



Building an Open () Data Hub

Laura Sepulveda and Taylor Christian

Project Background

The slide features several decorative teal lines. A thick line curves from the left edge towards the top right. Two thinner lines also curve from the left towards the top right, positioned below the thick line. A horizontal teal line runs across the bottom of the slide, with a short vertical line segment extending upwards from it on the left side.

2017

Aspen Institute
Dialogue Series
on Water Data

Vision

The Internet of Water envisions a nation engaged in equitable and resilient water management and stewardship enabled by shared and integrated water data and information.

Project Background




2017

Aspen Institute
Dialogue Series
on Water Data

2018

TWDB hosts a workshop
for the Texas water
data community



Findable

Metadata and data should be findable for both humans and computers

F

A

Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

I

R

Accessible

Once found, users need to know how the data can be accessed

Reusable

The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

Project Background

A timeline diagram showing the project's background from 2017 to 2020. The title 'Project Background' is at the top left. A thick teal horizontal line at the bottom represents the timeline. Four vertical teal lines mark the years 2017, 2018, 2019, and 2020. Above each year is a text box describing an event. Three curved teal lines rise from the timeline, starting from the 2017 mark and curving upwards towards the right, ending near the 2019 mark.

2017

Aspen Institute
Dialogue Series
on Water Data

2018

TWDB hosts a workshop
for the Texas water
data community

2019

Communication,
outreach, and prep
for the project begins











2020


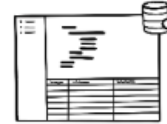










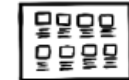

Project begins with a
focus on Human
Centered Design (HCD)

Human Centered Design (HCD)

- People-centered
- Understand and solve the right problems
- Everything is a system
- Small and simple interventions



 WATER DATA	 SEARCH BAR	 REAL-TIME	 RELATED
 RECOMMENDATIONS	 FILTERS	 UPDATES	 EXPLORE
 CATALOG	 LINKS	 QUERY	 DOWNLOAD
 TOOLS	 LOCATION		

 REAL-TIME	 QUERY	 TOOLS
 LOCATION	 EXPLORE	 RECOMMENDATIONS
 FILTERS	 UPDATES	 SEARCH BAR
 DOWNLOAD	 RELATED	 WATER DATA
		 CATALOG
		 LINKS

“I have to talk to a guy...ask him for a [data] dump. It's not even on their website.”

-Mark (996)

“It's not just 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)

“If we didn't have those contacts we wouldn't have necessarily known that that information was available.”

-Gene (1936)

“I have to talk to a guy...ask him for a [data] dump. It's not even on their website.”

-Mark (996)

“It's not just 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)

“If we didn't have those contacts we wouldn't have necessarily known that that information was available.”

-Gene (1936)

**Data access is dependent
on who you know**

“The other potential impediment to sharing is, it’s really important in my business to thoroughly document how that information has been collected.”

-Gene (1925)

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

-Sheila (1557)

“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)

“The other potential impediment to sharing is, it’s really important in my business to thoroughly document how that information has been collected.”

-Gene (1925)

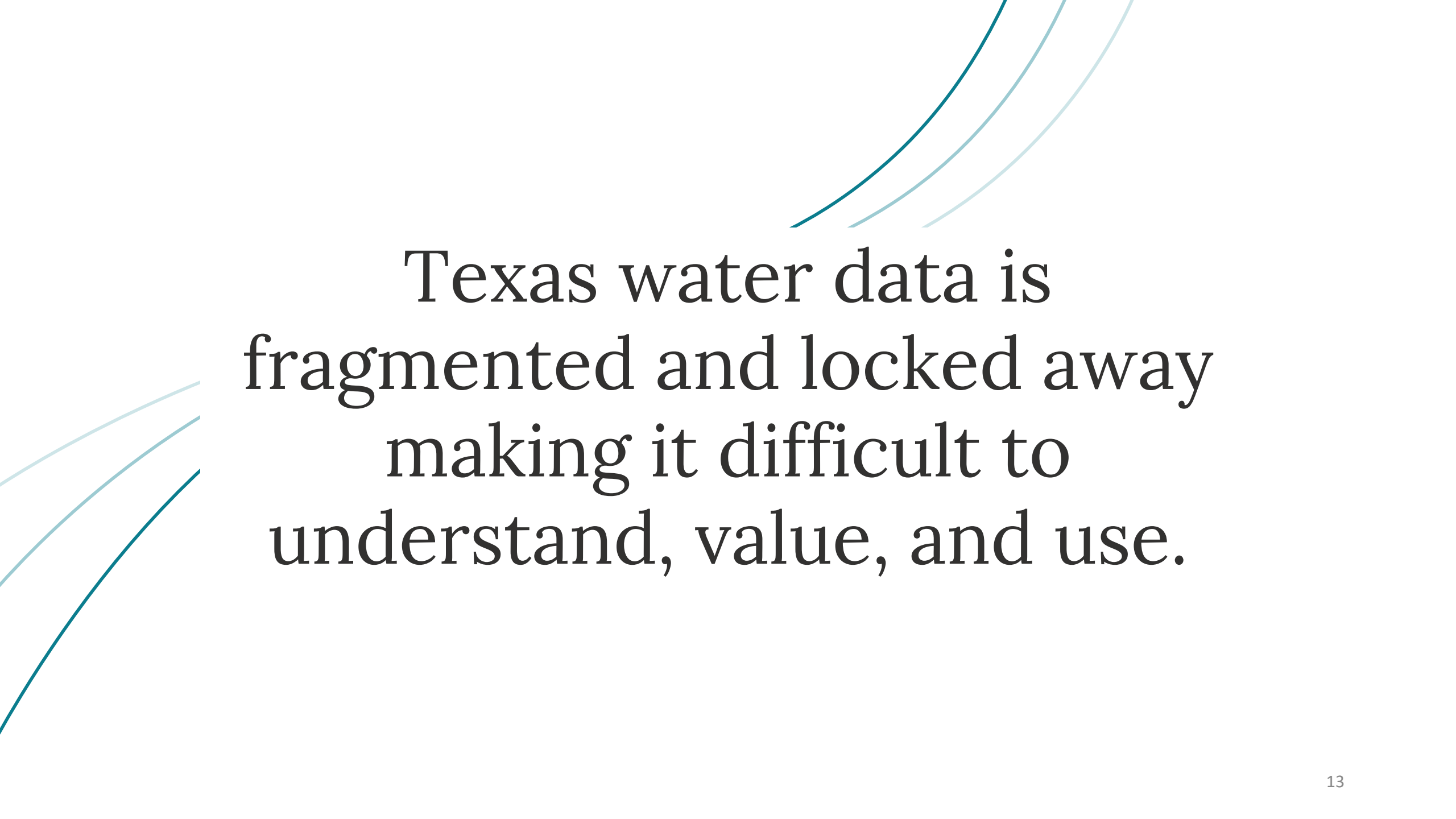
“We’re trying to get the QA/QC completed and we’re probably 85% complete. At the point that we’re pretty confident the data is mostly complete, we’ll start serving that data out.”

-Sheila (1557)

“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)

**Successful data sharing
depends on trust between
producers and users.**



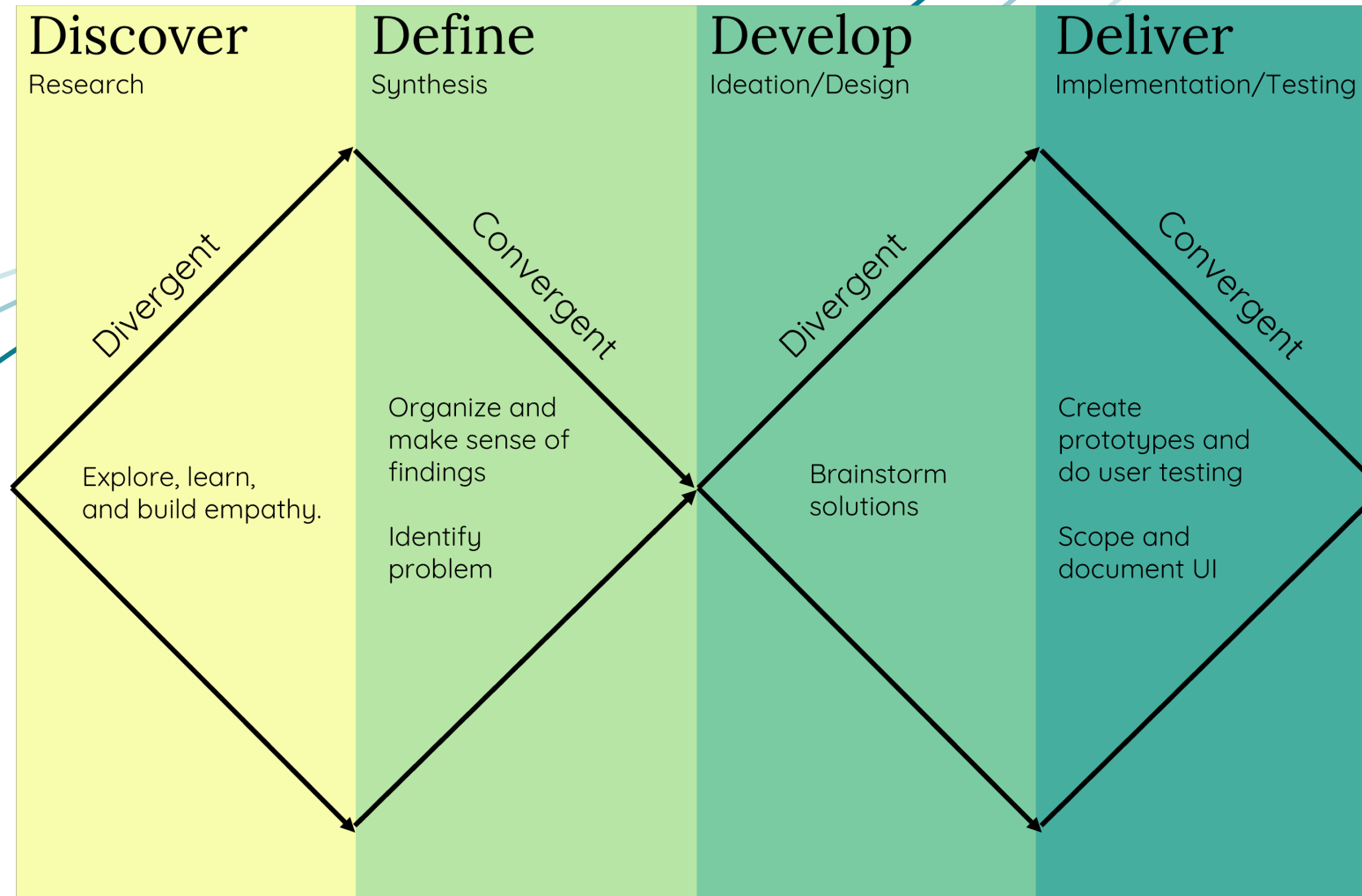
Texas water data is
fragmented and locked away
making it difficult to
understand, value, and use.

A water data hub should...



1. Provide a central location for data that reflects the entire 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 data interoperability efforts through standards and curated datasets.

Our Process



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



Data Inventory



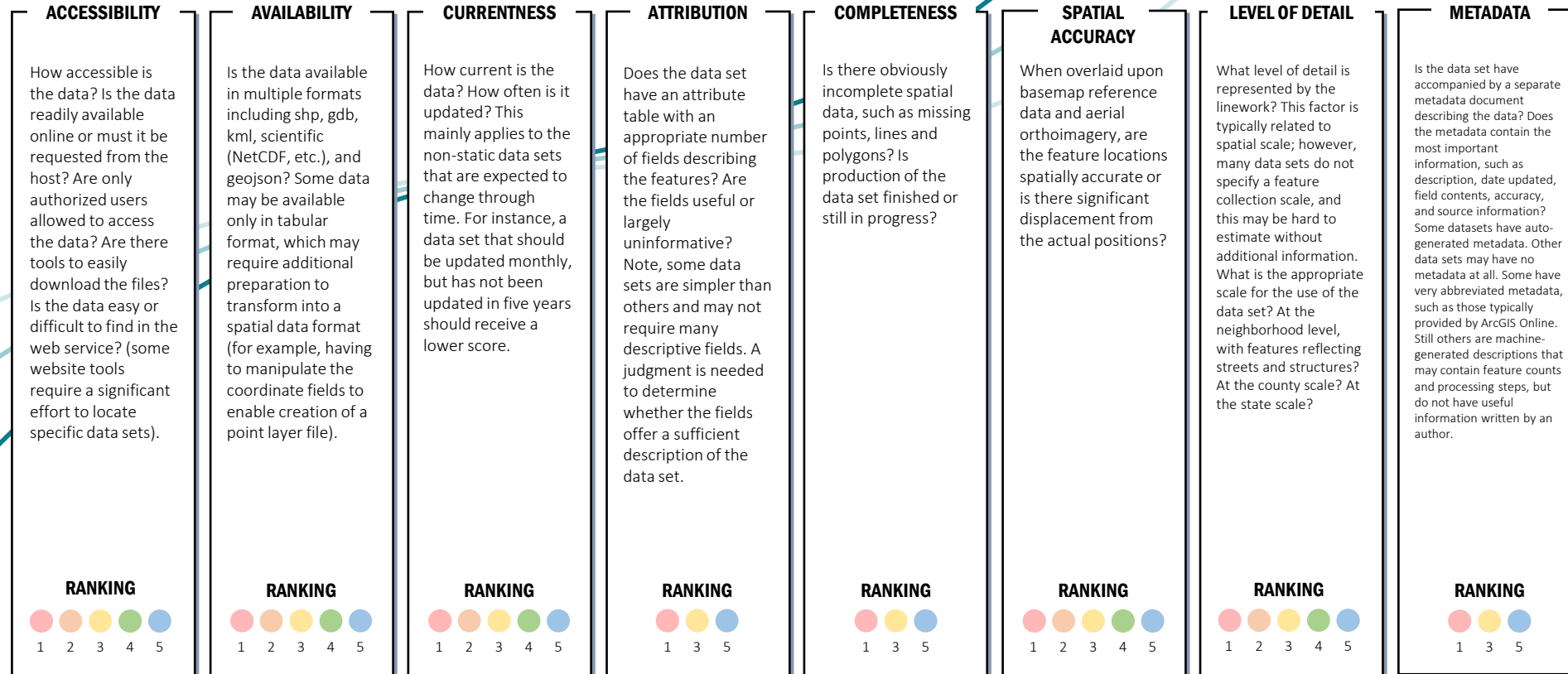
Start with what you know



Look to others for previous work

Data Assessment

Source: [CDR Data Management Plan](#)

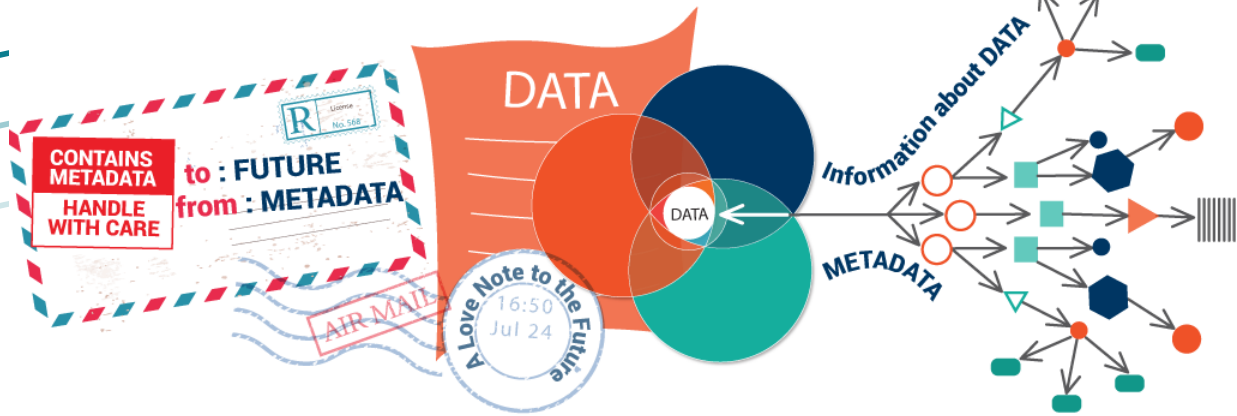


Stakeholder Input

Use the voting feature to pick 3 data categories that you would use. Think about why they're important for our next discussion.

<p>Groundwater Availability Models ⁹</p> <p>Groundwater datasets</p> <p>TWDB</p>	<p>Adopted Desired Future Conditions ⁴</p> <p>Groundwater datasets</p> <p>TWDB</p>	<p>Submitted Drillers Database ⁴</p> <p>Groundwater datasets</p> <p>TWDB</p>	<p>Aquifer Storage and Recovery Suitability ⁴</p> <p>Groundwater datasets</p> <p>TWDB</p>
<p>Pumping Data ⁴</p> <p>Groundwater datasets (various sources)</p> <p>TWDB, GWCDs</p>	<p>Springs Discharge ⁶</p> <p>Groundwater datasets</p> <p>TWDB, USGS</p>	<p>Daily Water Levels ⁵</p> <p>Groundwater datasets (feet below surface)</p> <p>TWDB, USGS</p>	<p>Water Use Survey ³</p> <p>Surface water datasets</p> <p>TWDB</p>
<p>Surface Water Quality ⁵</p> <p>Surface water datasets</p> <p>TCEQ</p>	<p>Water Rights</p> <p>Surface water datasets</p> <p>TCEQ</p>	<p>Stream Flow Conditions ⁵</p> <p>Surface water datasets</p> <p>USGS</p>	<p>Surface Water Quality - Quantity Data ⁴</p> <p>Surface water datasets (field to watershed data scales collected over 25 years)</p> <p>TIAER</p>
<p>Water Availability Model ⁷</p> <p>Surface water datasets</p> <p>TCEQ</p>	<p>Evapotranspiration ⁴</p> <p>Weather and environmental datasets (will launch this year)</p> <p>OpenET</p>	<p>Modeled Soil Moisture</p> <p>Weather and environmental datasets (using remote sensed product to create a modeled layer for the state)</p> <p>NASA</p>	<p>TexasMesoNet ⁶</p> <p>Weather and environmental datasets (precipitation, temperature, humidity, wind, soil moisture)</p> <p>TWDB</p>
<p>Groundwater Surface Water Interactions in 22 River Basins ²</p> <p>Academic literature from 1999</p> <p>TCEQ</p>	<p>Geodatabase for the Brazos River Alluvium Aquifer</p> <p>Academic literature from 2007</p> <p>USGS</p>	<p>Quantification of Surface Water/Groundwater Interaction In the Colorado River Basin ⁴</p> <p>Academic literature from 2017</p> <p>TWDB</p>	<p>Results of Streamflow Gain-Loss Studies in Texas ³</p> <p>Academic literature from 2002</p> <p>USGS</p>

Metadata Schema



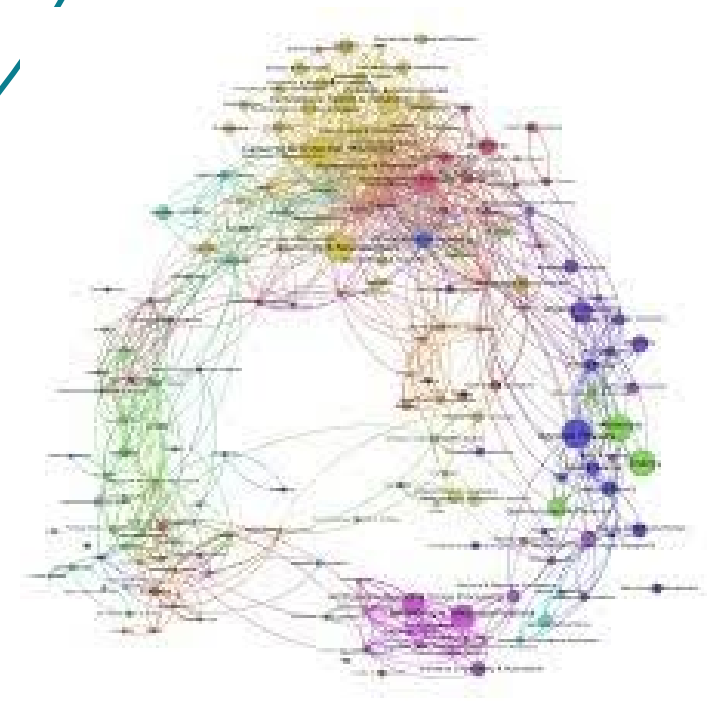
- Intuitive and semi-automated upload flow
 - What information can be pulled from the dataset and then verified by the user?
- Coordinate with others on transfer of data between systems
- Strong metadata schema upfront is critical to FAIR data and future interoperability work

Data Ontology

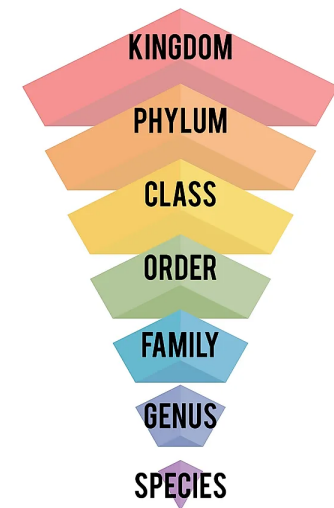
"In computer science and information science, an ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many, or all domains of discourse."

But what does that REALLY mean for the hub?

We want to know how datasets in similar categories are related, through field names, locations or other variables. This will lead to the development of data dictionaries, and common naming conventions. Additionally, it is the first step to making data ***INTEROPERABLE***.



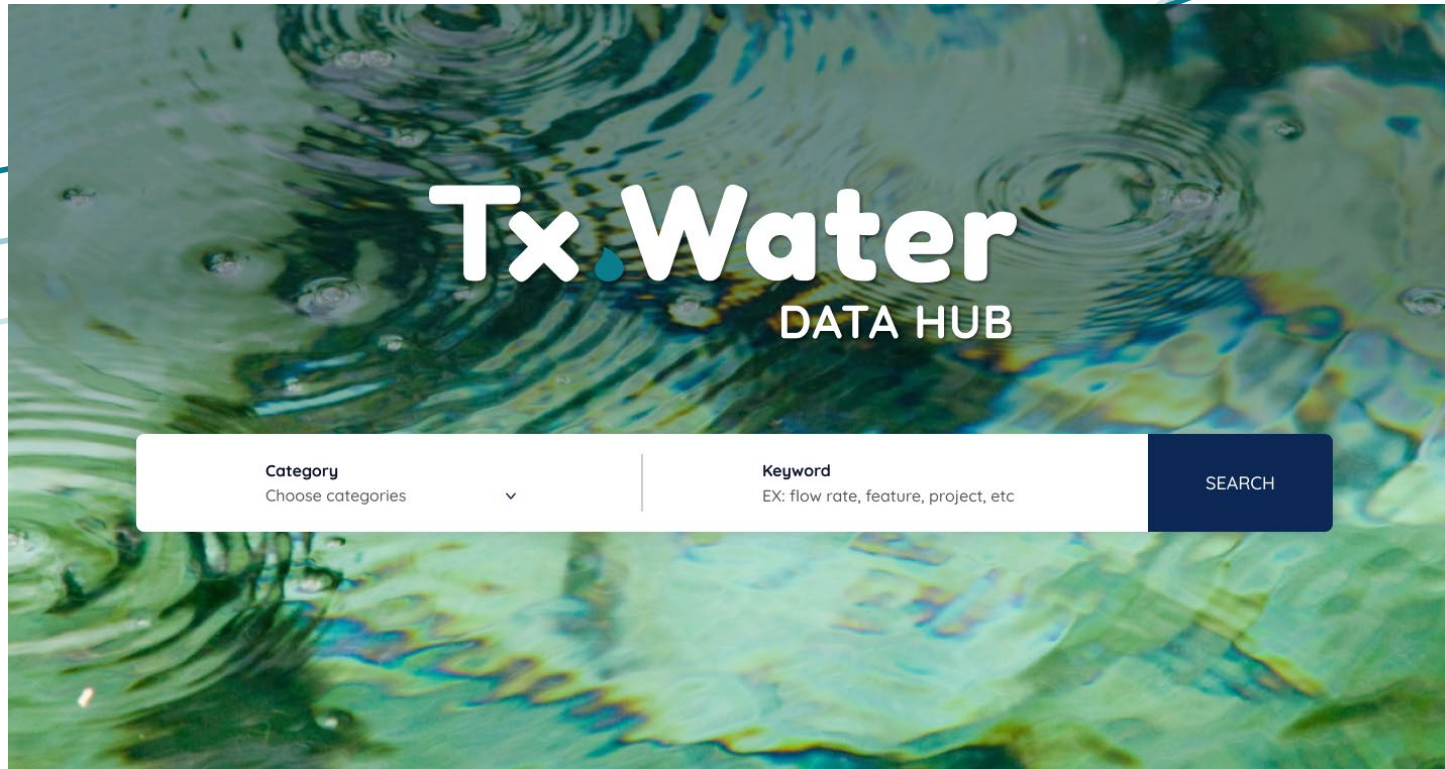
HIERARCHY OF BIOLOGICAL CLASSIFICATION



Data Governance



Thank you!



Keep an eye out for the beta version of the Hub later this year!

What to discuss more?

taylor.christian@twdb.texas.gov

laura.sepulveda@twdb.texas.gov