

# 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

# Participant Profiles

Profiles were created for each organization to share background and context.

These were for internal team use only and have been removed from this version of the document.

# 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 progress if the burden is too heavy.