWDB nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, including the warranties of merchantability and fitness for a particular purpose, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

TWDB makes no effort to independently verify, and does not exert editorial control over, information on pages outside of the Water Data for Texas Web site. TWDB advises site visitors to read the privacy policies of any third-party sites accessed through this site. TWDB does not endorse any of the products, vendors, consultants, or documentation referenced in this website. Any mention of vendors, products, or services is for informational purposes only.

Additional Information
The TWDB initiated the springs monitoring program in 2020 in conjunction with the groundwater quality sampling program. The purpose of the springs monitoring program is to monitor and inventory a consistent network of springs across the state on an annual basis.

Additional Links
Texas Springs Monitoring Program

Data Contact
rebecca.storms@twdb.texas.gov

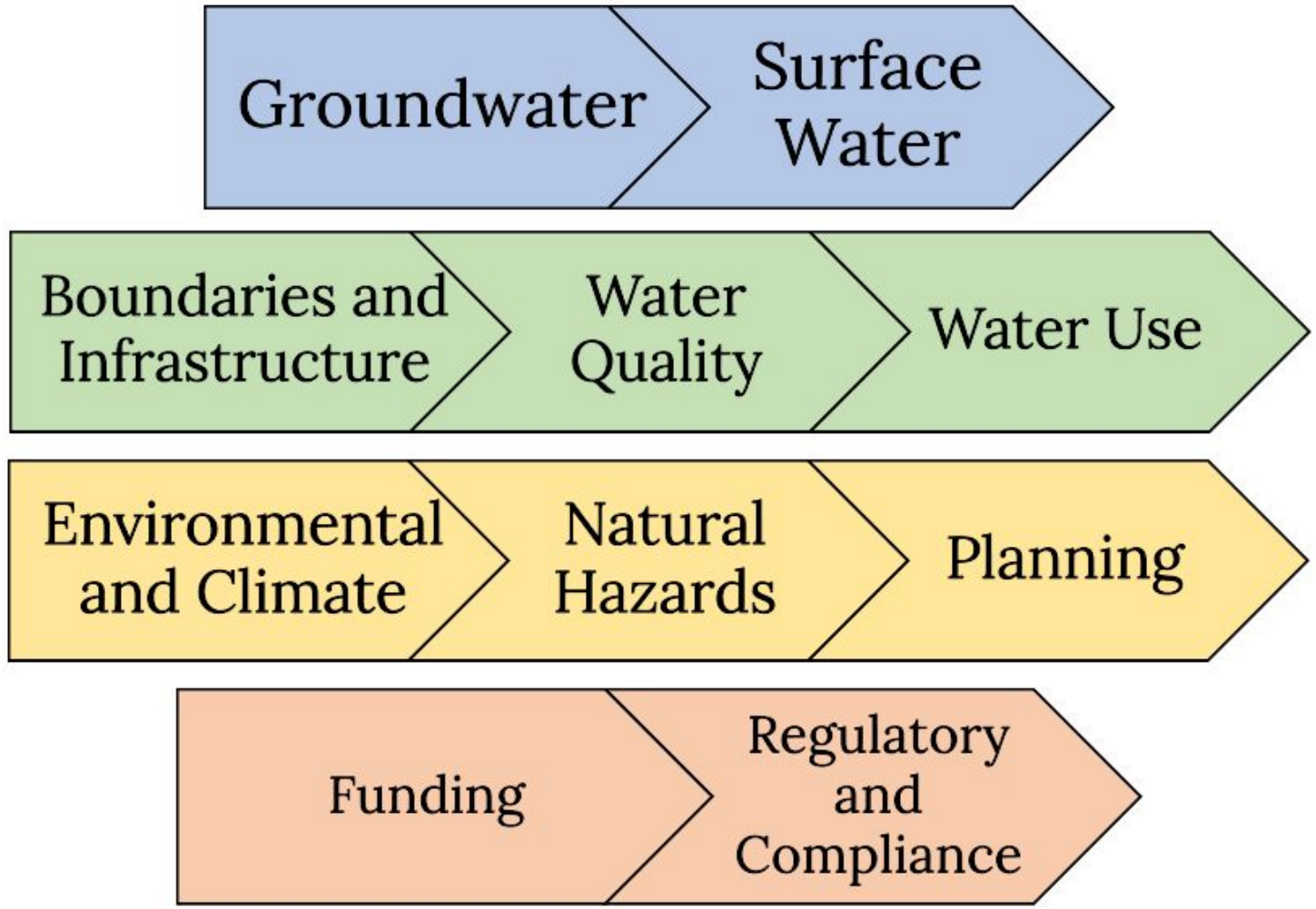
PRINT/SAVE CLOSE

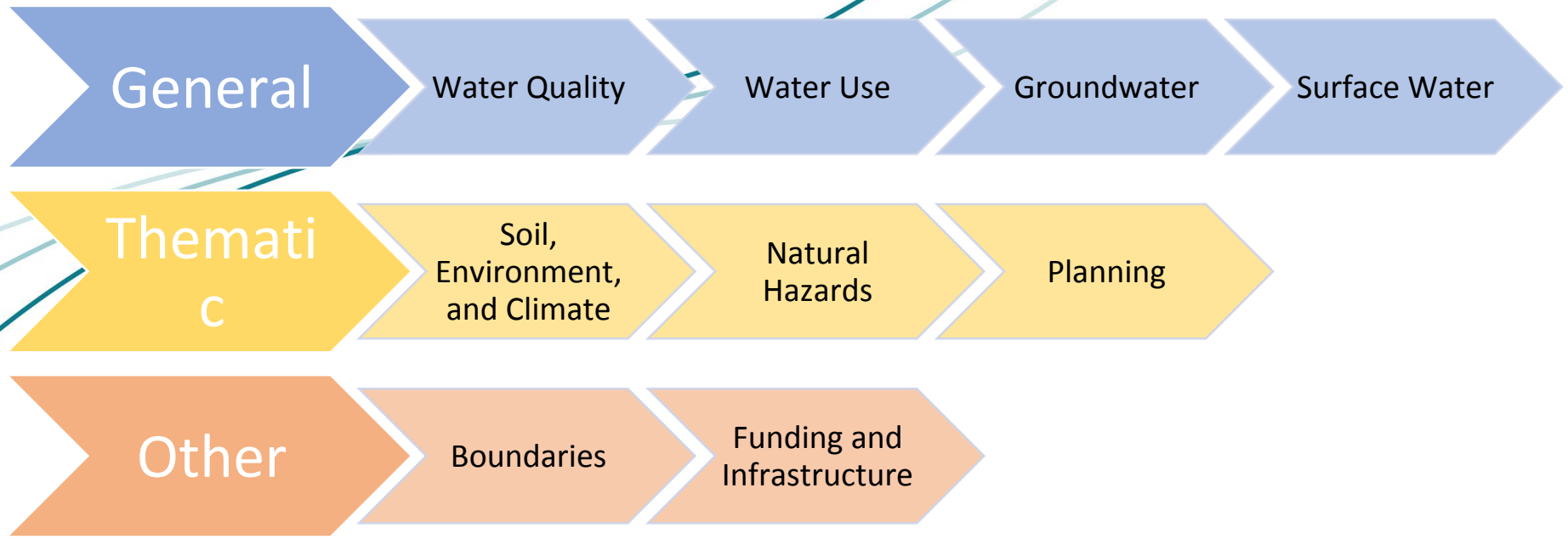
TWDH Metadata Field	Schema.org Field	
Title/Name	name	Automated Groundwater Level Wells
Abstract	abstract	Recent conditions for daily high-water level (feet below land surface) for groundwater monitoring recorder wells across the state.
Organization	maintainer	TWDB
Resource Type		tabular
Resource Type		geospatial
File Location	distribution	https://waterdatafortexas.org/groundwater/recent-conditions.csv
File Location	distribution	https://waterdatafortexas.org/groundwater/recent-conditions.json
File Location	distribution	https://waterdatafortexas.org/groundwater/wells.kmz
File Location	distribution	https://waterdatafortexas.org/groundwater/wells.geojson
Collection Method	measurementTechnique	Physical measurement- sensor reading
Category	about	Groundwater
Category	about	Water Use
Primary Tags	Keywords	Aquifer Water Levels
Primary Tags		Recorder Wells
Primary Tags		Groundwater Wells
Secondary Tags	Keywords	Aquifer
Supporting URL	url	https://waterdatafortexas.org/groundwater
Spatial coverage		Statewide
date range	temporalCoverage	Current Day
Dataset download size		CSV- 9KB
Dataset download size		KMZ- 56KB

Example of metadata from TWDH and Schema.org/dataset

This 'starter ontology' will give us insight on the water data landscape such as:

- Relationship of categories and keywords
  - spatial coverage
  - temporal coverage
- File storage and distribution types


Groundwater
Surface Water
Boundaries and Infrastructure
Water Quality
Water Use
Environmental and Climate
Natural Hazards
Planning
Funding
Regulatory and Compliance


General
Water Quality
Water Use
Groundwater
Surface Water
Thematic
Soil, Environment, and Climate
Natural Hazards
Planning
Other
Boundaries
Funding and Infrastructure

TWDI Purpose and Membership Discussion

Purpose Statement 1

Original

Foster an active, collaborative community of Texas water data stakeholders.

Revised

Foster an active, collaborative, **and diverse** community of Texas water data stakeholders.

Purpose Statement 2

Original

Promote interoperability of existing datasets to support state/local decision makers and stakeholders.

Revised

Advance interoperability of existing datasets to support **decision making on water.**

Purpose Statement 3

Original

Identify and prioritize water data gaps and opportunities.

Revised

Identify and prioritize water data gaps and opportunities **to address them.**

Purpose Statement 4

Original

Promote the adoption of common water data standards and definitions.

Purpose Statement 5

Original

Support **the development of** efforts to make Texas water data FAIR.

Revised

Support efforts to make Texas water data FAIR.

Purpose Statement Order

Option 1

1. Foster an active, collaborative community of Texas water data stakeholders.
2. Promote interoperability of existing datasets to support state/local decision makers and stakeholders.
3. Identify and prioritize water data gaps and opportunities.
4. Adopt common water data standards and definitions.
5. Support the development of efforts to make Texas water data FAIR.

Option 2

1. Foster an active, collaborative community of Texas water data stakeholders.
2. Identify and prioritize water data gaps and opportunities.
3. Adopt common water data standards and definitions.
4. Promote interoperability of existing datasets to support state/local decision makers and stakeholders.
5. Support the development of efforts to make Texas water data FAIR.

TWDI Membership Organizations

Government Agencies

Local

- San Antonio Water System

State

- Texas Parks and Wildlife
- General Land Office
- Texas Waer Development Board
- Texas Commission on Environmental Quality
- Railroad Commission

Federal

- US Geologic Survey

Academic

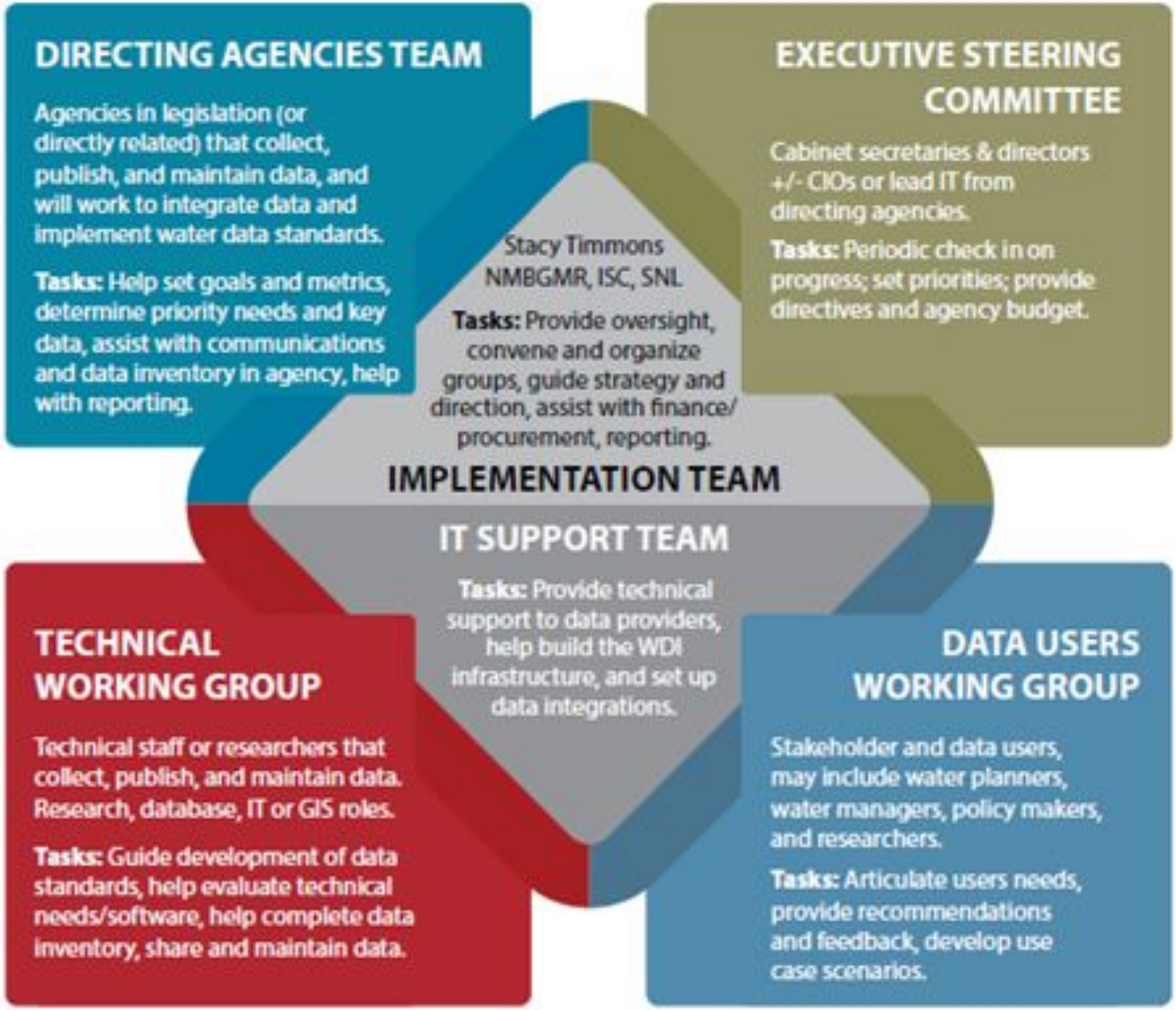
- Bureau of Economic Geology (UT)
- The Meadows Center (Texas State University)
- Institute for Disaster Research for Texas (Texas A&M)

Environmental and Non-Profit

- Cibolo Nature Center
- Environmental Defense Fund
- National Wildlife Foundation
- Cynthia and George Mitchell Foundation

Other

- Earth Genome
- Texas Groundwater Conservation Association
- Internet of Water
- Texas Water Foundation
- Texas Alliance of Groundwater Districts


DIRECTING AGENCIES TEAM
Agencies in legislation (or directly related) that collect, publish, and maintain data, and will work to integrate data and implement water data standards.
Tasks: Help set goals and metrics, determine priority needs and key data, assist with communications and data inventory in agency, help with reporting.
EXECUTIVE STEERING COMMITTEE
Cabinet secretaries & directors +/- CIOs or lead IT from directing agencies.
Tasks: Periodic check in on progress; set priorities; provide directives and agency budget.
Stacy Timmons
NMBGMR, ISC, SNL
Tasks: Provide oversight, convene and organize groups, guide strategy and direction, assist with finance/ procurement, reporting.
IMPLEMENTATION TEAM
IT SUPPORT TEAM
Tasks: Provide technical support to data providers, help build the WDI infrastructure, and set up data integrations.
TECHNICAL WORKING GROUP
Technical staff or researchers that collect, publish, and maintain data. Research, database, IT or GIS roles.
Tasks: Guide development of data standards, help evaluate technical needs/software, help complete data inventory, share and maintain data.
DATA USERS WORKING GROUP
Stakeholder and data users, may include water planners, water managers, policy makers, and researchers.
Tasks: Articulate users needs, provide recommendations and feedback, develop use case scenarios.

Appendix F

Texas Data Hub
User Research Plan

Overview

In order to explore and build an innovative and effective statewide hub for water data in Texas, the team will conduct virtual in-context research to understand how people think about and access data. This research method is intended to be participatory in nature and will include structured exercises and simulations in addition to interview style questions to better understand real human behaviors around this topic and provide participants with creative tools to give more in-depth and realistic information.

Goals

We are seeking to understand:

- How people search for and find data and information they need?
- How people use (water) data?
- What is important to people when evaluating a data source?
- How producers update and share data?

Participants

Research will be conducted with ~10+ participants who meet the below criteria:

- Be willing to sign a consent form, be audio/video recorded, and have screen shots taken.
- Open to virtually meeting from their place of work or home and sharing their environment with the interviewers
    - 5+ data users
    - 5+ data producers
    - 3+ work with water data specifically
    - 5+ from identified contributing agencies (TWDB, TCEQ, USGS, LCRA, BRA, SARA)
    - 1+ legacy data users
    - 1+ legacy data producers

A short screener survey will be used to quickly select potential participants and may be shared with specific groups or people as well as publicly.

Screener Link:
https://forms.office.com/Pages/ResponsePage.aspx?id=jE7QevWSIEuB1hD9iDJYmcjiEVET6elMhjeQIlIRcHNURThQVEY0TUVWMTdSNEFCTjExWkVRVTc4Sy4u

Method

1. Overview

Laura will serve as the facilitator for each interview. One additional core team member will join for each interview and serve as note taker and observer. The facilitator will begin with introductions for themselves, the observer, the research in general before explaining how the session will go and answering any initial questions from the participant.

2. Consent

   The participant will be asked to sign a consent form explaining that they agree to be recorded during the session, and that their information will be shared within the organization for research purposes but anonymized.

3. Background and Data Questions

   The facilitator will ask the participant a variety of questions about their background to get to know them better and give context to how they work with and think about data and water data.

4. Activity Idea 1

   Shadow user as they work with or update data in their normal context. Ask clarifying questions to better understand the workflows, behaviors and shortcuts data users/producers commonly use.

5. Activity Idea 2

   Ask user to search for, find, and select something they've done before and explain their thoughts, reasoning, and motives behind their actions. Goal of this is to give us insight into a user's thought process as they locate and evaluate items for their use, as well as what existing tools and features are comfortable and useful for them.

6. Activity Idea 3 - Card Sorting Game

   Present user with cards that contain different types of data formats, features, or related items and have them rank and discuss their experiences and feelings about each. List of card sort items/terms below:

Search bar	Explore	Catalog	Location
Filters	Related	Query	Tools
Updates	Water data	Download	Links
Real-time	Recommendations		

7. Closing

   Participants will be asked a high-level open-ended closing question about the topic. The observer will be given a final opportunity to ask any additional questions, and the participant will have an opportunity to ask any additional questions.

Script

[While written in a conversational tone, this script is a guide and will not be verbatim]

1. *Start Recording*
2. Introduction *[5 minutes]*

   Hi, [Participant Name]! My name is Laura and I'll be facilitating, and this is [additional team member] who will be helping with notetaking. First, thank you for taking the time to meet with us today.

   We're conducting these research sessions to better understand how people (like you) find, retrieve, and work with data- so thank you in advance for helping us! To be clear, there are no wrong answers and the more candid you are the better for us. The plan is for this meeting to run no more than 2 hours.

   Before we get going, you should have received and signed a consent form ahead of time that described that we will be recording this session for our own notes and recollection with your permission. This meeting is intended purely for research purposes, and we will make all efforts to maintain your anonymity. The form also described that you will not receive compensation. I have your signed form here, but would you please verbally verify that you reviewed and agreed to that consent form?

   [verifies]

   Great, thanks! Now, let me explain a bit about the plan for today. First, I'm going to start with some background questions to get to know you, and then a few questions to better understand how you use and interact with data in general. After we'll do a couple interactive activities and end with some final follow up questions.

   If I ask you anything you're not comfortable answering it's no problem. Just let me know and we can move on to the next one. Because we're trying to build a solution based on real world behaviors and needs, we are interested in hearing and understanding your personal methods and experiences when it comes to finding and accessing data. I just want to underscore that I'm not here to judge or comment, and again if something makes you uncomfortable you do not have to share it and just let me know so we can move on.

   Do you have any questions for me before we get started?

   [answer any questions]

3. Background Questions *[15-30 minutes]*

   Ok, great, so first…

   - Will you start by telling me a little bit about yourself?
   - What do you do for work?
     - How long have you been doing that?
     - How did you get into this type of work?

        - What do you enjoy most about this work?
    - Where are you working these days? Can you give us a tour of your workspace?
    - What do you like to do in your free time?

Thanks, you know, our goal in all this is to understand how people use and think about data. So, let's talk a bit about that…

- What types of data do you typically use or look for?
    - Will you show me how you access those data?
    - Any specific to water? How, why?
- How do you use data?
- What do you use the most?
    - Why do you use those the most?
- Any other types of data do you typically use or look for?
    - Will you show me how you access those data?
    - Any specific to water? How, why?
- How do you typically find and access the data you need?
    - Can you show me?
    - Why do you do it this way?
- What technology do you typically use to produce or analyze data?
    - Can you show me?
- Do you ever need to look for outside data or information?
    - Tell me about that- why, what, when, how?
    - Can you show me?
- What are some of the challenges you've faced with accessing relevant data?
    - Tell me about that- why, what, when, how?
    - Can you show me?
- What has been you experience with legacy data?
    - Can you show me?
    - Do you have any work arounds?
    - Plans for future updates or changes?

4. Activity 1 *[20-30 minutes]*
   Ok, for our first activity we'd just like to shadow you for a bit as you go about your normal data work. Will you please show us what you're currently working on and talk us through what you're doing as you do it? We'll mainly listen and ask you clarifying questions along the way.

   [Ask clarifying/follow up questions]

   Thank you for letting us in on the work you do and how you do it. That was both interesting to learn and helpful for us so we can build something that will take your real workflows into account.

5. Activity 2 *[15 minutes]*

So, for the next activity we're hoping to better understand how you might search and find things on your own, and I'm going to give you a very broad goal and leave the rest completely up to you. Remember there's no wrong way to go about this we're just hoping to see what you'd normally do.

So, for this activity, we'd like you to find and select a new book to read and walk us through how you do it. Please go about this how you normally would be using the sites, tools, and applications that you're familiar and comfortable with. The specifics are up to you, but please talk aloud to let us know what you are doing as you go through this. I may ask clarifying questions or prompt you to keep talking as we go.

[Ask follow up questions]

That was so helpful for us. Thank you so much.

6. Activity 3 *[20 minutes]*
   This next activity is a bit of a game. I'm going to share my screen and *virtually* lay out some cards with different data hub related items. Let me know when you can see these. Now, take a couple minutes to review them.
   - Which ones stand out to you?
     - Tell me about that.
     - Tell me about your experience with these and what they mean to you.
     - If you were to assign emotions to these cards what would they be?
     - Is there a specific time you're thinking of? Please tell me about it.

   [re-orders cards]

   Now, I'd like you to rank each of these in importance from highest to lowest for you in a data hub context.

   [re-orders cards]

   Seeing these ranked, why do you think you prioritized or grouped some of these higher than others?

7. Closing Questions *[5-10 minutes]*

   Today we've talked about data and how it's found, accessed, updated, and used. Take a minute to think about everything we've discussed about how you interact with data and what your needs are. Now, image you have a magic wand that you could use to change anything about how you find and access the data that you need in any context. What would you change or create for yourself? Why?

   [answer and any follow up questions]

Thank you for your time today. We really appreciate you being open to sharing your experiences and opinions to help us better understand data users needs.

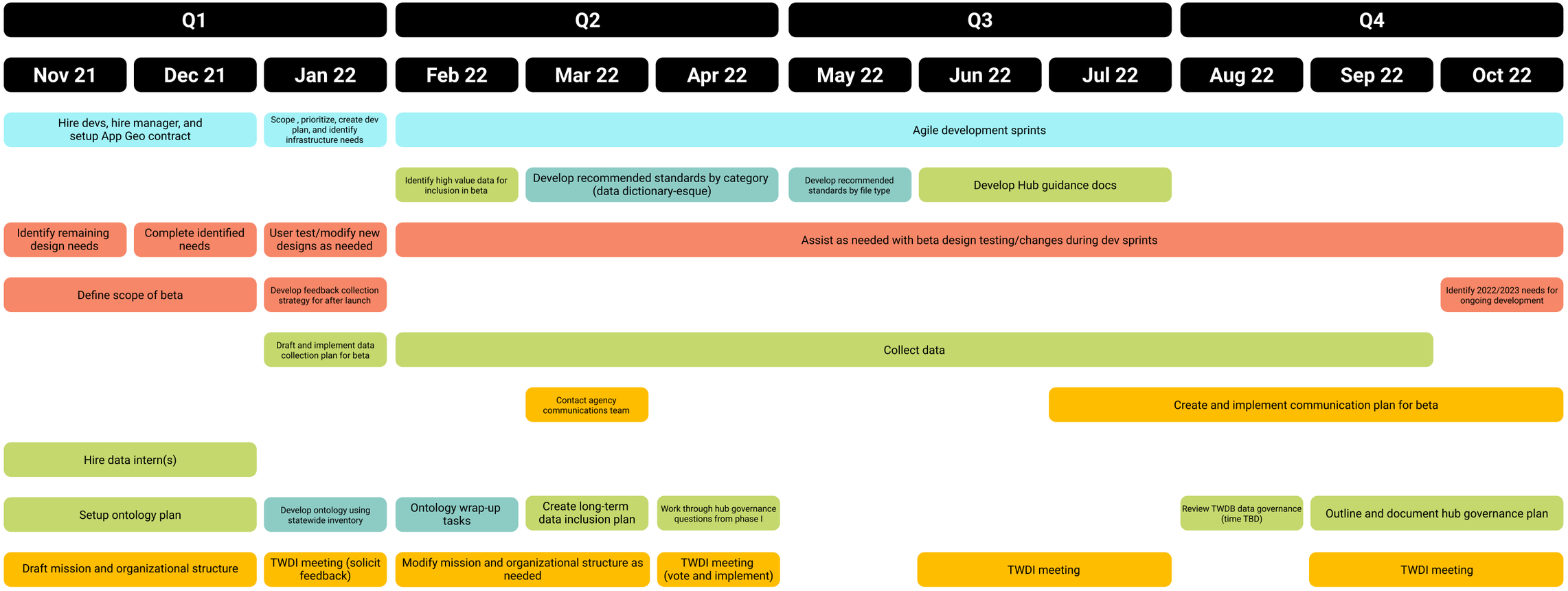
Materials

- Script
- Video meeting and recording software (Microsoft Teams or Skype)
- Consent form

Appendix G

Texas Water Data Hub Phase II (2021-2022) Roadmap

Draft date: 10/27/21



Legend

- Tech
- Data
- Data intern
- Design
- Mixed bag + Leadership

Goals

1. **Build Beta:** Build-out of the hub user interface and technical framework (front- and back-end).
2. **Recommended Data Standards:** Formalization of recommended data standards in guidance documents designed to support hub participants.
3. **Design:** Collaboration with data users and producers to inform framework design and development.
4. **Promote and Launch Beta:** Launch of a beta version of the hub for public or limited public review and use.
5. **Data Governance:** Development of a long-term data governance structure to support the hub over time.
6. **Water Data Stakeholder Outreach:** Increased connection among water data stakeholders across the state, including increased awareness of the hub as a resource.

Appendix H

Keywords	
Absolute Humidity	Flood
ac-ft	Flood Damages
Acidity	Flood Plain
Active Life	Flood Risk
Active Management Area	Flood Warning
Administrative	Forecast
Advanced Wastewater Treatment	Funding
Agency	Geology
Agricultural	GIS
Agriculture	Historical Data
Alert	Hydraulics
Algae	Hydrology
Alluvial	Imagery
Alluvium	Lakes
Annual Groundwater Allotment	Landcover
Application	Landsat
Aquatic Ecosystem	Maps
Aquifer	Marine life
Artesian Well	Measurements
Artisian Aquifer	Monitoring
ASCII	Navigation
Atmospheric	Observed Data
AWRA	Ocean
Bank	Oceanic
Basin	Oceanography
Basin	Petroleum
Bathymetry	Pollution
Bay	Population
Bay	Port
Beach Closure	Precipitation
Bed	Recreation
Beneficial Use	Regulatory
Bioregion	Report
Blackwater	Riverine
Brackish	Rivers
Bridge	Safety
Built Environment	Soil Moisture
Climate	Spatial
Coastal	Storm Surge
Contaminants	Streamflow
Critical Facilities	Subsidence
Critical Infrastructure	Survey
Dam	Transportation
Dams	Usage
Database	Wastewater
Dredging	Water Modeling
Drought	Water Resources
Economic	Water Supply
Emergency Response	Watershed
Energy	Weather
Environment	Wells
Erosion	Wetlands
Estuary	Wildlife

Appendix I

Texas Water Data Hub

Design Criteria

Before we began user research, we questioned if water data are any different than other types of data…

Before we began user research, we questioned if water data are any different than other types of data…

"I'm not working on esoteric, ivory tower academic questions. **I'm looking at things that matter to people in their daily lives.** Being able to look at those kind of problems and be able to bring science in to help…is just wonderfully rewarding."
(Gerald, 565)

"If we can bring all that data, bring it to bear and be able to point it at somebody and laser beam it in a particular person to answer a question, that's where it's at. **Why are we here if we're not helping people?**"
(Edgar, 1407)

"Always really satisfying when you got that call from somebody that was looking for that water well on their grandfather's property or something…and then **finding that and giving it to them, and they're so excited…That was really cool.**"
(Jimmy, 342)

"I think I've always had a connection with nature, but it wasn't until I…got to see how water was managed and what hydrology looks like on the landscape, **living so closely to it and understanding the factors that play into that flow on a day to day level**."
(Duncan, 751-752)

Working with water data is more than just a job.

It’s important and personal.

We also heard some of the struggles water data practitioners are facing…

We also heard some of the struggles water data practitioners are facing…

"It's still fairly **time consuming to go through and do this exercise. It's just not very efficient**…and then the other challenge is sometimes **you just don't know what's out there.**"
(Erik, 268-269)

"They've developed those naturalized flows a different way from the guys at X…who's done it differently to the folks in the [other] basin…and **if we didn't have those contacts we wouldn't have necessarily known that information was available.**"
(Gene, 1936)

"I simply **don't have the time to…learn how to get into the database files and pull everything I need** out of these 10, 15, maybe some of them are even 20-year-old GAMs."
(Gerald, 633)

"It seemed like working with that **data had a lot of legacy issues, like there were different managers, there were different procedures, and all of them presented their own issues.**

Where they may be able to negotiate those issues…**I don't have the working knowledge to do it, so they were real hesitant to give it to me."**
(Duncan, 819-821)

Texas water data is fragmented and locked away making it difficult to understand, value, and use.

Design Criteria

Five explicit goals the project must achieve in order to be successful.

01

A water data hub should provide a central location for water data that reflects the entire Texas water landscape.

KEY INSIGHTS

- Data access is dependent on who you know.
- Project and legacy data are especially siloed.
- Practitioners often lack a way to publicly display or share their work.
- Data are accessed through many different locations and methods.

Data access is dependent on who you know

Without an easy way to understand what data exists and how it may be accessed, data searches rely heavily on contacting data owners for personalized assistance. For those not in “the know” this can turn into a guessing game or seemingly impossible task.

Data owners also lose valuable time responding to vague, informal, and unplanned requests.

> “It's not available on their website. We reach out to different people trying to hunt down- find the right person to get in contact with, and they email us the data.”
>
> (Julie, 43)

> “It starts with a person, ‘hey, we're looking for sediment reduction projects’, and then someone will go, ‘oh something's ringing a bell, look at cycle 24, this funding here and go over there and this is where you look in that particular location’.”
>
> (Edgar, 1427)

> “Theirs is not online at all. I have to talk to a guy, he's called their Tron…and I ask him for a dump, which is just a file, a flat file…so, they don't even have it. It's not even on their website.”
>
> (Mark, 996)

> “You know, and if we didn't have those contacts we wouldn't have necessarily known that that information was available.”
>
> (Gene, 1936)

> “I just start emailing… ‘I need this. You know, I can't find this model on the page. Can I have this?’ And they're inevitably, I'm sure tired of hearing from me, but they inevitably get back to me.”
>
> (Gerald, 611)

> “But there's probably available data out there that I don't have that same working relationship with, or knowledge of the organization.”
>
> (Duncan, 779)

01

Project and legacy data are especially siloed.

Legacy and project data are often stored away and archived. Due to the specific nature or format of these data, owners may not see a wider value or need to advertise or share the information. These data often just end up taking up space and may never be used beyond their initial purpose creating gaps or duplication of work.

> “There's an agency scanning project. So, that is kind of part and parcel. We have a whole room full of legacy data that's paper copies.”
>
> (Mark, 1036)

> “We don't have a database -- a public facing database, it kind of just lives in my world. And of course there's a lot of other people in [employer] that are probably doing the same thing.”
>
> (Duncan, 772)

> “We tend to do things very study-specific. So, I build -- I go through and people spend the time of putting the two data sets together. And then once that study is done, that database is archived and moved over.”
>
> (Thomas, 1170)

> “We archive it for ourselves and for the client. But we typically don't share that with anyone else. There's a lot of that kind of data that's collected. Unfortunately, it's just not -- people don't know about it…therefore, it's not available.”
>
> (Gene, 1924)

01

Practitioners often lack a way to publicly display or share their work

Practitioners are doing amazing work they are proud of. As the ones producing and maintaining the data they are often in the ideal place to share and explain the work, but may not have a clear or easy way of doing so.

Intermediate data may be overlooked due to it's transitional nature, and final products may be aggregated or formatted in a way that is not easily accessible.

"Yeah, this is hanging up in the cube window. So, this is an example of one big aspect of my job…And so, that means scientific reports, that means technical reports, presentations, stakeholder meetings, things of that sort."

(Mark, 968)

"Your idea of being able to upload data to it, I think that's a really good idea. And if I had an opportunity to upload…data I collected, I'd love to take part in that."

(Duncan, 922)

"We want this data to be out there like some Broadway star kicking those legs up or whatever."

(Edgar, 1452)

"The hope is that we could get it out there just for everybody to use and anybody that needs that data..."

(Jimmy, 366)

"It is a final report, so it is available. I pass it out like candy on Halloween to people who are interested in it."

(Edgar, 1465)

"Supporting data can always be really helpful. Giving the final data to a third party is great, but sometimes they want to see how can they recreate it or you need those in between steps to show how to get there."

(Julie, 9)

01

Data are accessed through many different locations and methods

Data become difficult to track and maintain when accessed from many different sources, methods, and people. Without a dedicated tracking system practitioners must do the work to create their own or risk losing important source information for documentation, updates, and collaboration.

> "I think us in the community, we are accessing the data through all the different canals and methods that we can and often bringing those together for our own separate databases so that we can then manipulate the data."
>
> (Edgar, 1325)

> "I honestly can't remember where we mined this data out of. I don't know if it was… I'm sorry, apologize, I can't remember exactly where we mined this data from."
>
> (Chuck, 1775)

> "If you'd log on to this water data portal, you're thinking 'OK this is where I need to go, this is home, this is where I'm going to get my data.' You get that happy feeling, because you're not looking in ten different places for your data. You know you're going to be able to get it here."
>
> (Gene, 2014)

> "A one stop shop to connect to everything rather than bookmarking, you know, all of our variety of data sources, that would be pretty cool."
>
> (Julie, 142)

02

A water data hub should establish automatic and easy ways to share data and updates.

KEY INSIGHTS

- Users expect immediate access to the latest and greatest data to do their work.
- Unexpected changes disrupt workflows and checking for updates is a waste of time and effort.
- Consistent data sharing is only possible through formalized agreements and automated procedures.
- Data producers often lack the time, resources, and technical abilities required for data retrieval and sharing.

Users expect immediate access to the latest and greatest data to do their work.

People have become accustomed to on-demand access to the information they need. Especially as more organizations modernize their reporting and data entry forms, turn-around time for published data should be decreasing as well. Data is becoming a key part of reporting, and operational decisions and as such needs to be available immediately and account for on-the-fly changes and corrections, when necessary.

> "I really think if I could click a magic button to say give me what's new in terms of someone's data, that's something that I need. I don't really ever see that, you know, being up to date."
>
> (Mark, 1087)

> "When you're in a workflow, the last thing you want to do is to wait around for an answer that should really be readily at your fingertips. Often we'll just go, we'll do the ticket, and then call the person direct to get the answer, try to get a workaround solution right away."
>
> (Edgar, 1413)

> "We have these different alarms set up…like if it gets above a certain flow…that means we're getting some pretty significant runoff and rainfall. We need to start looking at it."
>
> (Chuck, 1798)

> "If it's a direct connection to a display portal, a dissemination viewer, it makes it really great because as long as it's always calling and refreshing the data, you always know that you're getting any updates, or any changes that's been made to the source data. You don't have to worry about maintaining. You just worry about the stuff that you're collecting."
>
> (Jimmy, 433)

02

Unexpected changes disrupt workflows and checking for updates is a waste of time and effort.

When users do not have clear update information they must spend time checking, searching, and tracking their data. Practitioners often build code and workflows around data specifics which means that unexpected or unplanned updates can break things and force users to take time re-adjusting their setup to accommodate.

> “This one's actually really great because they put their release date and their next, their most recent release date and their next release date, so we know…we'll just come back in June of 2021.”
>
> (Julie, 48)

> “When we make updates to the X, we actually have a listserv that we use. I'll send an email that says this model has been updated for these specific things and give people contact information if they want that data. Yeah, we don't just update data sets and not tell anybody. That usually doesn't work very well when you do that.”
>
> (Sheila, 1656)

> “The other big issue I have just for me is I have to report to the ledge who took salinity data on the coast and where. Was it operational all quarter? And these links that I've made to other people's databases, they don't tell me when they add a new station or remove a station or do anything. I find out haphazardly that they did…”
>
> (Mark, 1003)

> “The way we've dealt with that is when our contractors go out and pick up data from X, just give us the date that you got it. If something happens later on…it's required some going back and doing double work, which as a state agency that you have limited resources, really. It's been a challenge in that respect.”
>
> (Sheila, 1571)

02

Consistent data sharing is only possible through formalized agreements and automated procedures.

Data sharing based on informal agreements or personal relationships may be inconsistent over time due to availability and other unknown factors. Automated and established data sharing paths are more reliable and take less effort in the long run.

> “There are so many different groups, and you can't make X do anything. A lot of the frustration I have is, frankly, some of them won't even give me the time of day. I've emailed certain people from other agencies, and I've never got a response back...”
>
> (Mark, 1125)

> “We work with the International Boundary and Water Commission. It's a treaty. The river is governed by a treaty. The data that gets put out is agreed to by both countries.”
>
> (Sheila, 1533)

> “When we know we have something that the other group doesn’t, that's doing this very similar work, we will try to go out of our way to share it, but…it's a pretty informal process.”
>
> (Erik, 275)

> “We're using more code to make things so things can be repeated in a similar fashion over time, and that code is put in a repository so people can get it.”
>
> (Thomas, 1172)

> “We share data with the Corp constantly. We have it set up I think to hourly. If anything gets updated, the Corp picks it up, and then the Corps sends us their data.”
>
> (Chuck, 1736)

02

Data producers often lack the time, resources, and technical abilities required for data retrieval and sharing.

Data practitioners rely heavily on IT professionals to assist them with data manipulation and technology setup of their own and outside data. While data users are often extremely knowledgeable in their own specialty or realm that doesn't always translate to other technologies or analysis.

> "The technology, I really rely on the PM, and unfortunately we have to use a small platoon of IT experts to help us access the data."
>
> (Edgar, 1408)

> "Then our IT department developed a report that pulls that data in from our database that can tell how much the aggregate of all of those..."
>
> (Chuck, 1766)

> "This one is a little bit more complicated when it comes, once we download it. It actually, it comes in a strange format…Normally, we have to work with, when we reach out to IT DBAs to get the data, because it typically…it comes in this really strange format that we have to ask them to put it together into an Access database because we aren't able to connect it typically."
>
> (Julie, 52)

> "In the older methodology, we actually worked with a consulting group and had them look at our methodology, and we made updates accordingly. We didn't implement everything that they said. It just depended upon resources and time and things like that, what we could implement."
>
> (Julie, 98)

03

A water data hub should provide intuitive methods to efficiently search and download data.

KEY INSIGHTS

- People don't trust high level search bars, and instead prefer to navigate groupings or search at a record level.
- Location based search and indexing is key.
- People are on a mission to find and download data as quickly as possible.
- Data searches are specific and users don’t want unnecessary information.

People don't trust high level search bars, and instead prefer to navigate groupings or search at a record level.

Google has set a high standard for search bars and users are disappointed when others don’t measure up. Practitioners prefer instead to use search bars within datasets to find specific records because search criteria may be more easily identified. Categories and filters are a more intuitive way for users to locate datasets at a high level.

> “Search bar, that seems a good -- a good thing if it's connected to the data…If my search results return information about the data and not something else…but people put search bars in every website these days and you don't always -- you're not searching what you think you're searching sometimes.”
>
> (Duncan, 893)

> “Or is it going to do like a lot of the search bars do, and it goes out and it also gives you results from outside the website, and suddenly, it's giving me BEG reports and groundwater model reports, and just a whole bunch of stuff that again, that's not why I'm here at this website. I'm looking for data. So, it could be useful, but it could be just annoying and cluttered.”
>
> (Gerald, 712)

> “The search bar, I think, is the saddest one…my last resort, you're forcing me to just magically do a search, and I mean, have you ever done a search on our own internal web or our own website? It's probably the saddest results, it's the closest to soul crushing.”
>
> (Mark, 1088)

> “You would think search function would be your easiest kind of one stop shop to find stuff but really my typing in a word to a search function has really, rarely provided great results for me.”
>
> (Edgar, 1499)

> “You know, surface water people often just do surface water stuff, right? Reservoir flows and management and stream flows and all that, and groundwater people, well, I'm looking at a mile underground. I don't care about all, you know? And then, the atmospheric people are kind of the same way.”
>
> (Gerald, 736)

03

Location based search and indexing is key.

Location is a key component for searching and evaluating data because users can quickly view in-context what is available. Being able to assess spatial relationships within or between datasets is also extremely important. Water is not siloed in the real world and is influenced by relational nearby factors which may be identified through a map interface.

> “How cool would it be to have this data portal where, you know, you're interested in flooding for example. You click on the flooding tab and it pulls up this map and you zoom in to your area of interest. Maybe you draw a rectangle and that downloads all of the information on floodplain maps for that particular region.”
>
> (Gene, 1984)

> “You just click on that grid, and you get anything and everything that was related to that grid. All the old historical reports that are related to that grid, all the water wells in that grid, all the reservoirs on that grid...”
>
> (Jimmy, 459)

> “I wanted a couple discharge and stream temperature, at the least, so I found all the stream temperature sites and discharge sites that were coupled, and also looked at dissolved oxygen, if those were available as well.”
>
> (Duncan, 766)

> “To efficiently be able to go and identify which sites are nearby that impact my sites or have a relationship with my sites for the various agencies and then to query that data out using the same time frame that I'm looking for…then that would be magic.”
>
> (Edgar, 1451)

> “Well, that's a pretty data-sparse environment, right? So, you know, a location tool like that would be kind of cool to know quickly and you could just see, ‘oh, we have tons of data for that location. Or, hey, sorry. We've got, you know, basically, nothing for that place.’”
>
> (Gerald, 709)

People are on a mission to find and download data as quickly as possible.

Most data users do not have the time to aimlessly browse or explore the water data world. We should make the process of finding and accessing needed data as short and sweet as possible. Usually, the goal is to download data to bring it into a personal workspace where it may be manipulated, combined, and analyzed. We should provide the ability to download quickly and easily without long drawn-out workflows.

> "What I want is I want some data. I know I want data at this place, at this time. I just need to know where to get it from. Does your site have that or not? I don't really care that you house data, water data for this or that or this or that. I could care less. That's not what I'm looking at your website for, you know?"
>
> (Mark, 1101)

> "Yeah, it can be frustrating. I worked there and sometimes I'm going to go find that thing again, and I get lost in the web of links. I'm like 'where is that thing again?' Luckily, I have a couple things bookmarked, but sometimes they're buried. And again, the search bar doesn't work very well to find it."
>
> (Erik, 316)

> "...and to be able to go through a discrete section of disqualifying data through a filter to where you're eventually only looking at what is specific and of interest to you and if that process could be done quickly in such a way that you wouldn't be drowning in, you know, a thousand line Excel spreadsheet or something like that..."
>
> (Edgar, 1504)

> "Or maybe you do find the data and download it. Then you want to get back and you got to click like six backwards -- six times backwards, in order to get back to where you started. It would -- I don't know that it necessarily needs to be perfectly seamless. Something better than some sites provide right now would be awesome."
>
> (Gene, 2050)

> "Then you could favorite your reports so that they're always at the top. Then everything else is just underneath that. I think something like that, or even if it tracks- if the datahub had your most recently accessed or most frequently accessed or something like..."
>
> (Julie, 174)

03

Data searches are specific and users don’t want unnecessary information.

Data practitioners usually have a specific need or goal in mind for what they are looking for, and do not want to waste time exploring beyond these needs or download extraneous information. Additional information clutters up data work and could lead to misunderstandings. The ability to make small changes and cut unnecessary information out before download is valuable.

“You got to be able to get it in a format, in a format that you can use and utilize, and filter out all the stuff you don't need, especially if it's an extremely large database.”

(Chuck, 1864)

“When you're looking for a data set, you don't necessarily want everything that's available associated with that. You want to be able to filter down and only download what you need.”

(Gene, 2013)

“You don't want to download everything. I think it's important to be able to filter things down before you get to the download aspect of it.”

(Jimmy, 520)

“It's all about what I'm here for, like my objectives. I'm not, I mean, showing me related things or explore or whatever, that might be cool, but it's just not what I'm here for.”

(Gerald, 727)

“Most of the time, I don't spend a whole lot of time exploring and figuring out what data is out there and how it could benefit me. That would be great, but I just don't have the time in my workload.”

(Julie, 159)

“I want to explore your site to find do you have data I want to grab you know? Is this even the right site…or is it one of 10 other websites you’ve got?”

(Mark, 1091)

04

A water data hub should emphasize clear communication and documentation to build trust and understanding.

KEY INSIGHTS

- Recurrent and predictable workflows lack useful documentation if decisions are not captured in the moment.
- Defined roles and direct communication reduce confusion and potential mistakes.
- Successful data sharing depends on trust between producers and users.

Recurrent and predictable workflows lack useful documentation if decisions are not captured in the moment.

Users learn and make decisions in the moment as they are working with and exploring data. Without a way to document or save this information it becomes difficult to remember later when the workflow or task is repeated. Lack of documentation can make collaboration with others difficult.

> “But a lot of the workloads that we've done to put that database together aren't really captured and therefore whenever that next study comes up 10 years down the line or five years down the line, that entire workflow is reinvented…”
>
> (Thomas, 1171)

> “We're going to go through and QC all of this data again when we do the process. I realized that I should be keeping up with all these notes because we're just going to sit down and do this all over again. And that's going to be a little frustrating.”
>
> (Julie, 91)

> “Giving the final data to a third party is great, but sometimes they want to see how can they recreate it or you need those in between steps to show how to get there.”
>
> (Julie, 9)

> “I keep my phone on me, as most people do, and I've got the notes section. If someone tells me about a good movie, or a good series or a good book, then I'll write that and refer back to it later. Because I can never remember by myself.”
>
> (Gene, 2003)

> “I don't know if it'd be helpful. The process doesn't really tell you how we -- what's going on in the background. I'll open this up. I have some notes here on kind of how to do it.”
>
> (Chuck, 1801)

> “Within the code I'm writing text along with that describing my steps. That Word document that I showed you is basically just helping me keep track of the various iterations of the code.”
>
> (Duncan, 846)

Defined roles and direct communication reduce confusion and potential mistakes.

Users need to directly communicate with data producers or owners for questions or concerns. An added benefit is more eyes to catch and correct errors. Defined roles and responsibilities will be important in the hub to make follow up possible, and to ensure maintenance and updates stay on schedule.

> “I work very closely with my colleagues on X team because they are the ones that actually reach out if we think that something's wrong and we think it needs to be corrected.”
>
> (Julie, 36)

> “We had higher confidence in what was in our database…what came from small entities we didn't necessarily have high confidence in so we had to go back to those entities and, ‘hey, is this really this? Is this location correct? Just finding out some of that information over time.”
>
> (Thomas, 1284)

> “That's part of what the contractor's responsibility is to make sure that if it's a major water user, they'll reach out and say, ‘hey, you reported like two in this month but you're a city of 100,000 people. So, it was a typo. The water plant was down, but I actually reported it over here. OK, that's fine.’ Those kind of things.”
>
> (Sheila, 1615)

> “And so eventually, after wading through that soup, I found out that no single entity had the resources or capacity, nor did any state agency or regional planning authority want to do that.”
>
> (Edgar, 1438)

> “Good delegation of work. The way they did delegate with…my position, there was a very clear division. It's like your job is data management. Sure, if you have some free time, learn some geology, but don't be thinking you're going to be-- like this is your job here. And I'm like yes, that's important. Then there's not someone feeling bad like ‘oh, why am I doing more of the programming, and they're doing more geology.’”
>
> (Erik, 221)

04

Successful data sharing depends on trust between producers and users.

Data users often look for official data sources from owners they trust to provide authenticated information. Data producers are hesitant to share data if they do not trust that it will be used appropriately. Complete documentation, metadata, and explanations of caveats and uses is necessary to build trust on both sides.

"I would look for, you know, official kind of sources, right? Like, dot-gov kind of addresses, FDA. I happen to know that this one is FDA, right? I would look for official kinds of sources and I would ignore/skim through sales kind of stuff, shopping kind of stuff."

(Gerald, 681)

"If you look at weathermen, people always poopoo on the future. Well, it says it's going to rain or 50 percent is going to rain and it never rains. Or that said, there's going to be this huge, massive storm that's going to hit me. That's what the model predicted, but it didn't. And then it's like, "Oh, well, why should I believe anything that those prediction models say?"

(Jimmy, 454)

"In some instances, it will be that the client doesn't want us to share that information. I think the other potential impediment to sharing that is, it's really important in my business to thoroughly document how that information has been collected."

(Gene, 1925)

"The data that I've collected has not been made public yet. I do want it to be, it just hasn't matured in that way yet."

(Duncan, 774)

"We're trying to get the QA/QC completed and we're probably 85% completed. At the point that we're pretty confident that the data is mostly complete, we'll start serving that data out..."

(Sheila, 1557)

05

A water data hub should assist statewide data interoperability efforts through standards and curated datasets.

KEY INSIGHTS

- Standardized datasets and common identifiers provide quick access, joins, and comparisons.
- Without a coordinated effort across agencies to standardize data, practitioners must manually clean, re-organize, and re-format data.
- The handling of legacy and unstructured data is deprioritized and piecemeal because of unclear standards and value.
- Development of decision support tools and other innovations are not possible without good data.

Standardized datasets and common identifiers provide quick access, joins, and comparisons.

Large statewide standardized datasets help users quickly understand where data is available and access them. It is usually much simpler to remove unnecessary data than it is to add missing data. Many users are looking to join and compare data from different sources or types to gain insights, context, or supplement their existing dataset.

> “The more that you can get those into one common area where you can query them together, you know, that they're kind of in the same format, the same environment where you can look for those commonalities...”
>
> (Jimmy, 417)

> “You know, we basically are creating a bridge table of all the different things that you could name of the site. So that way, when we identify, okay, this is, you know, USGS calls it this, so I need to make sure to enter my database, this is the identifier that USGS uses for this…”
>
> (Jimmy, 484)

> “You know, working with data from New Mexico, it was relatively clean identifying, you know, because they didn't identify across those state lines for the most part. But then the identifiers from one state to the next were completely different.”
>
> (Thomas, 1169)

> “Sometimes it's just browsing around in areas where I don't know there's something…also, I will type in a number- generally not a name, because that's never reliable. But a number, the unique number. Like the state well number for water wells. Or the API number for oil and gas wells. I'll often put those in a search box to take me to the location and look at what's around it.”
>
> (Erik, 304)

> “There's the datasets. This one is much easier because we can literally just…click on our state, and it just downloads…So that one's definitely very easy for us to access.”
>
> (Julie, 47)

Without a coordinated effort across agencies to standardize data, practitioners must manually clean, re-organize, and re-format data.

Manual manipulation of source data in preparation for use is a time consuming but necessary task. While some sort of adjustment will likely always be necessary there is a large and valuable opportunity for the hub to reduce the amount of time, effort, and potential duplication that exists for setup data massaging.

05

> “And it seems like that cleaning of data workflow is always just such a large step before you can do anything. Then you find that your whole budget is blown because you've spent the last year cleaning the data.”
>
> (Thomas, 1302)

> “I know there's a bunch of data layers just for water segments. That's always a funny question because each agency has a different way that they express water segments...”
>
> (Edgar, 1356)

> “It takes some effort…It's easy if you have all the data in a consistent format with clear -- this is exactly what this data is. That's always helpful. You know what you're getting, in the format you're getting it in.”
>
> (Sheila, 1574)

> “It's not convenient to go to different -- all these different entities when you're talking about these large datasets, or having multiple sites, because they're in different formats, and they take a lot of care to clean and to organize in a similar way…if they have data, that's great, but each different entity requires handling. The relationship, but also the data. It just takes time.”
>
> (Duncan, 783)

> “Well, if I was a king for the day, I would force everyone that collects data to collect data in a unified way and that they actually come in agreement for what data is recorded, what needs to be recorded and to organize it in a fashion. Separately, if you can't organize it that way, you write a conversion program to convert it in this way.”
>
> (Mark, 1118)

> “That one, it's a bit more complicated. It's easy to get a hold of the data, but once we have it, the amount of work that we have to put into it to make it make sense for the state, it's a lot.”
>
> (Julie, 26)

The handling of legacy and unstructured data is deprioritized and piecemeal because of unclear standards and value.

Organizations and users are dealing with legacy and unstructured data differently with varying levels of success. Standards and best practices for how to handle these data types would benefit the entire community and make these data more accessible.

> “And he's essentially taking publications and just going through them line by line and copying and pasting. And I'm sure he's automated it, but that's what it would take.”
>
> (Duncan, 824)

> “A lot of the reports that we write are on our websites, so you can go and look at our reports. They're not peer reviewed or anything, they're just reviewed internally and then put on the web. This particular study has not been -- there has not been a report and the data has not been made accessible.”
>
> (Duncan, 771)

> “Then I would go to the same information, except they would be the scanned paper maps, like I say, I'm not going to pull them into here because it's a pain. It will freeze up the computer just because there's so many of them and they're so big...”
>
> (Erik, 255)

> “So, we have all these studies that are either not digitized or only part of it is digitized…I'm redoing these studies, and in these studies, the data lives here off of this website. This is all hardcoded stuff, these are images. These are not -- this isn't dynamic of any sort.”
>
> (Mark, 1038)

Development of decision support tools and other innovations are not possible without good data.

Quality data and access is the foundation of decision support and visualization tools which are often more specialized for specific audiences or events. The hub can best support the water data community's work by focusing first on data.

> “As that information has become more readily available and more extensive and, models and the associated infrastructure are quicker and quicker, those tools have improved a lot.”
>
> (Gene, 1955)

> “This data, the use of it is to make a decision on whether someone can get a water right, to make a decision on whether a project in the state and regional water plan can go forward, and to be able to defend those decisions in court, because our staff has very strong legal component.”
>
> (Sheila, 1606)

> “But on the other hand, if we, if maybe this, would be looking to inform a piece of a statement, or we're looking at policy changes, we would look at other cites, as in citations, that we can then lean on as something that's been peer reviewed, published, that's already been kind of chewed through the ringer of experts who've already vetted that information. So that if we want to use something to inform a policy element, we'll often look for something that's been fully vetted and used in their output to then direct towards our particular purposes where we can find a fit.”
>
> (Edgar, 1377)

> “I don't know if the state realized this when they embarked upon this LIDAR data collection initiative, holy cow. The information is going to be so useful for so many applications, not least of which is going to be on the floodplain mapping stuff...”
>
> (Gene, 1949)

05

A water data hub should…

1. Provide a central location for water data that reflects the entire Texas water landscape.

2. Establish automatic and easy ways to share data and updates.

3. Provide intuitive methods to efficiently search and download data.

4. Emphasize clear communication and documentation to build trust and understanding.

5. Assist statewide data interoperability efforts through standards and curated datasets.