Building an Open (💧) Data Hub

Laura Sepulveda and Taylor Christian

Project Background

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.

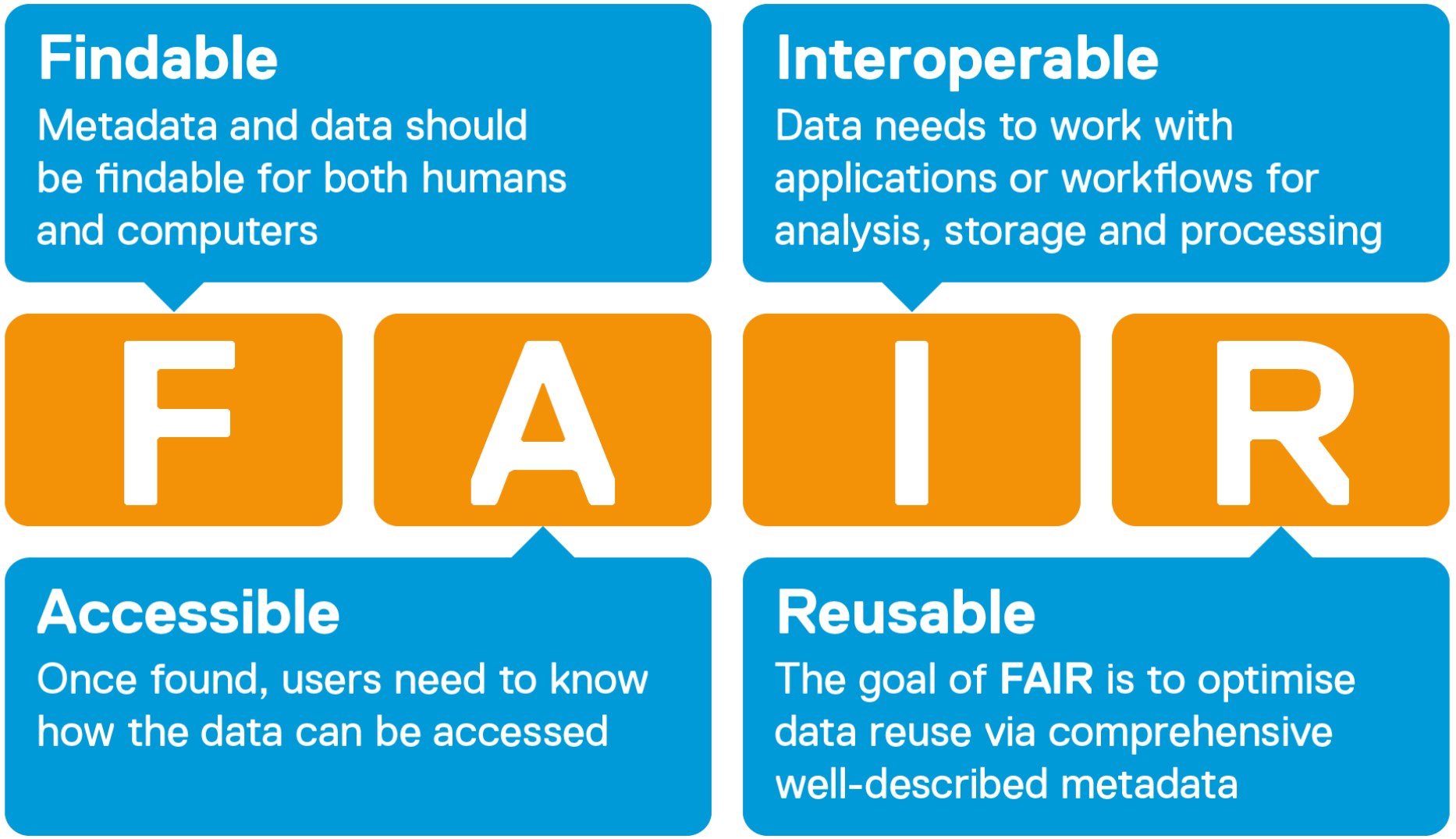
Project Background

2017

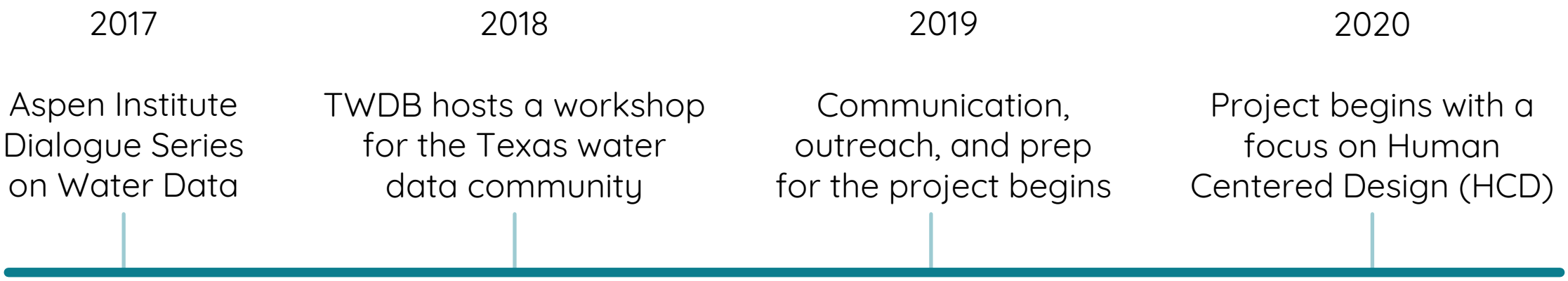
Aspen Institute Dialogue Series on Water Data

2018

TWDB hosts a workshop for the Texas water data community

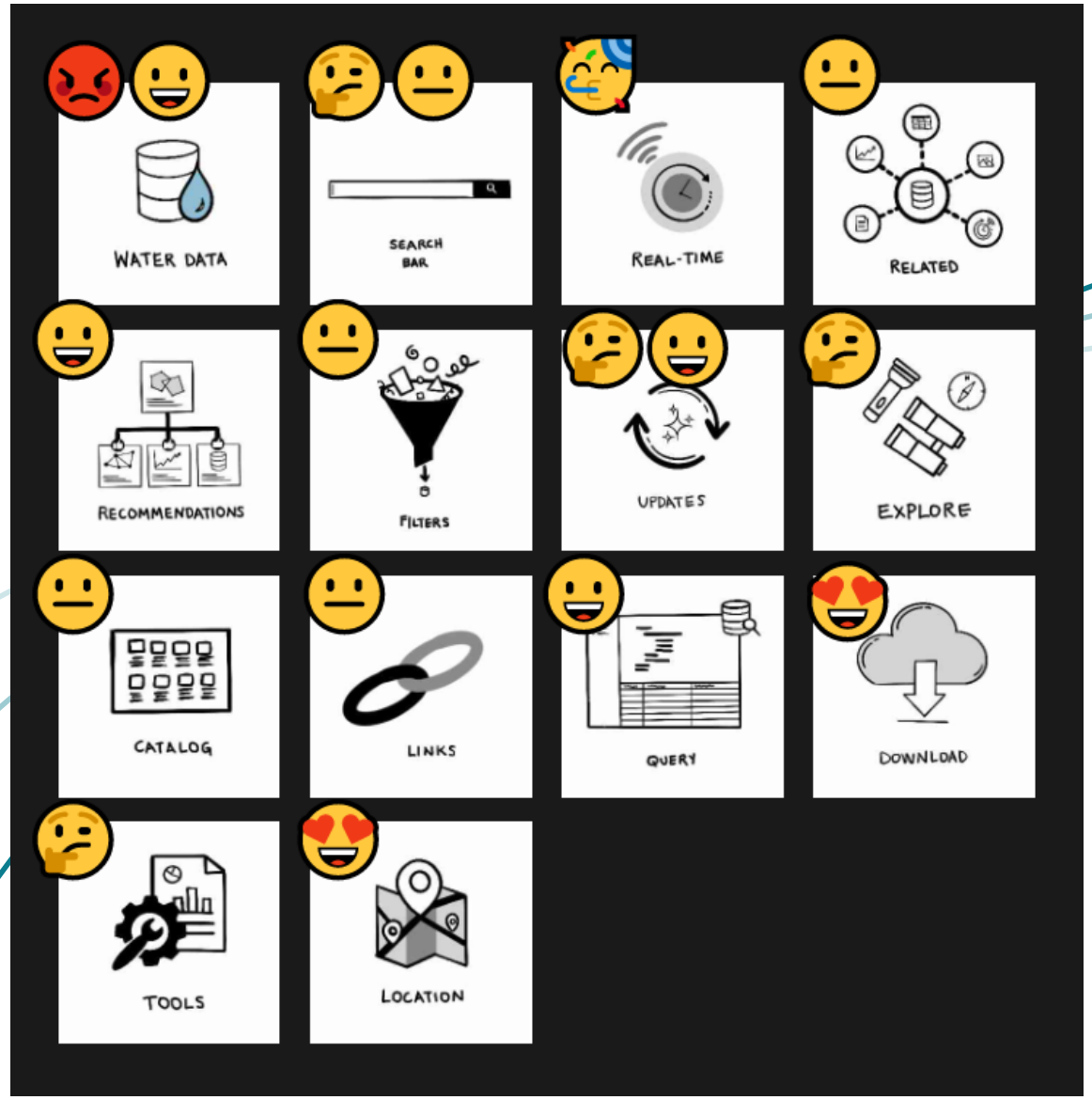

Findable
Metadata and data should be findable for both humans and computers
Interoperable
Data needs to work with applications or workflows for analysis, storage and processing
F
A
I
R
Accessible
Once found, users need to know how the data can be accessed
Reusable
The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

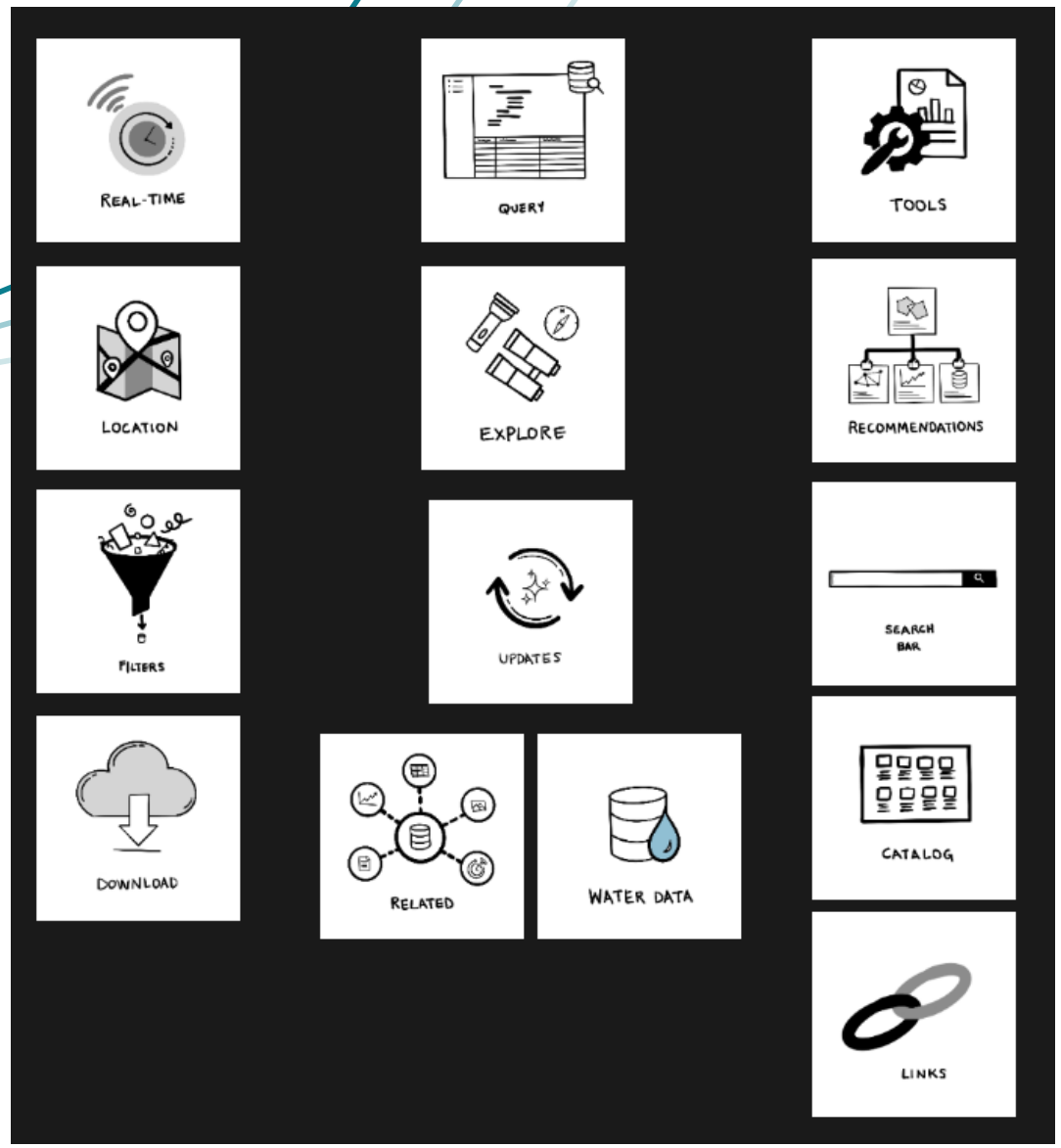
Project Background



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)

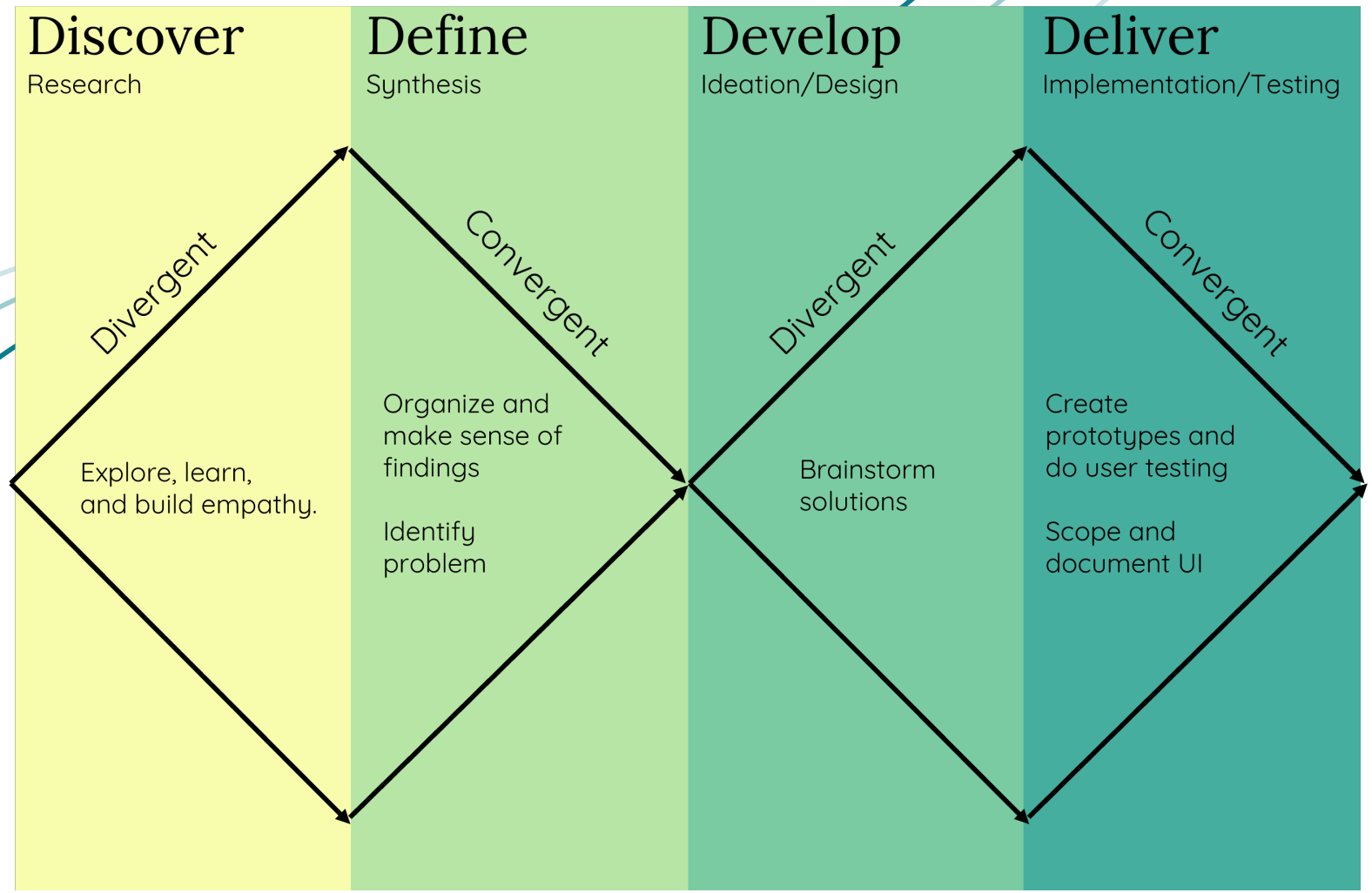
Successful data sharing depends on trust between producers and users.

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


Discover
Research
Define
Synthesis
Develop
Ideation/Design
Deliver
Implementation/Testing
Divergent
Convergent
Divergent
Convergent
Explore, learn, and build empathy.
Organize and make sense of findings
Identify problem
Brainstorm solutions
Create prototypes and do user testing
Scope and document UI

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


GET POINTED IN THE RIGHT DIRECTION
000-099 GENERALITIES
100-199 PHILOSOPHY AND PSYCHOLOGY
200-299 RELIGION AND MYTHOLOGY
300-399 SOCIAL SCIENCES
400-499 LANGUAGE
500-599 NATURAL SCIENCES
600-699 TECHNOLOGY
700-799 ARTS
800-899 LITERATURE
900-999 GEOGRAPHY AND HISTORY

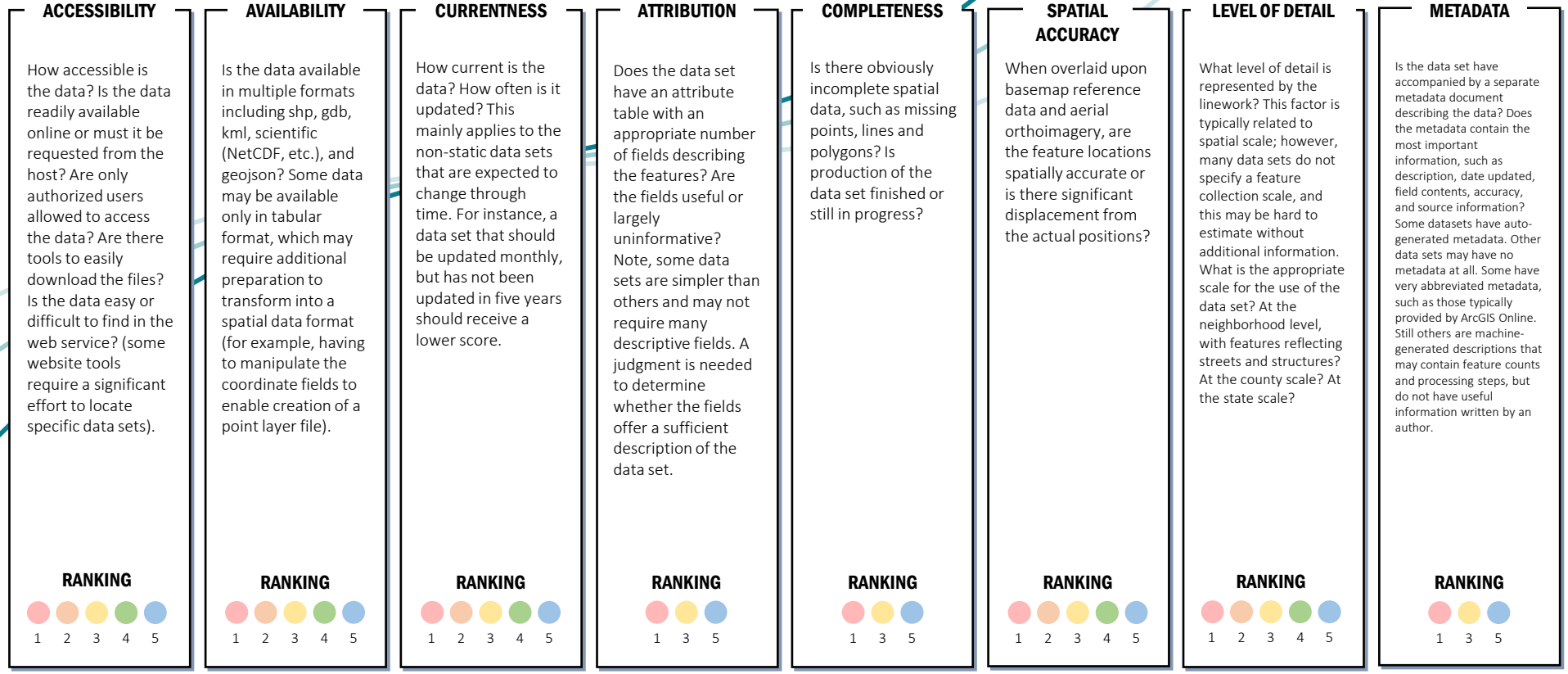
Data Inventory

Start with what you know

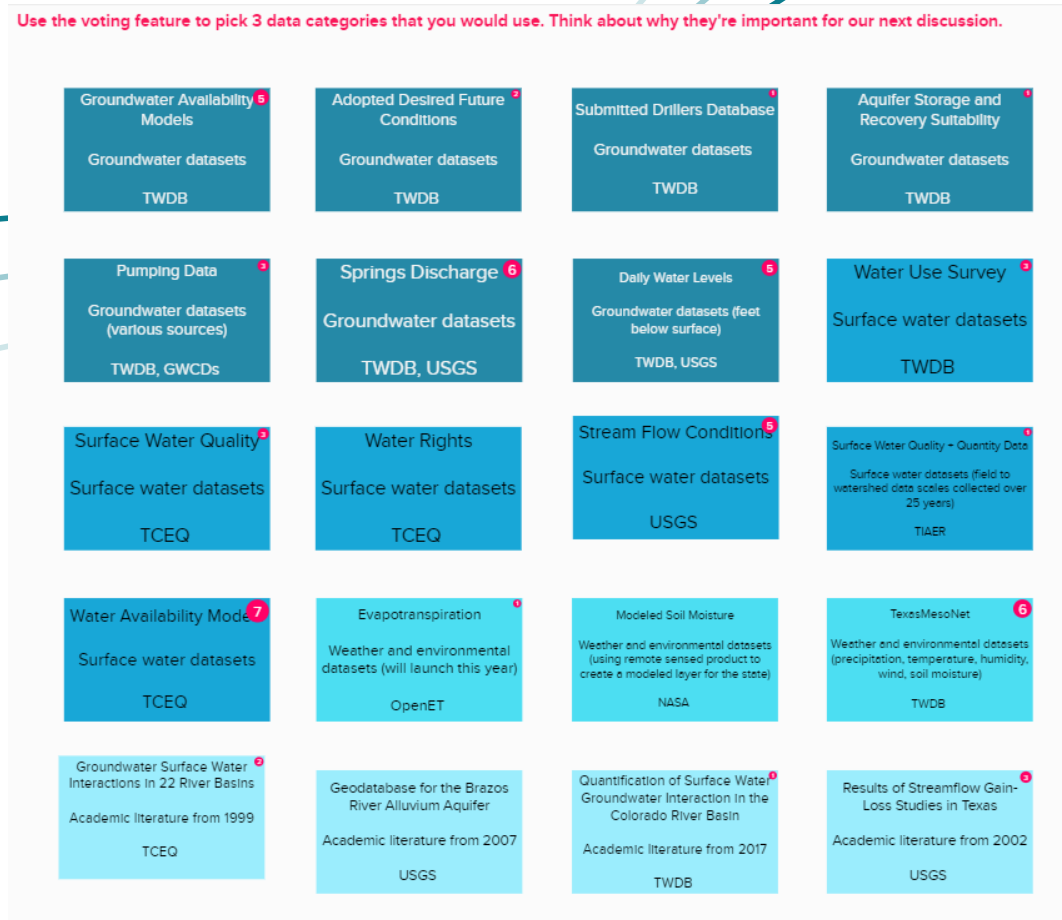
Look to others for previous work

Data Assessment

Source: CDR Data Management Plan



Stakeholder Input

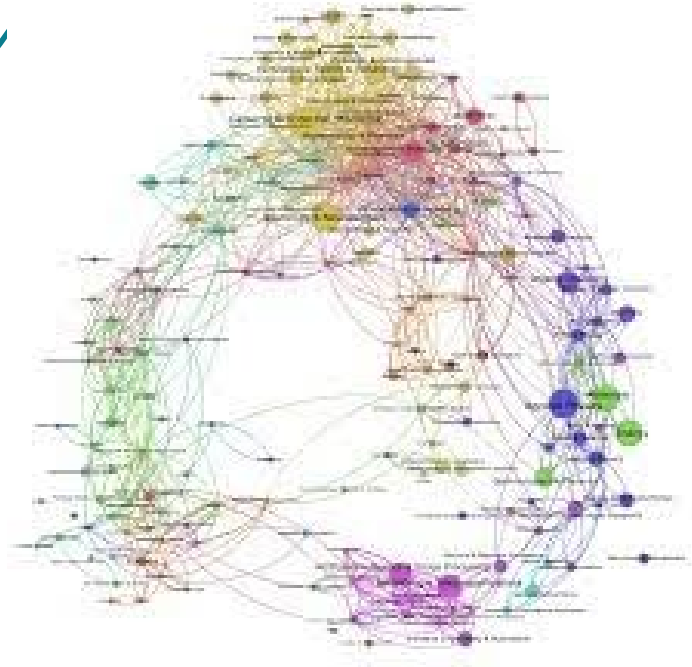


Metadata Schema

- 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

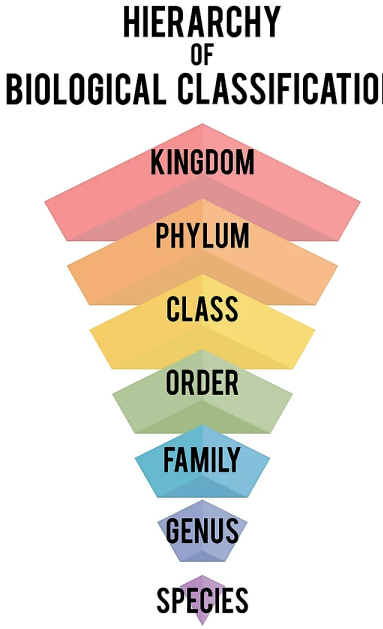
Data Ontology

"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."



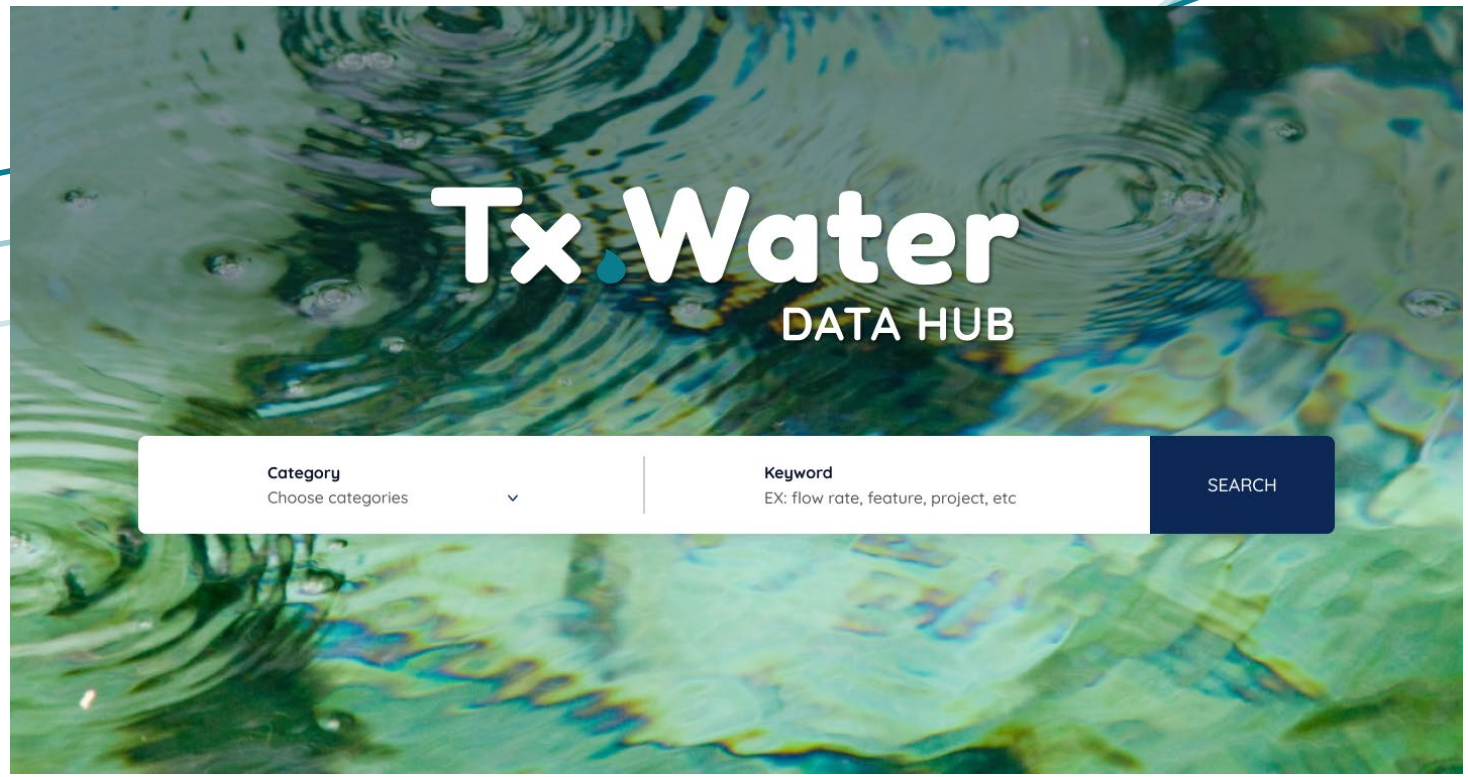
But what does that REALLY mean for the hub?

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*.



Data Governance

Thank you!



Keep and 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