

# WHICH CAME FIRST: THE DATA OR THE DECISION?

HOW AI/ML HELPS THE USGS DELIVER DECISION-AGNOSTIC WATER  
DATA THAT INFORMS EVERY LINK IN THE DECISION QUALITY CHAIN

KATRINA ALGER<sup>1</sup>, JACOB ZWART<sup>1</sup>, JARED SMITH<sup>1</sup>, GERICKE COOK<sup>1</sup>, JULIEN  
MARTIN<sup>1</sup>, AND LAURA BRANDT<sup>2</sup>

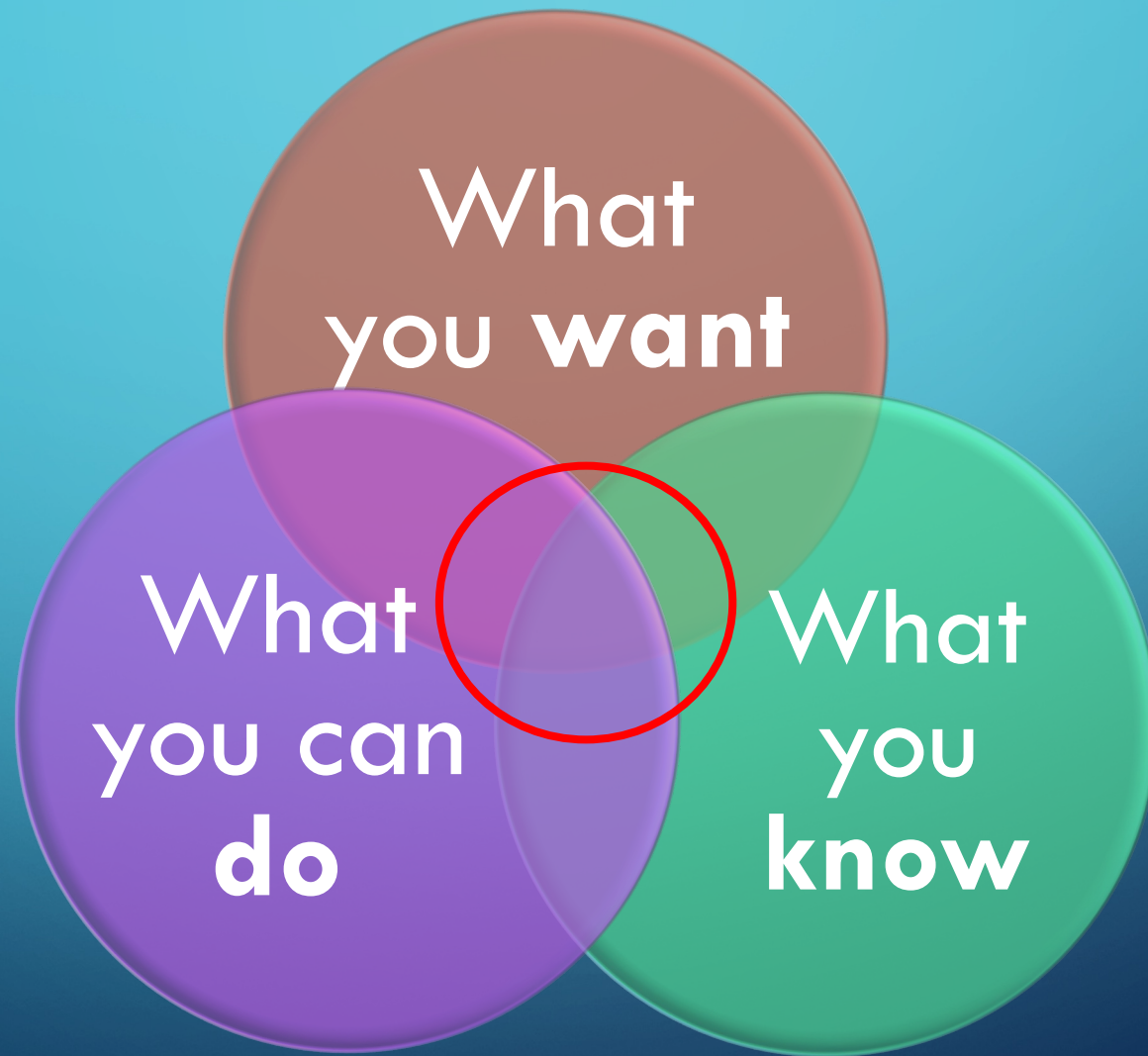
<sup>1</sup>U.S. GEOLOGICAL SURVEY

<sup>2</sup>U.S. FISH AND WILDLIFE SERVICE

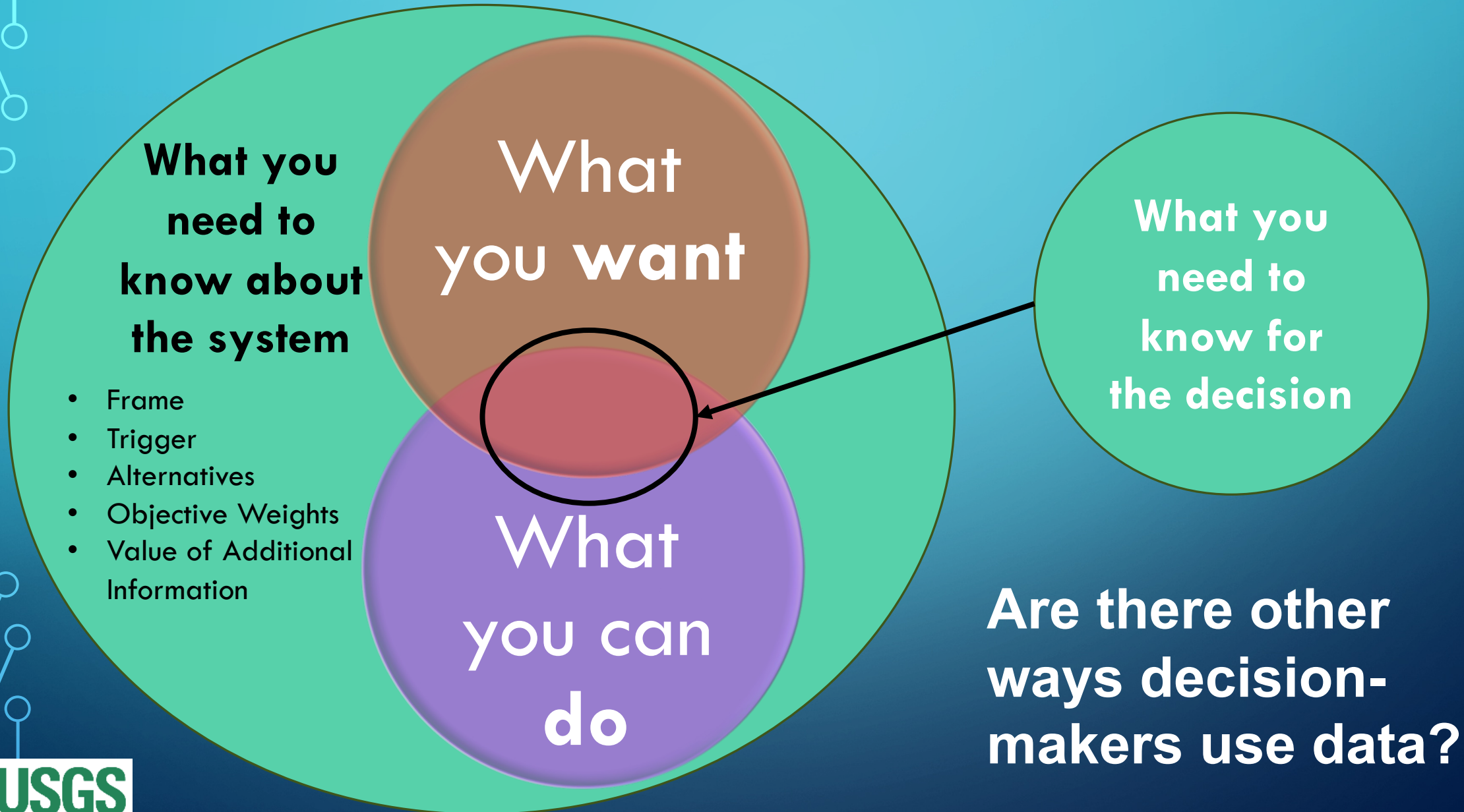
2024 SOCIETY FOR DECISION PROFESSIONALS  
*EFFECTIVE DECISION MAKING IN A DYNAMIC WORLD*  
ARLINGTON, VA

THIS INFORMATION IS PRELIMINARY AND IS SUBJECT TO REVISION. IT IS BEING PROVIDED TO MEET THE NEED FOR TIMELY BEST SINCE. THE INFORMATION IS PROVIDED ON THE CONDITION THAT NEITHER THE U.S. GEOLOGICAL SURVEY NOR THE U.S. GOVERNMENT SHALL BE HELD LIABLE FOR ANY DAMAGES RESULTING FROM THE AUTHORIZED OR UNAUTHORIZED USE OF THE INFORMATION.

# WHERE IS DATA NORMALLY SLOTTED IN THE DECISION-MAKING PROCESS?



# DATA FOR DECISIONS



# THE CHALLENGE



Reiner, Rob. (1984) *This is Spinal Tap*

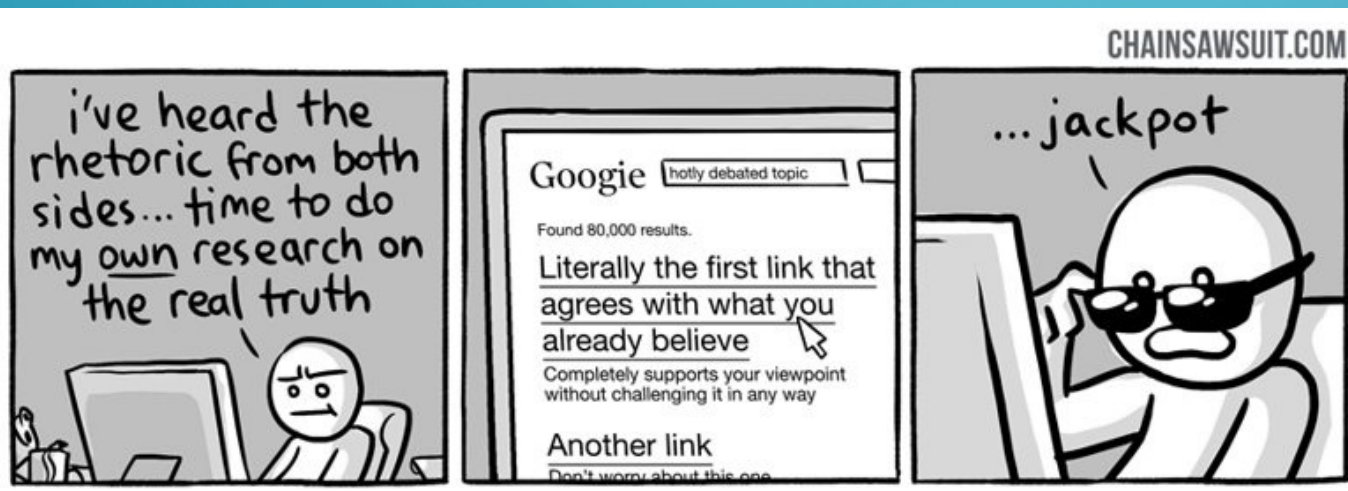


There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know.

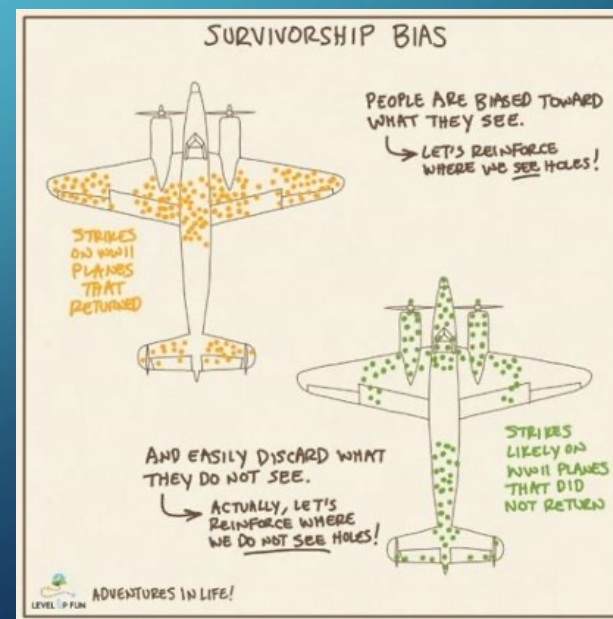
— Donald Rumsfeld —

AZ QUOTES

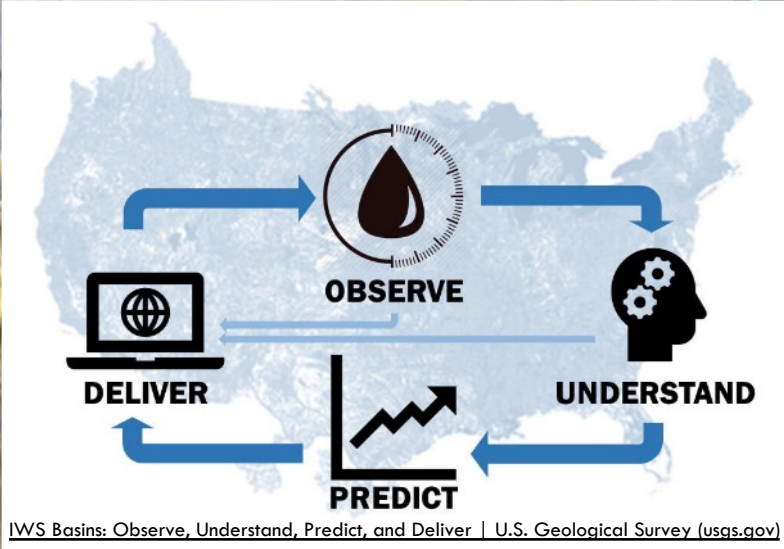
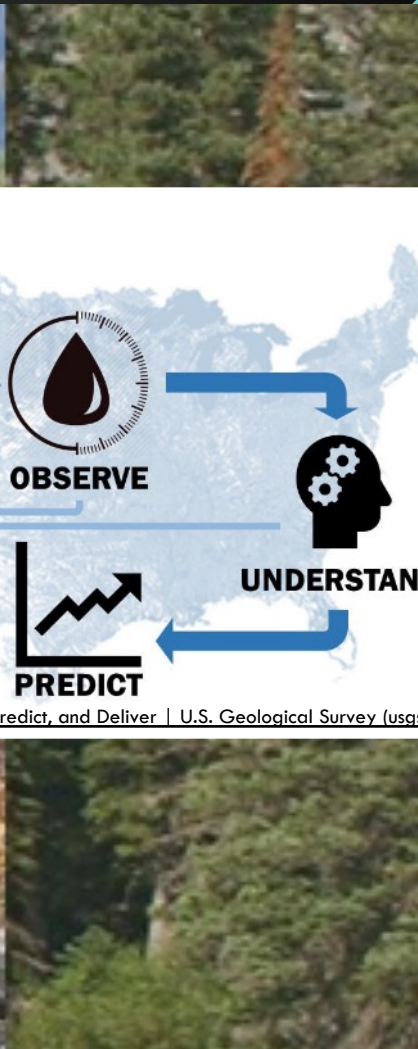
<https://www.azquotes.com/quote/254214>



Kris Straub (2014), [Chainsawsuit](http://Chainsawsuit.com)



<https://www.levelupfun.com/articles/survivorship-bias>



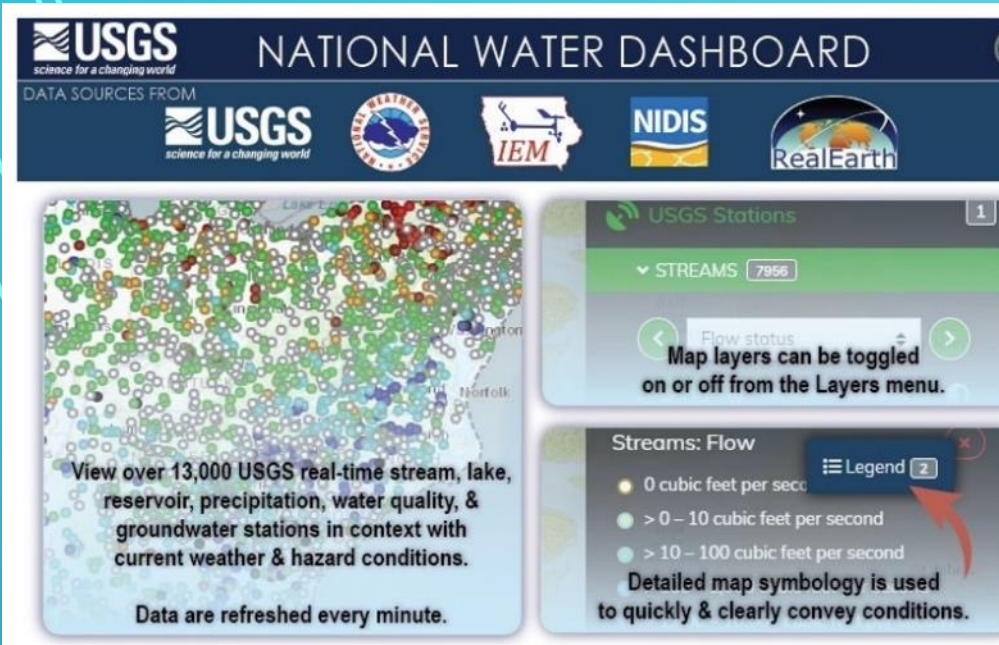
UNREVIEWED DRAFT - DO NOT DISSEMINATE

**U.S. Geological Survey Water Science Strategy—  
Observing, Understanding, Predicting, and  
Delivering Water Science to the Nation**

Circular 1383-G

U.S. Department of the Interior  
U.S. Geological Survey

PRELIMINARY INFORMATION – SUBJECT TO REVISION. NOT FOR CITATION OR DISTRIBUTION

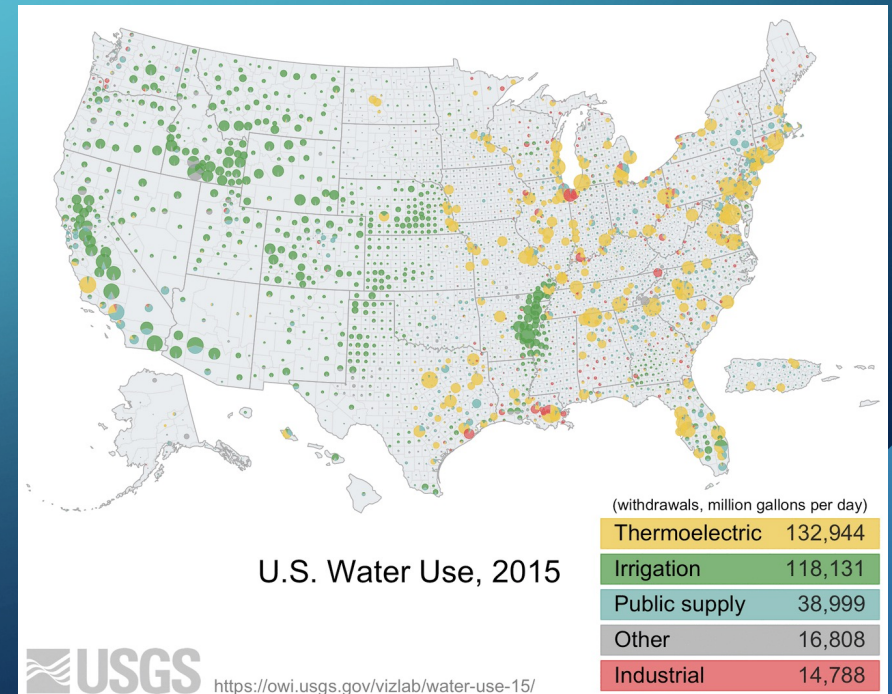


[USGS | National Water Dashboard](https://labs.waterdata.usgs.gov/)



ALL PHOTOS PUBLIC DOMAIN UNLESS OTHERWISE NOTED

<https://labs.waterdata.usgs.gov/visualizations/water-use-15>

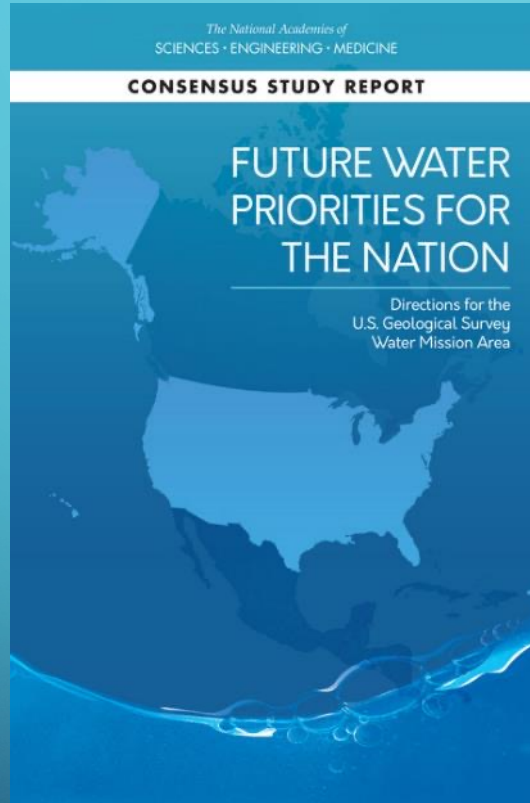




<https://bobdylancenter.com/about/biography/>

YOU BETTER  
START SWIMMIN'  
OR YOU'LL SINK  
**LIKE A STONE**  
FOR THE TIMES  
THEY ARE A-CHANGIN'  
BOB DYLAN

# 2009 SECURE Water Act PL 111-11 Section 9508



To evaluate:

- The *status* of water resources in the United States
- The *quantity* of water available
- The *quality* of water available
- Long-term *trends* in water availability

And ultimately, to *forecast* water availability for future economic, energy, and environmental uses



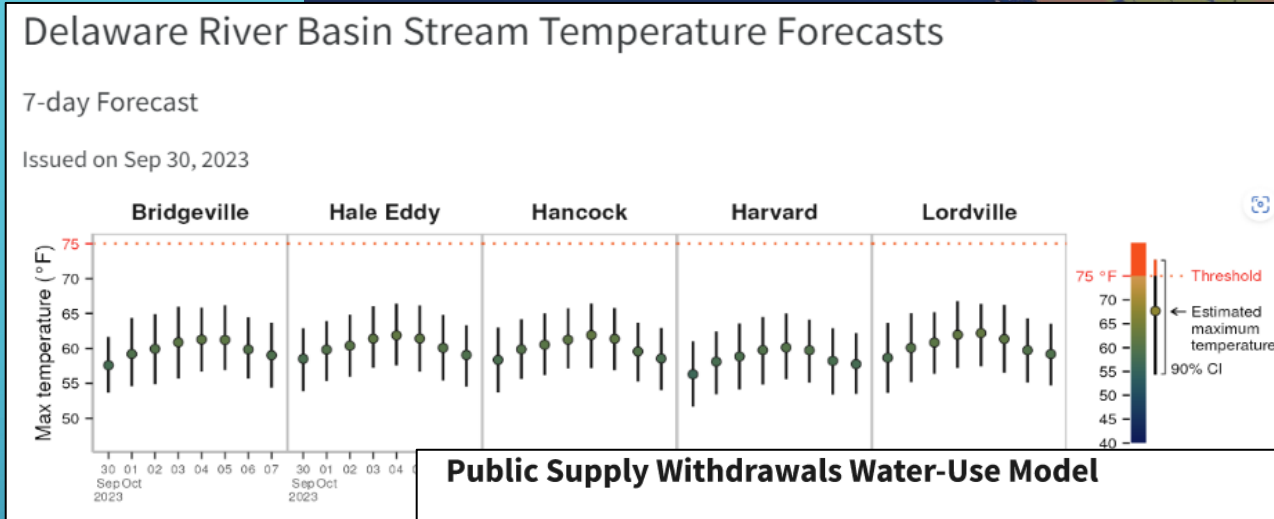
U.S. Geological Survey

## National Modeled Water Atlas

The National Modeled Water Atlas (NMWA) is a centralized website and near real-time data delivery system for water availability, use, quality, and aquatic ecosystem conditions derived from U.S. Geological Survey (USGS) scientific models.

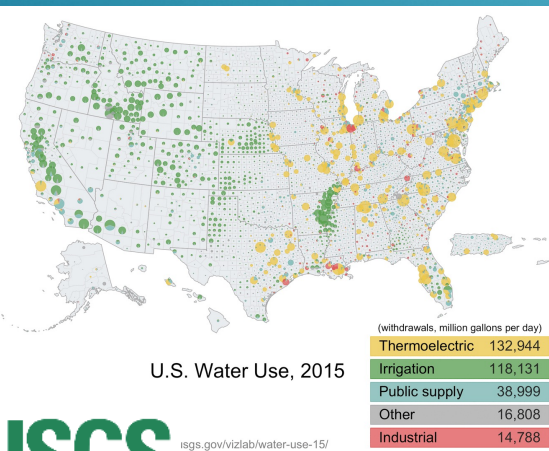
**PROVISIONAL – NOT FOR DISTRIBUTION OR CITATION**

[Learn More](#)



[Using Artificial Intelligence to Help Protect Trout and Drinking Water Supplies | U.S. Geological Survey \(usgs.gov\)](#)

[Stream Temperature Forecasts for Sites in the Delaware River Basin - Delaware River Basin Stream Temperature Forecasts \(usgs.gov\)](#)



### Public Supply Withdrawals Water-Use Model

**Data Access**  
Data from the most current model version can be accessed through the Data Download and Web Service URL Generation Tool via the button below.

[Subset & Download](#)

**Description**

**Summary**  
This dataset contains the amount of water withdrawn for public supply each month over the period from 2000-2020 for all 12-digit USGS Watershed Eoun (units of Subwatersheds (HUC12) in the conterminous United States (CONUS). The machine learning model that generated this dataset combines historical information and an array of different factors that drive public supply withdrawal, including hydroclimatic, demographic, economic, geographic, and social variables. Ultimate concepts will include public supply system losses, cellular, withdrawal, consumption.

**Technical Description**  
A national data-driven public supply water use model was developed using a complex array of driving factors that include hydroclimatic, demographic, economic, geographic, and social factors. The public supply water use model predicts annual and monthly public supply withdrawals. Historic water use data was compiled for the period 2000-2020 and combined with explanatory factors at different temporal scales (annual, monthly, and daily) to train and validate a national-scale machine learning model. The water use model was developed using the XGBOOST machine learning library (Chen and Guestrin, 2016) and was formulated to predict monthly and annual per capita water use for all public utility service areas in the conterminous US that contain populations greater than

**Overview**

**Sector**  
WATER USE

**Release Information**  
**Release Status**  
Formally published or released

**Update Frequency**  
Annual  
Last updated: YYYY-MM-DD

**Originator**  
U.S. Geological Survey

**License**  
INFO TO BE ADDED

**Full Data Access**  
[Open File Browser](#)

**Temporal Information**

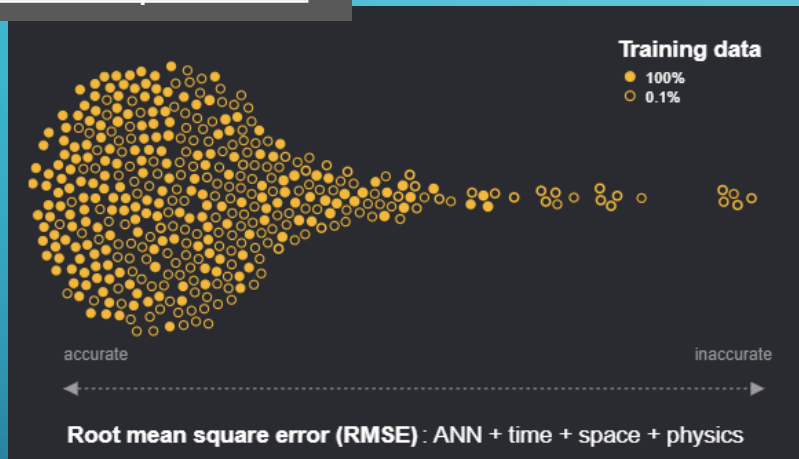
**PROVISIONAL – NOT FOR DISTRIBUTION OR CITATION**



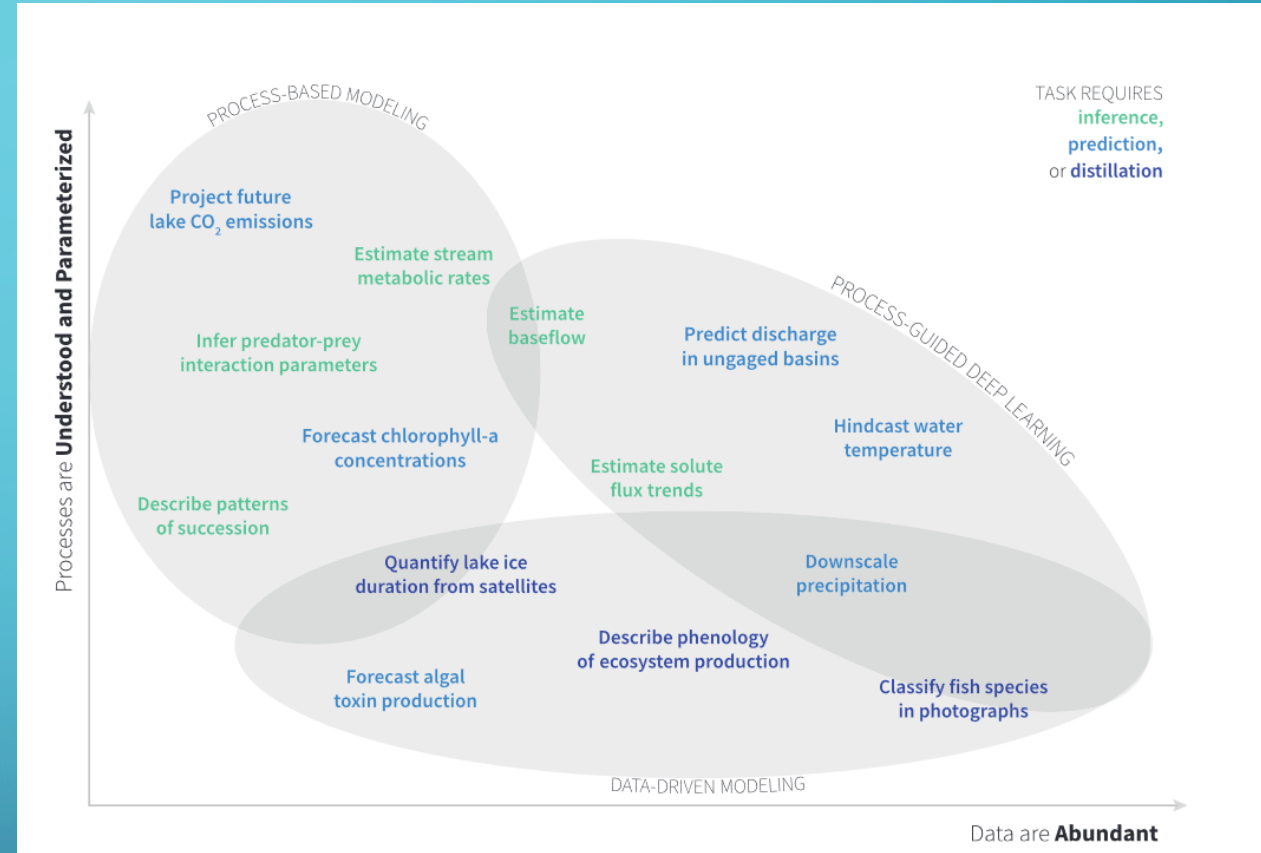


# AI/ML COMPLIMENTS OTHER MODELING FRAMEWORKS

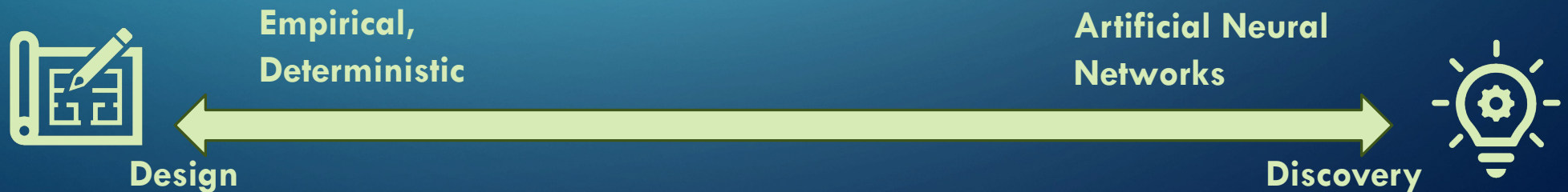
<https://labs.waterdata.usgs.gov/visualizations/temperature-prediction>



Figures by H. Corson-Dosch; data from Jia et al. 2021



From [Appling et al. 2022](#)



From [Appling et al. 2022](#)



istockphoto.com

# 3 KEY QUALITIES OF DECISION-AGNOSTIC DATA



**1. RELIABLE**



**2. USEFUL**

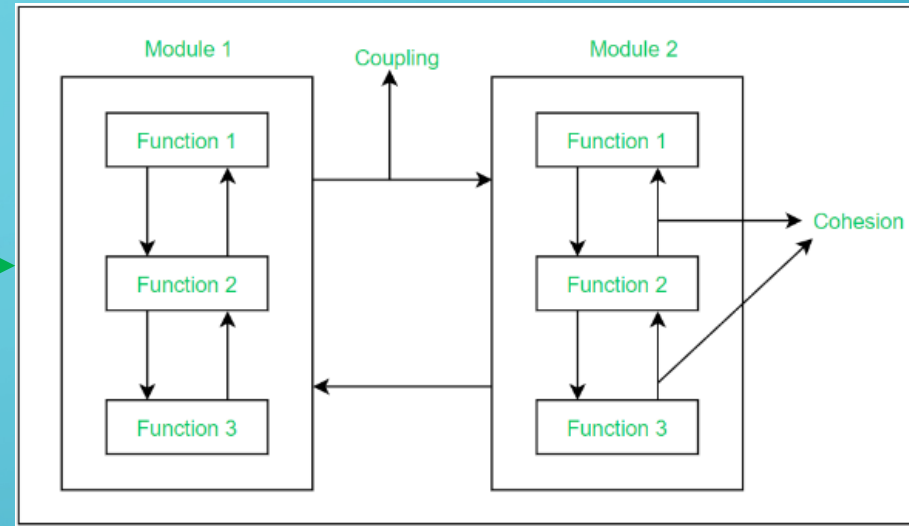
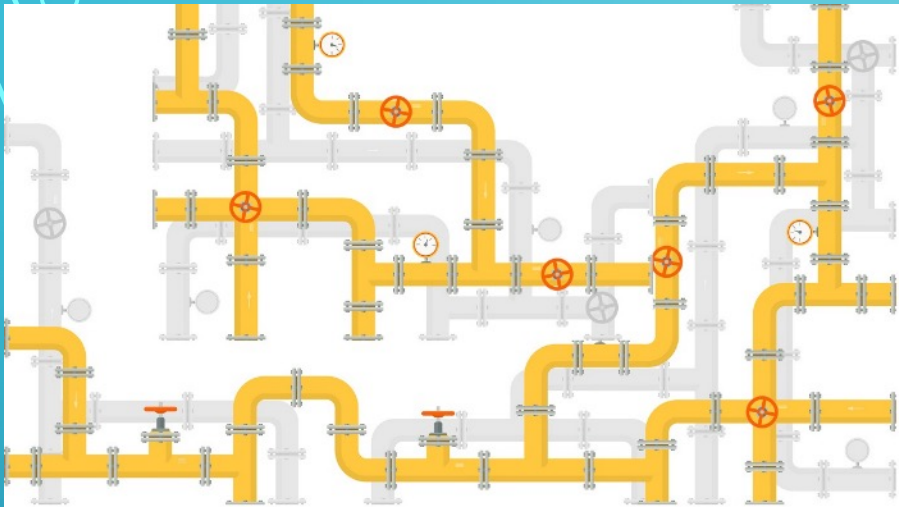


**3. ENHANCE  
UNDERSTANDING**



photo.co

# ENSURING BEST PRACTICES FROM DEVELOPMENT TO DELIVERY

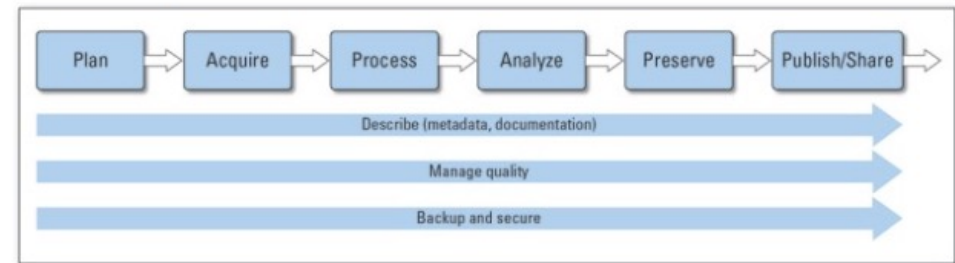


USGS Celebrates the Year of Open Science

15<sup>th</sup> Year of Open Science

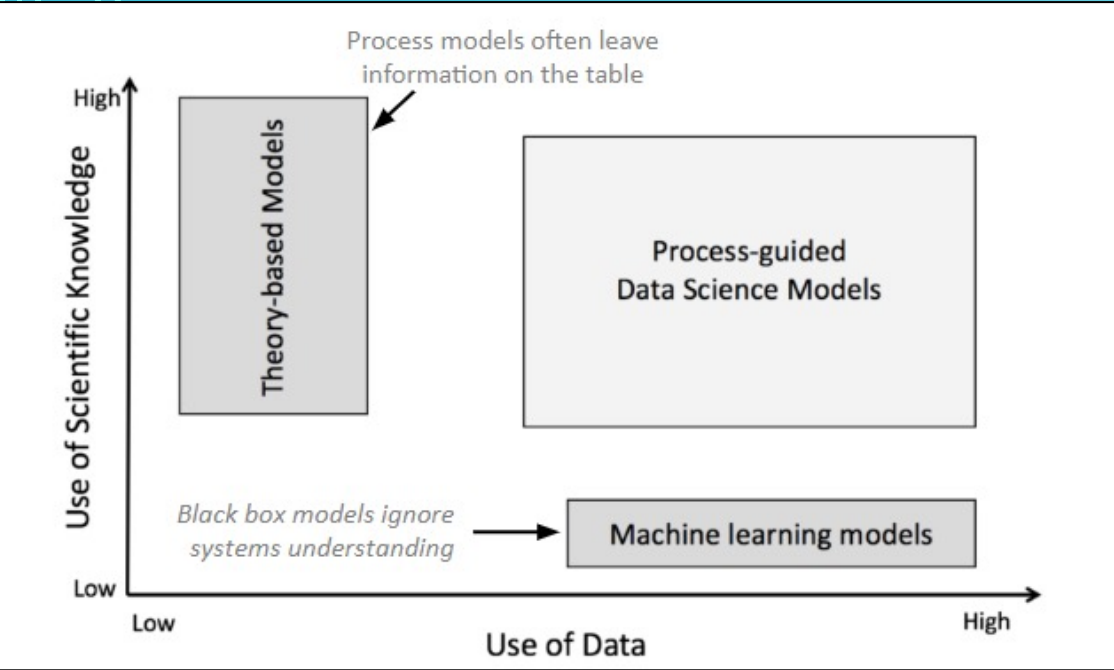


Machine Learning Best Practices Framework

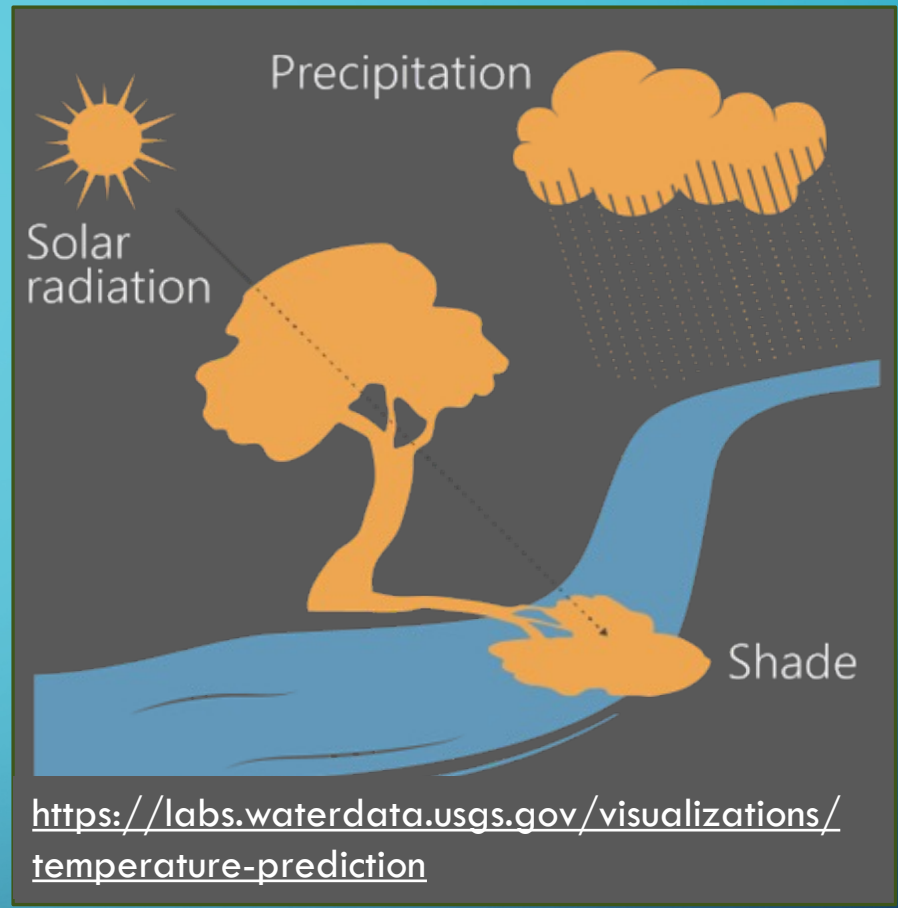


We are organizing our best practices into the data lifecycle categories.

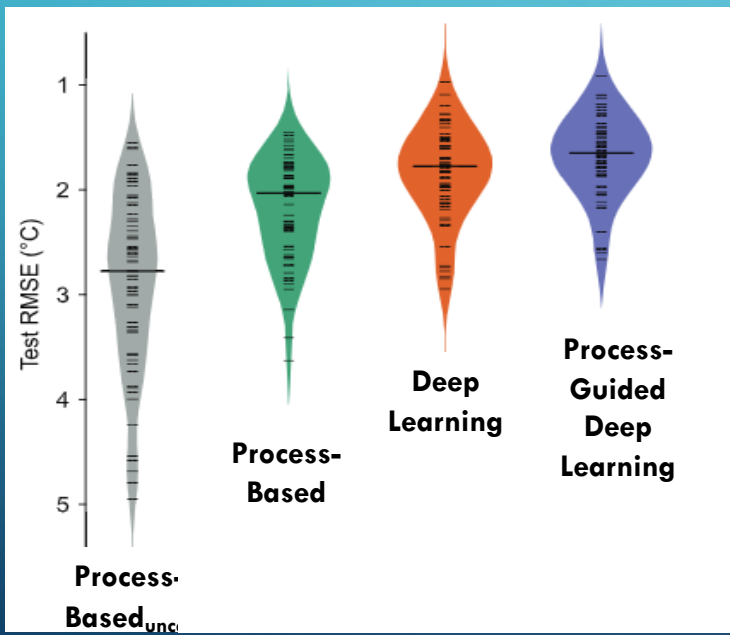




[Data Science for Water Resources | U.S. Geological Survey \(usgs.gov\)](https://www.usgs.gov/data-science)



Figures by H. Corson-Dosch; data from Jia et al. 2021

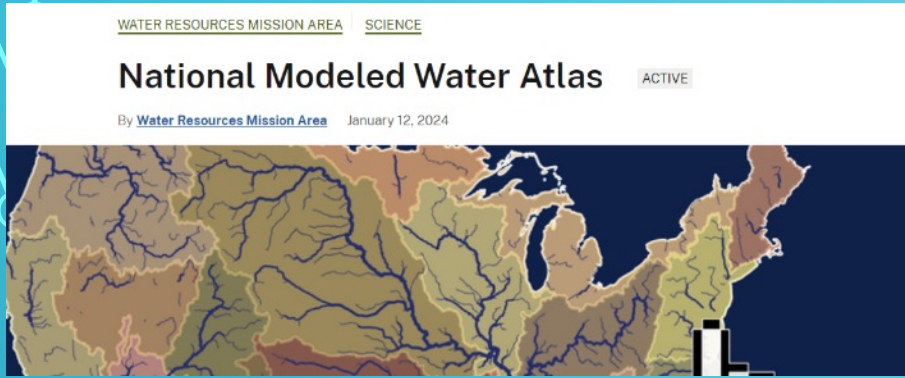


[Process-Guided Deep Learning Predictions of Lake Water Temperature - Read - 2019 - Water Resources Research - Wiley Online Library](#)

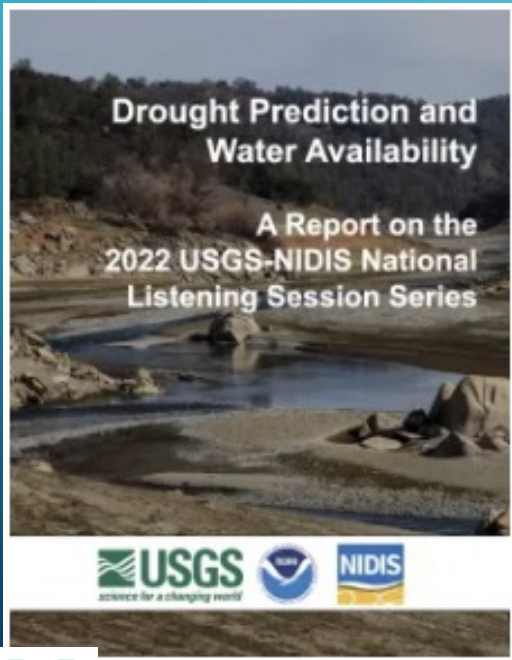
Deep Learning model “learns” physics through pre-training with thermodynamic model predictions



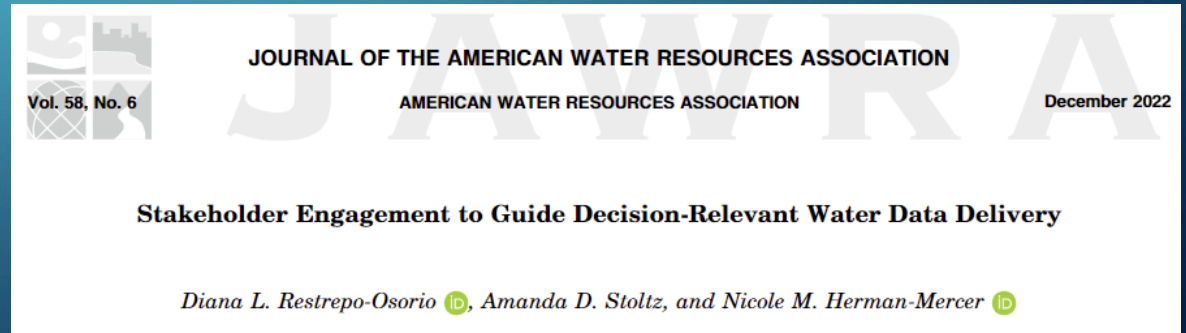
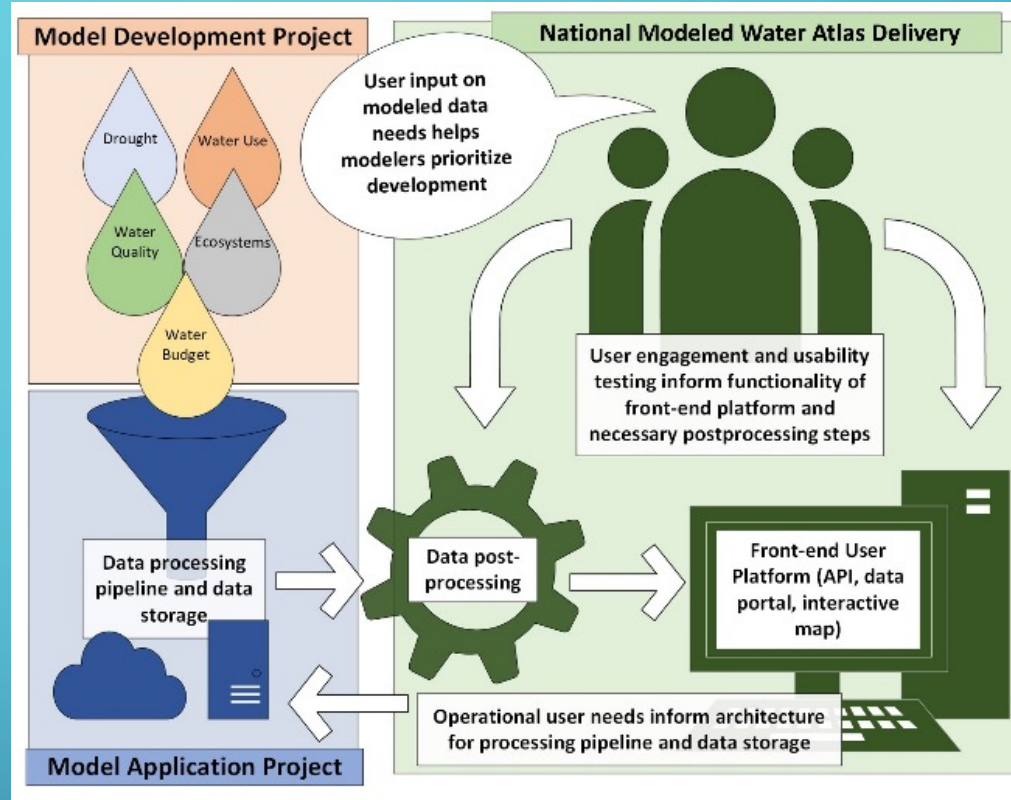
# WHO NEEDS WHAT DATA IN WHAT FORMAT TO MAKE WHAT DECISION?



[National Modeled Water Atlas | U.S. Geological Survey \(usgs.gov\)](https://www.usgs.gov/national-modeled-water-atlas)



[Drought Prediction and Water Availability: A Report on the 2022 USGS-NIDIS National Listening Session Series | Drought.gov](https://drought.gov)

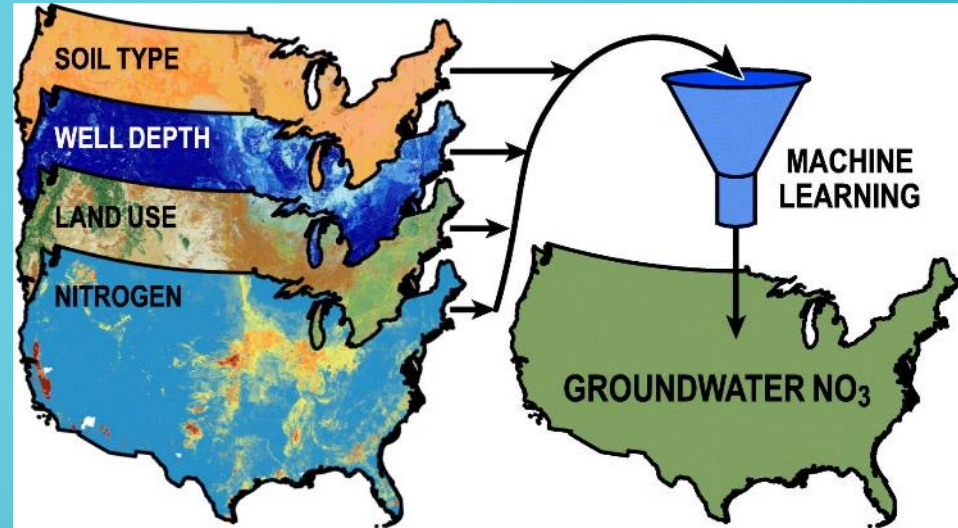


[Stakeholder Engagement to Guide Decision-Relevant Water Data Delivery \(wiley.com\)](https://onlinelibrary.wiley.com/doi/10.1111/jawra.12588)

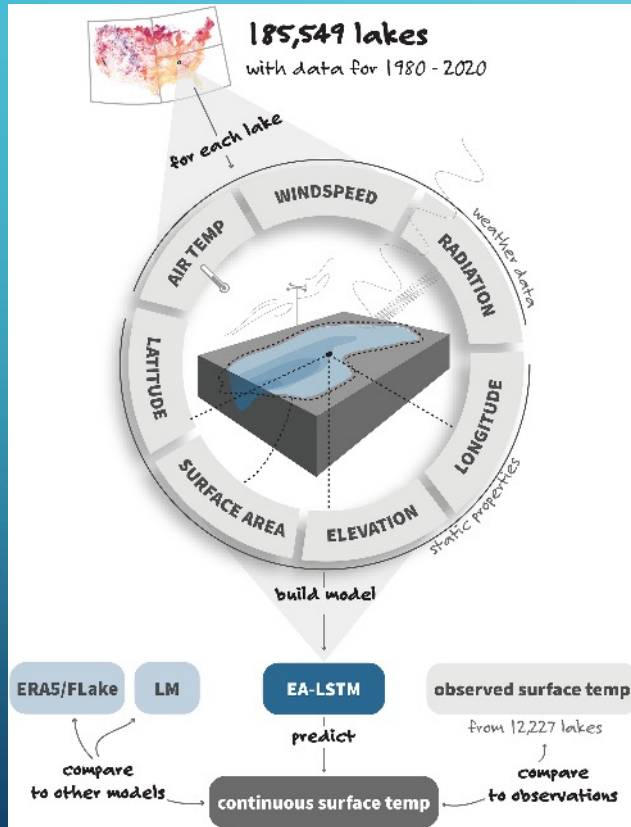


# Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States

K.M. Ransom <sup>a</sup>, B.T. Nolan <sup>b</sup>, P.E. Stackelberg <sup>c</sup>, K. Belitz <sup>d</sup>, M.S. Fram <sup>a</sup>



[Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States - ScienceDirect](#)



LIMNOLOGY AND OCEANOGRAPHY  
**LETTERS**  
ASLO  
Open Access

Data Article | Open Access |

## Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020)

Jared D. Willard , Jordan S. Read, Simon Topp, Gretchen J. A. Hansen, Vipin Kumar

First published: 17 March 2022 | <https://doi.org/10.1002/lol2.10249> | Citations: 5

[Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning \(1980–2020\) - Willard - 2022 - Limnology and Oceanography Letters - Wiley Online Library](#)

# USGS Vizlab

water data visualizations

## Counts of bottling facilities in Texas by county

597 Bottled water facilities

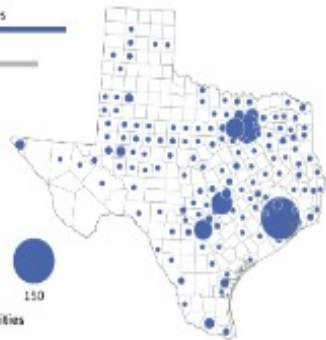
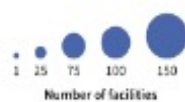
250 Breweries

339 Distilleries

57 Ice facilities

168 Soft drink facilities

250 Wineries



Water bottling across the U.S.

View

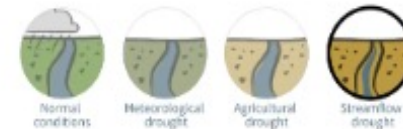
Code



Five droughts that changed history

View

Code



What is streamflow drought?

View

Code



The Water Cycle

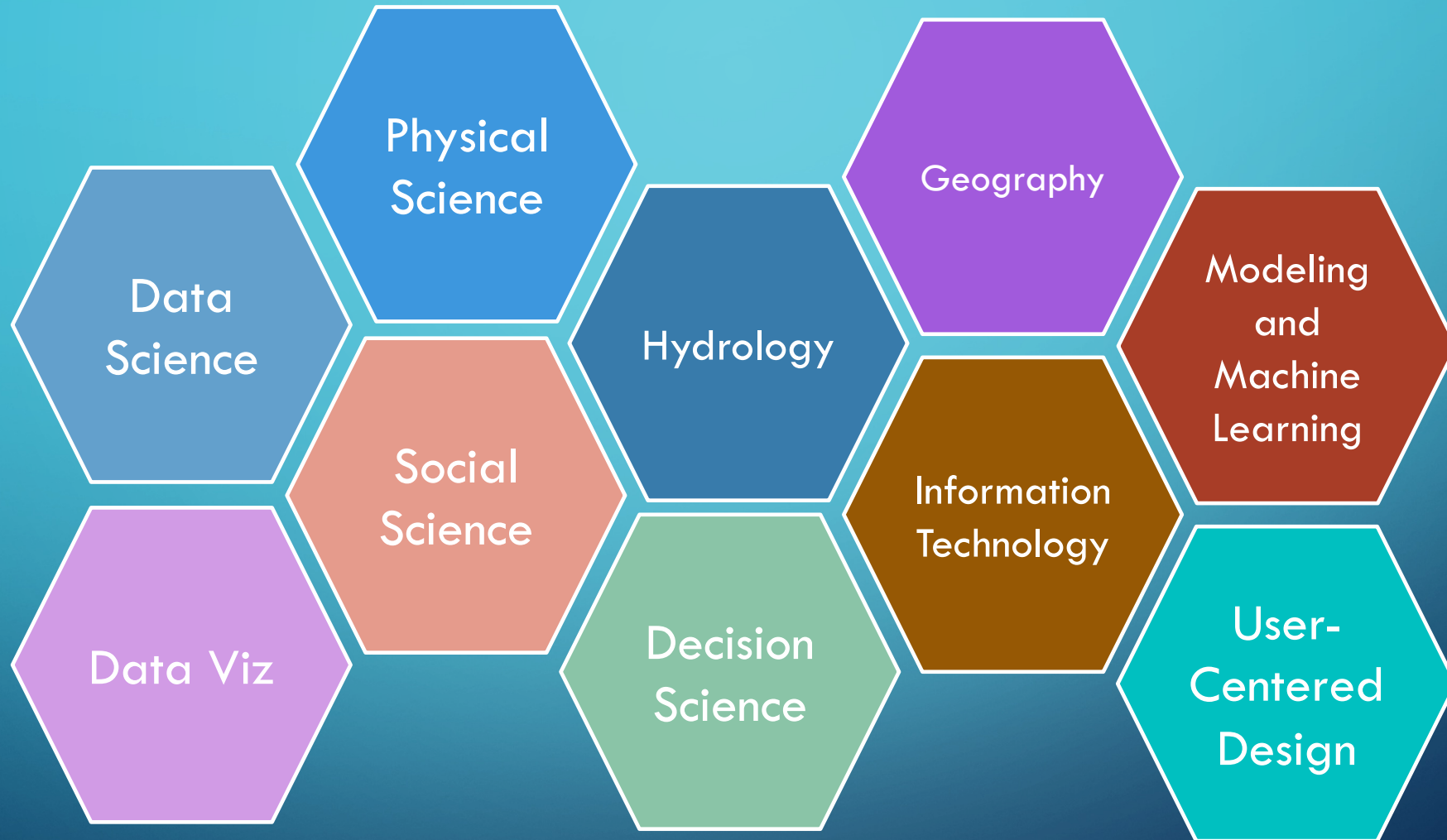
View

Code

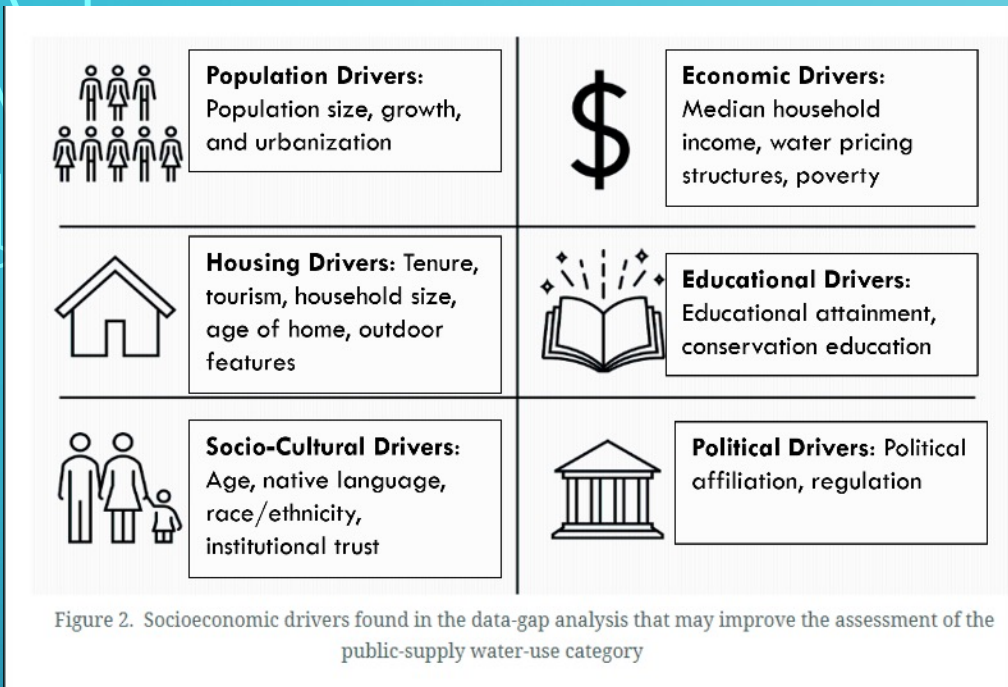
<https://labs.waterdata.usgs.gov/visualizations>



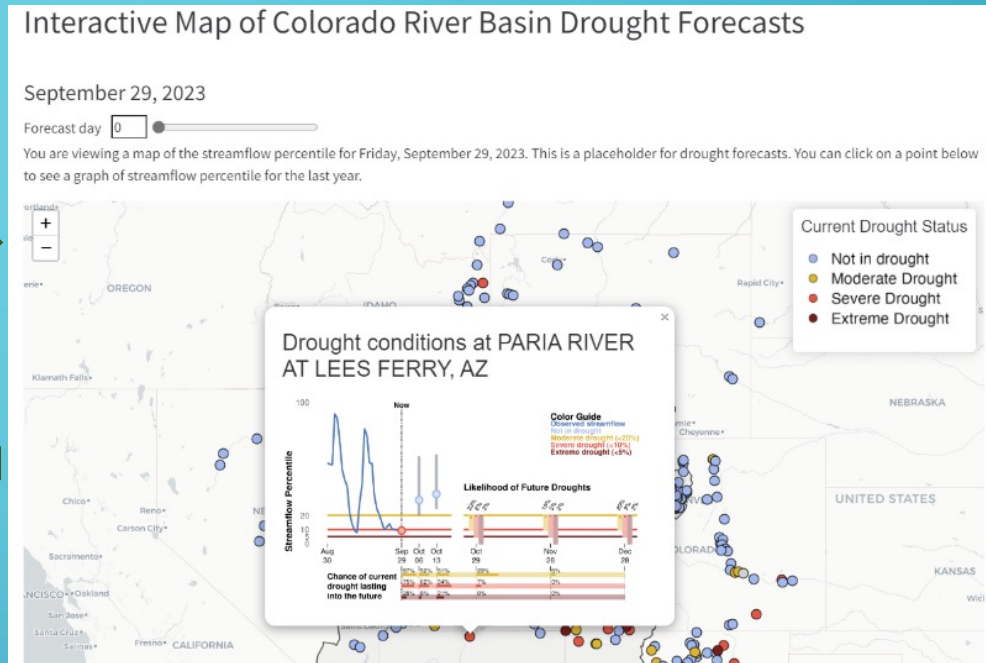
# INTERDISCIPLINARY APPROACHES TO COMPLEX SYSTEMS



# SOPHISTICATED MODELING APPROACHES TO COMPLEX SYSTEMS



From "Water-Use Data Gap Analysis," USGS GeoNarrative



Colorado River Basin Drought Forecasts - Interactive Map of Colorado River Basin Drought Forecasts (usgs.gov), Internal Only

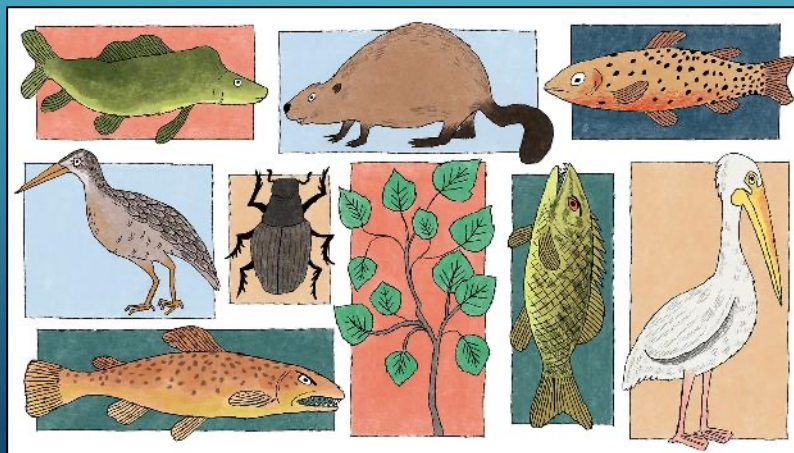


Illustration by Millie von Platen, Vox 2023

<b>Overall Vulnerability</b>	<b>Socioeconomic Status</b>	<ul style="list-style-type: none"> <li>Below 150% Poverty</li> <li>Unemployed</li> <li>Housing Cost Burden</li> <li>No High School Diploma</li> <li>No Health Insurance</li> </ul>
	<b>Household Characteristics</b>	<ul style="list-style-type: none"> <li>Aged 65 &amp; Older</li> <li>Aged 17 &amp; Younger</li> <li>Civilian with a Disability</li> <li>Single-Parent Households</li> <li>English Language Proficiency</li> </ul>
	<b>Racial &amp; Ethnic Minority Status</b>	<ul style="list-style-type: none"> <li>Hispanic or Latino (of any race)</li> <li>Black or African American, Not Hispanic or Latino</li> <li>Asian, Not Hispanic or Latino</li> <li>American Indian or Alaska Native, Not Hispanic or Latino</li> <li>Native Hawaiian or Pacific Islander, Not Hispanic or Latino</li> <li>Two or More Races, Not Hispanic or Latino</li> <li>Other Races, Not Hispanic or Latino</li> </ul>
	<b>Housing Type &amp; Transportation</b>	<ul style="list-style-type: none"> <li>Multi-Unit Structures</li> <li>Mobile Homes</li> <li>Crowding</li> <li>No Vehicle</li> <li>Group Quarters</li> </ul>

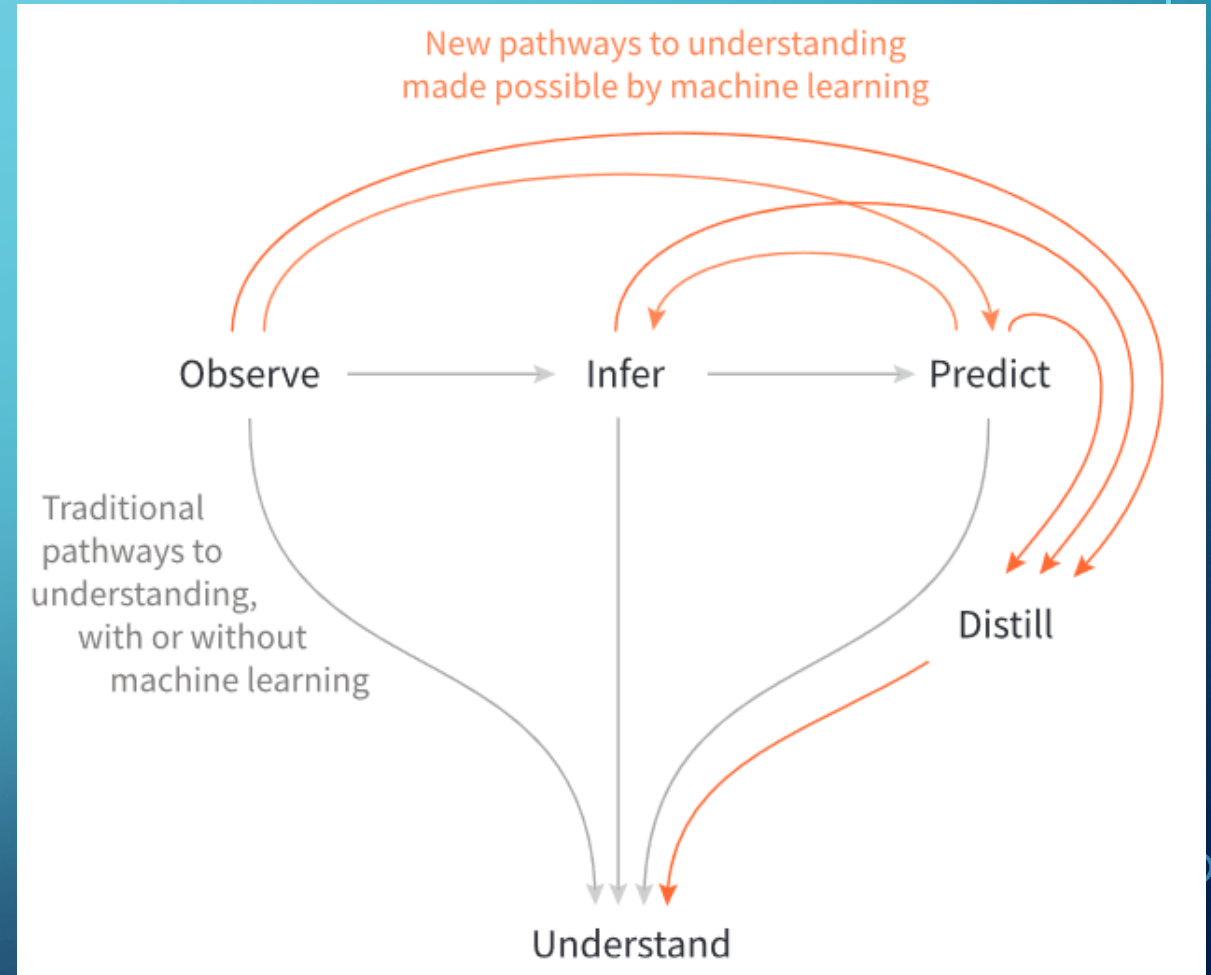
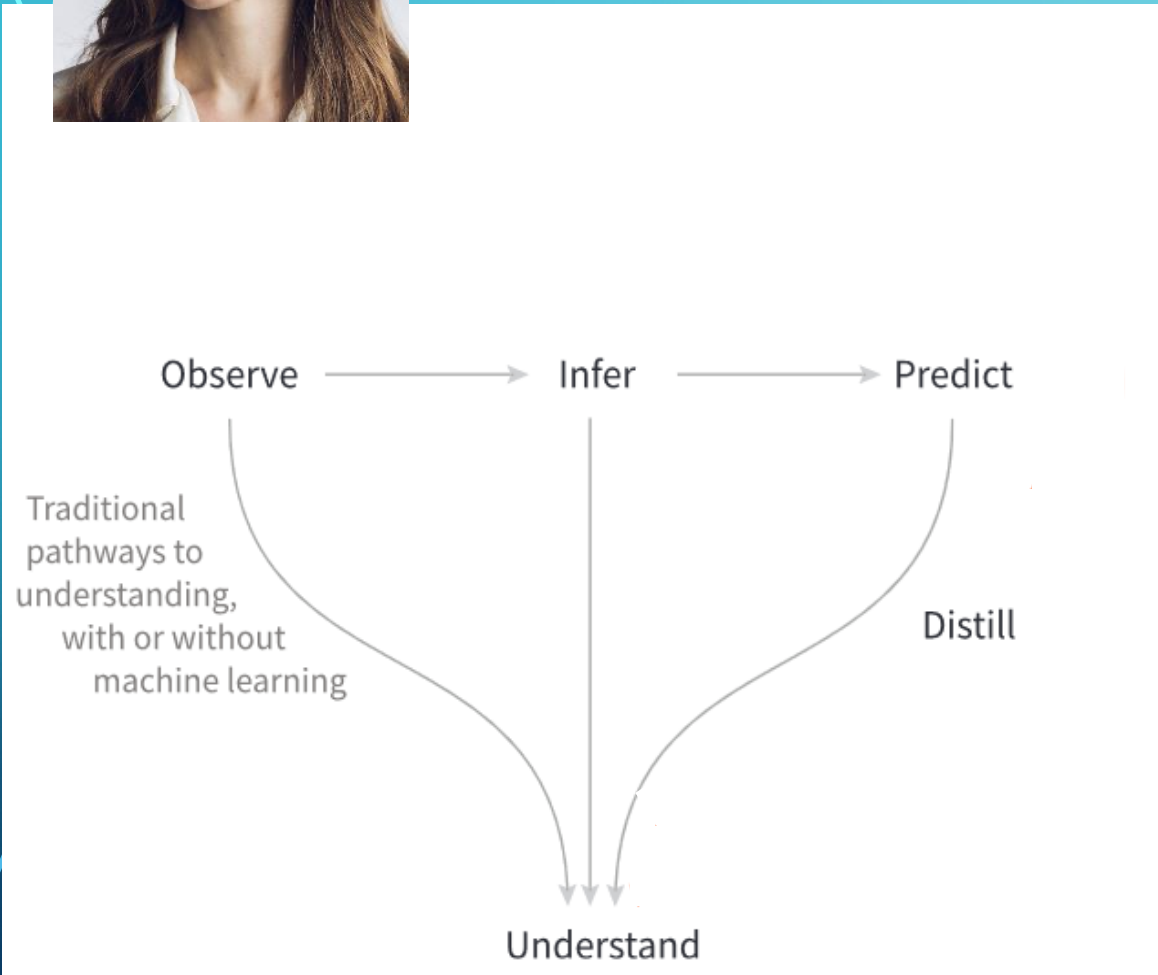
CDC Social Vulnerability Index, 2020





Cassie Kozyrkov,  
Decision Scientist

# AI = “AUTOMATED INSPIRATION”



Figures from [Appling et al. 2022](#)

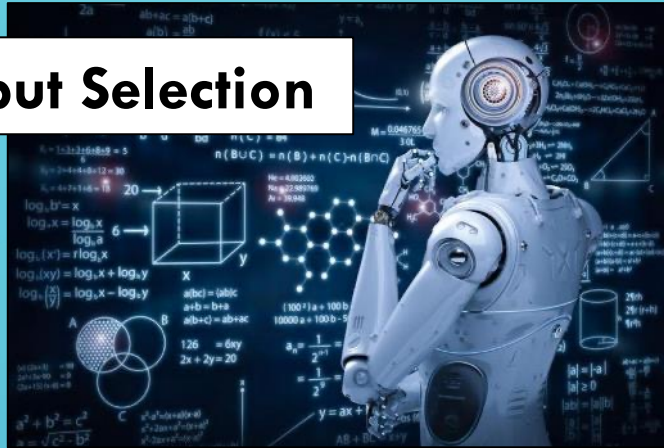
# AI/ML IS NOT “ONE-SIZE-FITS-ALL”

**Preprocessing**



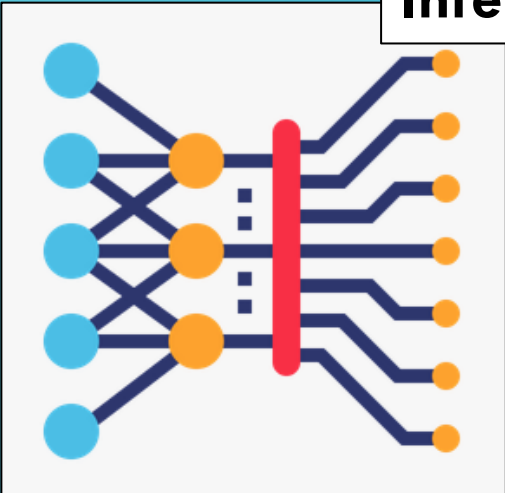
Icon designed by [Ida Desi Mariana](#) for FlatIcon.com

**Input Selection**



<https://viso.ai/deep-learning/data-preprocessing-techniques-for-machine-learning-with-python/>

**Inference**

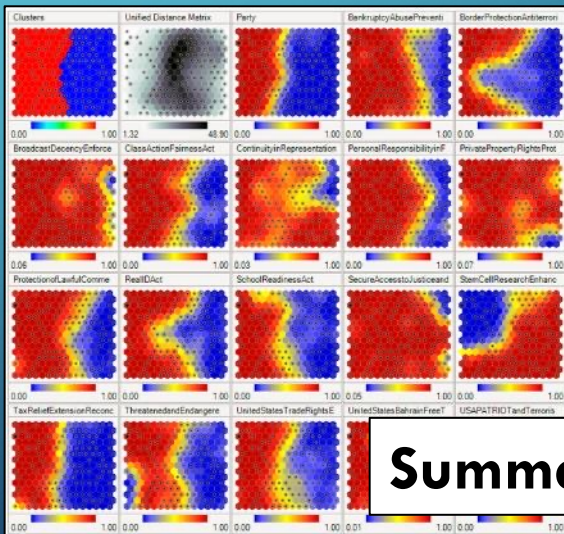


Icon designed by [Becris](#) for FlatIcon.com

**Forecasting**



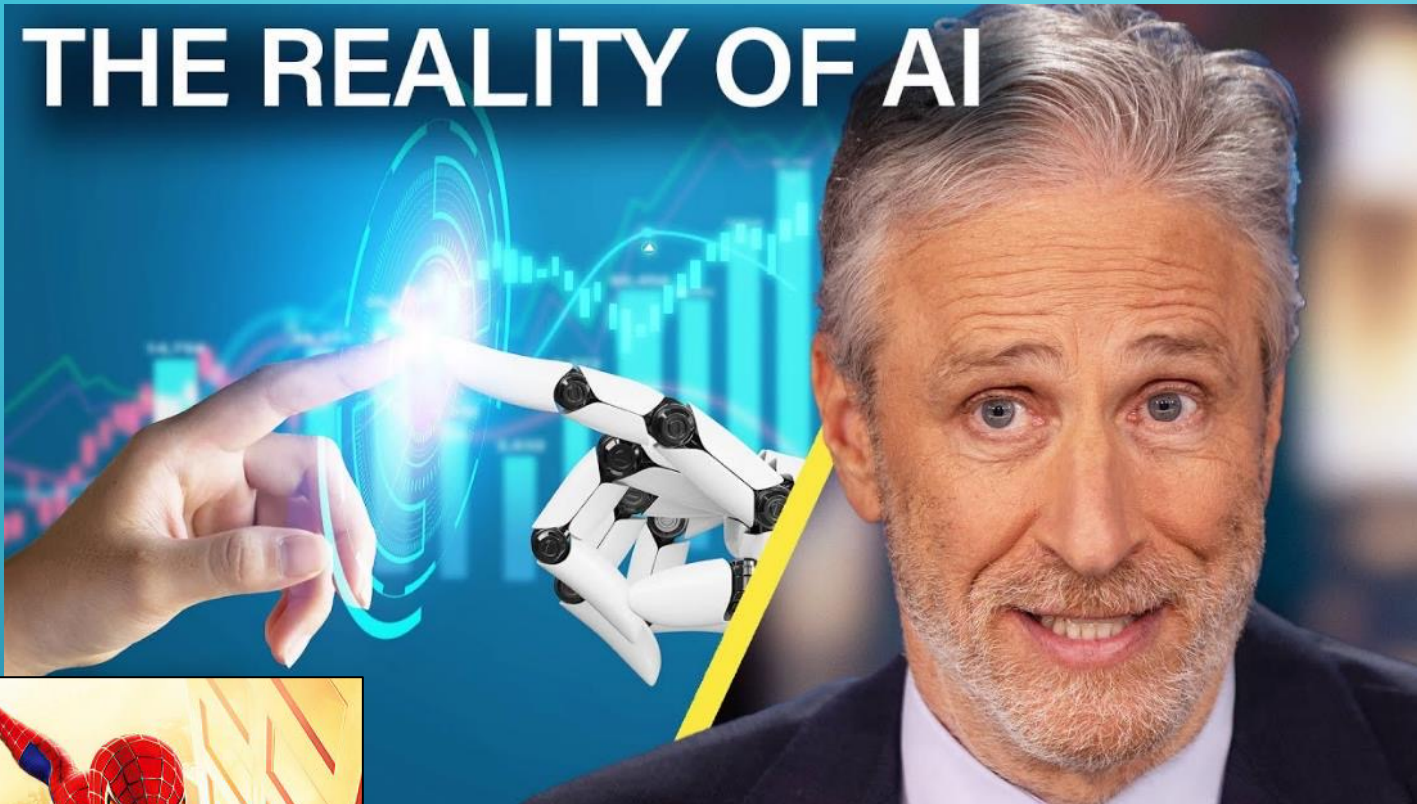
Icon designed by [Parzival' 1997](#) for FlatIcon.com



**Summarization**

[https://en.wikipedia.org/wiki/Self-organizing\\_map](https://en.wikipedia.org/wiki/Self-organizing_map)

# THE REALITY OF AI



The Daily Show, April 1, 2024

“As a society, we’ve been through technological advances before. And they’ve all promised a utopian life without the drudgery, but the reality is they come for our jobs...Whether it’s globalization or industrialization or now artificial intelligence, the way of life that you are accustomed to is no match for the promise of more profits and new markets...at least those other disruptions took place over a century or decades. AI is gonna be ready to take over by Thursday...”

Jon Stewart





THANK YOU!

# Google

can artificial intelligence

can artificial intelligence

can artificial intelligence **be dangerous**

can artificial intelligence **be dangerous explain with evidence**

can artificial intelligence **replace human intelligence**

can artificial intelligence **take over the world**

can artificial intelligence **lie**

can artificial intelligence **replace humans debate**

can artificial intelligence **become self aware**

can artificial intelligence **write a book**

can artificial intelligence **write code**

Google Search

I'm Feeling Lucky

*Report inappropriate predictions*

# QUESTIONS?

Katrina Alger  
USGS Water Mission Area  
Integrated Information  
Dissemination Division  
kalger@usgs.gov