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 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 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)

03

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)

04

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)

05

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 are 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)

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.